Asymptotic Behavior of Bayesian Learners with Misspecified Models *

Ignacio Esponda (UC Santa Barbara)
Demian Pouzo (UC Berkeley)
Yuichi Yamamoto (UPenn)

May 28, 2019

Abstract

We consider an agent who represents uncertainty about her environment via a possibly misspecified model. Each period, the agent takes an action, observes a consequence, and uses Bayes' rule to update her belief about the environment. This framework has become increasingly popular in economics to study behavior driven by incorrect or biased beliefs. Current literature has either characterized asymptotic behavior in general settings under the assumption that the agent's action converges (which sometimes does not) or has established convergence of the action in specific applications. By noting that the key element to predict the agent's behavior is the frequency of her past actions, we are able to characterize asymptotic behavior in general settings in terms of the solutions of a generalization of a differential equation that describes the evolution of the frequency of actions. Among other results, we provide a new interpretation of mixing in terms of convergence of the frequency of actions, and we also show that convergent frequencies of actions are not necessarily captured by previous Nash-like equilibrium concepts.

---

*Preliminary; please do not distribute or post online. Acknowledgements to be added. Esponda: Department of Economics, UC Santa Barbara, 2127 North Hall, Santa Barbara, CA 93106, iesponda@ucsb.edu; Pouzo: Department of Economics, UC Berkeley, 530-1 Evans Hall #3880, Berkeley, CA 94720, dpouzo@econ.berkeley.edu. Yamamoto: Office 504, University of Pennsylvania Department of Economics, The Ronald O. Perelman Center for Political Science and Economics, 133 South 36th Street Philadelphia, PA 19104.

Contents

1 Introduction

Any model is an approximation to the truth. Not surprisingly, a large statistics and econometrics literature studies how to tackle model misspecification. A more recent literature studies models where the agents themselves, not just the econometrician, have misspecified models of their environments. The world is complex and it is natural for economic agents to represent uncertainty about the world with parsimonious models that are likely to be misspecified. Examples include:

* Several firms compete by setting prices, and their sales depend randomly on their own and their competitors' prices. Prices are not observed in this industry, and one of the largest firms decides to ignore other firms' prices and estimate a demand model where sales depend only on its own price. (Arrow and Green (1973), Kirman (1975)).

* An instructor observes the initial performance of a student, decides to either praise or criticize her, and then observes the student's final performance. The two performances are independent, but the instructor does not understand regression to the mean and believes that the student's improvement from one performance to the other depends on the decision to praise or criticize the student. (Tversky and Kahneman (1973), Esponda and Pouzo (2016)).

* A seller thinks that she faces a constant-elasticity demand function, but does not know this elasticity. In reality, the elasticity is not constant but is high for low prices and low for high prices. (Nyarko (1991), Esponda and Pouzo (2016), Fudenberg, Romanyuk and Strack (2017)).

* A person faces a nonlinear tax schedule where the marginal tax rate is increasing in income. The person, however, behaves as if she were in a simpler world where she faces a linear tax schedule with a constant marginal tax. (Sobel (1984), Liebman and Zeckhauser (2004), Esponda and Pouzo (2016)).

* Performance pay increases productivity due to an incentive effect and a sorting (or selection) effect (Lazear (2000)). The sorting effect occurs over time and is due to higher turnover for less productive workers and higher productivity of new hires. A firm learns how its current workers respond to incentives but fails to account for the sorting effect. (Esponda (2008))

* A worker must decide how much work to delegate to a teammate. The expected output of the team depends on the worker's own ability, her teammate's ability, and the extent of delegation. The worker learns about her teammate's ability from observations of output, but this learning is biased by overconfident beliefs about her own ability. (Heidhues, Kőszegi and Strack (2018))

* A person wishing to improve his health considers abstaining from a food he likes. In reality, the person's choice and health are independent, but each of them separately have some influence on some chemical in the blood. The person, however, inverts the causal link and believes that his decision affects the chemical reading, which in turn affects his health. (Spiegler (2016))

In all of these examples, the agent processes information through the lens of a simple model that misses some aspect of reality. The main question in the literature is what happens to the agent's behavior as time goes by and she uses feedback to update her belief about the model's primitives.[1] The agent will of course never learn the true primitives, but one would like to know if behavior will remain suboptimal, and, if so, what is the direction of the bias. An appealing feature of these examples is that the direction of the bias is not ex ante obvious. The agent's behavior affects the feedback she observes. This feedback is processed via the agent's misspecified model, which leads to updated beliefs and subsequent changes in behavior, which in turn lead to changes in beliefs, and so on.

The literature has developed several approaches to study the agent's asymptotic belief and behavior in the kinds of examples described above. Typically, progress has been made one example at a time, and the main contribution of this paper is to develop tools that can be used to characterize asymptotic behavior for a large class of environments. Before describing our main contributions in more detail, we first describe the environment, previous results, and existing challenges.

Time is discrete and there is a single, infinitely-lived agent who discounts the future and must take an action in each period. The agent's action potentially affects the distribution of an observable variable, which we call a consequence. Her per-period payoff depends on the agent's action and the realized consequence. The true distribution over consequences as a function of an action $x \in \mathbb{X}$ is given by $Q(\cdot \mid x) \in \Delta(\mathbb{Y})$, where $\mathbb{Y}$ is the set of consequences. The agent, however, does not know Q. She has a parametric model of it, given by $(Q_\theta(\cdot \mid x))_{x \in \mathbb{X}}$, where parameter values, such as θ, belong to a parameter space Θ. The agent is Bayesian, so she has a prior over Θ and updates her prior in each period after observing the realized consequence. The agent's model is misspecified if the support of her prior does not include

[1]There are at least three reasons why focusing on asymptotic behavior is interesting. First, there are many instances where it is not surprising that people initially make incorrect decisions; the more interesting question is what types of biases persist with experience. Second, systematic patterns tend to arise as time goes by, while initial behavior tends to be more dependent on random draws. Finally, there is a long tradition in statistic and economics focusing on asymptotic or equilibrium behavior, and so we can use existing tools as well as compare our results to existing results in these literatures.

the true distribution Q, and she is correctly specified otherwise.[2]

The literature has examined this type of problem from two complementary perspectives. The first perspective focuses on charactering steady states, also known as equilibria.[3] Esponda and Pouzo (2016; henceforth EP2016) define a Berk-Nash equilibrium to be a probability distribution over actions, $\sigma \in \Delta(\mathbb{X})$, with the property that there exists a belief over Θ such that (i) any action in the support of σ is myopically optimal given the belief, meaning that it maximizes the perceived expected per-period payoff, and (ii) any parameter value θ in the support of the belief minimizes the Kullback-Leibler divergence (henceforth, KLD) between the true distribution over action-consequence pairs and the parameterized distribution, where these distributions in turn depend on σ.

EP2016 show that, if the agent's action converges, then it converges to a pure (i.e., degenerate) Berk-Nash equilibrium distribution. There are cases where such pure equilibria do not exist, and so the agent's action does not converge. A mixed-strategy (i.e., nondegenerate) equilibria, however, always exists. To see how behavior might converge to a mixed strategy, EP2016 add random perturbations to the agent's payoff (as in Harsanyi (1973) and Fudenberg and Kreps (1993)). With payoff perturbations, the agent's behavior may be characterized by a non-degenerate distribution over actions in each time period. EP2016 show that if the distribution of actions converges to a non-degenerate distribution, then the convergent distribution must be a mixed-strategy Berk-Nash equilibrium.[4]

While payoff perturbations are natural in many instances, a remaining question is what happens in the many cases where the action does not converge and we do not assume the existence of payoff perturbations. Our first contribution is to answer this question and to show that we can still characterize important regularities in behavior. As a byproduct, we gain additional insights about the nature of the concept of equilibrium as it is commonly applied in economics.

Our approach deviates from existing literature by considering the main object of interest to be the frequency of actions, not the actions themselves. So, even if the agent's action does not converge, it might be that the frequency of actions converges, thus providing a predictable picture regarding steady-state behavior. We show that, if the frequency of actions happens

[2]The correctly-specified version of this environment was originally studied by Easley and Kiefer (1988) and Aghion, Bolton, Harris and Jullien (1991).

[3]There are many examples of boundedly-rational equilibrium concepts that abstract away from the question of dynamics and convergence, including Jehiel (1995, 2005), Osborne and Rubinstein (1998), Eyster and Rabin (2005), Esponda (2008), Jehiel and Koessler (2008), and Spiegler (2016, 2017). 2017).

[4]EP2016 define Berk-Nash equilibrium for an environment with multiple agents; the appropriate comparison for us is with respect to the special case of a single agent.

to converge to some $\sigma \in \Delta(\mathbb{X})$, then for each action x in the support of σ there is a belief μ_x over Θ such that the following conditions are satisfied: (i) x is an optimal action in the dynamic optimization problem where the agent maximizes discounted expected utility with current belief state μ_x, and (ii) any θ in the support of μ_x minimizes KLD as described earlier. We call any σ satisfying these conditions an equilibrium.[5]

This notion of equilibrium differs from others notions, such as Nash equilibrium or its generalization, Berk-Nash equilibrium, in two regards. First, each action may be supported by a different belief. Second, optimality is defined with respect to the agent's dynamic problem, where the belief is a state variable that may change in the next period. The reason for these features is that, under model misspecification, beliefs may not converge. So, even though the frequency of actions may converge, different actions might be taken at different times and under different beliefs. Also, since the agent anticipates that her belief might change in the future, she chooses an action that is dynamically optimal in the sense of taking the continuation value, with a possibly different belief, into account.

We next investigate conditions such that the two convenient properties that are standard in most equilibrium concepts (unique supporting belief and myopic, not dynamic, optimality) do indeed hold in our environment. We show that these two properties hold if the agent's model is weakly identified, which roughly means that the agent's belief is uniquely determined for actions in the support of the convergent frequency, but leaves open the possibility of multiple beliefs for off-the-path actions. In other words, under weak identification, the appropriate equilibrium concept for frequencies of actions is Berk-Nash equilibrium.

The perspective of focusing on steady states is attractive because the results hold very generally and equilibrium analysis has a long tradition in economics. But steady-state analysis alone cannot tell us if if the actions, or even their frequency, will converge or not. The second perspective in the literature is to study asymptotic behavior, and not just steady states, to determine if behavior converges or not. Success in this regard has been achieved in a few specific applications, and even in these specific cases the focus has been on convergence (or not) of actions, not frequencies.

Nyarko (1991) studies the example of a monopolist learning its demand function. The monopolist chooses one of two prices and updates between two possible demand models, neither of which is the true model. Nyarko shows that the monopolist's price does not converge. Fudenberg, Romanyuk and Strack (2017) consider a more general model where the agent has a finite number of actions but still updates between two possible models (i.e., Θ has two

[5]In the paper, we consider the more general case where the agent follows a policy that may not be optimal. We focus on optimal policies in the introduction in order to contrast our work with previous literature.

elements). They provide a full characterization of asymptotic actions and beliefs, including cases where the action converges and cases where it does not. Their model is in continuous time and they exploit the fact that the belief over Θ follows a single-dimensional stochastic differential equation. Heidhues, Kőszegi and Strack (2018) study a model of an agent whose overconfidence biases his learning of a fundamental that is relevant for determining the optimal action. They are able to establish convergence by exploiting the monotone structure of their environment.

Our second main contribution is to propose a method to study asymptotic behavior that holds generally across environments and does away with many specific assumptions. Our key point of departure, once again, is the focus on the evolution of the frequency of actions rather than on actions alone or on the agent's belief. The frequency of actions at time $t+1$ can be written recursively as a function of the frequency at time t plus some innovation term that depends on the agent's action at time $t+1$. The action at time $t+1$, however, depends on the agent's belief at time t, and one challenge is to be able to write this belief as a function of frequencies of actions so as to make this recursion depend exclusively on frequencies, not beliefs.

Extending results by Berk (1966) and EP2016, we show that eventually the posterior at time t roughly concentrates on the set of parameter values that minimize KLD given the frequency of actions up to time t. This result allows us to write the evolution of frequencies of actions recursively as a function of the past frequency alone, excluding the belief. We then apply techniques from stochastic approximation developed by Benaïm, Hofbauer and Sorin (2005) to show that the continuous-time approximation of the frequency of actions can be essentially characterized as a solution to a generalization of a differential equation.[6] This approach can be applied to a wide range of problems, and we illustrate its value by revisiting existing applications in the literature.[7]

Misspecified learning has also been studied in other types of environments. Rabin and Vayanos (2010) study an environment where shocks are i.i.d. but agents believe them to be

[6] The type of differential equation is called a differential inclusion in the literature. It differs from a differential equation in that there may be multiple derivatives at certain points and therefore multiple trajectories that solve the equation. Multiplicity arises in our environment because there are certain beliefs at which the agent is indifferent between different actions, and we need to keep track of what would happen to beliefs and subsequent actions if the agent were to follow any one of these.

[7] Tools from stochastic approximation have been previously applied in economics, including the literature on learning in games (e.g., Fudenberg and Kreps (1993), Benaim and Hirsch (1999), and Hofbauer and Sandholm (2002)) and learning in macroeconomics (e.g., Sargent (1993)). Our approach is inspired by Fudenberg and Kreps's (1993) model of stochastic fictitious play. In that environment, the frequency of past actions exactly represents the agents' beliefs about other agents' strategies. In our environment, we characterize beliefs to be a function of the frequency of actions.

autoregressive. Esponda and Pouzo (2017b) extend Berk-Nash equilibrium to Markov decision problems, where a state variable other than a belief affects continuation values. Molavi (2018) studies a general-equilibrium framework that nests a large class of macroeconomic models where agents learn with misspecified models. Finally, Bohren and Hauser (2018), and Frick, Iijima and Ishii (2019) characterize asymptotic behavior in social learning environments with model misspecification.[8] The results from this paper should also be useful in these other settings.

We present the model in Section 2, characterize asymptotic beliefs in Section 3, carry out the analysis of steady states in Section 4, and provide a characterization of asymptotic behavior in Section 5. We conclude in Section 6 by discussing directions for further research.

2 The environment

Objective environment. There is a single agent facing the following infinitely repeated problem. Each period $t = 1, 2, \ldots$, the agent must choose an action from a finite set $\mathbb{X}$. She then receives a consequence according to the consequence function $Q : \mathbb{X} \to \Delta(\mathbb{Y})$, where $\mathbb{Y}$ is the set of consequences and $\Delta(\mathbb{Y})$ is the set of all (Borel) probability measures over it. Finally, the payoff function $\pi : \mathbb{X} \times \mathbb{Y} \to \mathbb{R}$ determines the agent's current payoff. In particular, if $x_t \in \mathbb{X}$ is the agent's choice at time t, then $y_t \in \mathbb{Y}$ is drawn according to the probability measure $Q(\cdot \mid x) \in \Delta(\mathbb{Y})$, and the agent's payoff at time t is $\pi(x_t, y_t)$.

Assumption 1. *$\mathbb{Y}$ is a compact subset of Euclidean space, and, for all $x \in \mathbb{X}$, the support of $Q(\cdot \mid x)$ is contained in $\mathbb{Y}$; (ii) There exists a Borel probability measure $\nu \in \Delta(\mathbb{Y})$ such that, for all $x \in \mathbb{X}$, $Q(\cdot|x) \ll \nu$, i.e., $Q(\cdot|x)$ is absolutely continuous with respect to ν (an implication is the existence of densities $q(\cdot \mid x) \in L^1(\mathbb{Y}, \mathbb{R}, \nu)$ such that $\int_A q(y \mid x)\nu(dy) = Q(A|x)$ for any $A \subseteq \mathbb{Y}$ Borel); (iii) For all $x \in \mathbb{X}$, $\pi(x, \cdot) \in L^1(\mathbb{Y}, \mathbb{R}, Q(\cdot \mid x))$.*[9]

Assumption 1 collects some standard technical conditions. It includes both the case where the consequence is a continuous variable (ν is the Lebesgue measure and $q(\cdot \mid x)$ is the density function) and the case where it is discrete (ν is the counting measure and $q(\cdot \mid x)$ is the probability mass function).

In the special case in which the agent knows the primitives and wishes to maximize discounted expected utility, she chooses an action in each period from the set of actions that

[8] See also Eyster and Rabin (2010), Bohren (2016), and Gagnon-Bartsch and Rabin (2017).

[9] As usual, $L^p(\mathbb{Y}, \mathbb{R}, \nu)$ denotes the space of all functions $f : \mathbb{Y} \to \mathbb{R}$ such that $\int |f(y)|^p \nu(dy) < \infty$.

maximizes

$$\int_{\mathbb{Y}} \pi(x,y)Q(dy \mid x) = \int_{\mathbb{Y}} \pi(x,y)q(y|x)\nu(dy).$$

We will study the case where the agent does not know the consequence function Q.

Subjective model. The agent is endowed with a family of consequence functions, $\mathcal{Q}_\Theta = \{Q_\theta : \theta \in \Theta\}$, where each $Q_\theta : \mathbb{X} \to \Delta(\mathbb{Y})$ is indexed by a parameter value $\theta \in \Theta$. We refer to $\mathcal{Q}_\Theta$ as the subjective model and say that it is *correctly specified* if $Q \in \mathcal{Q}_\Theta$ and *misspecified* otherwise.

Assumption 2. *(i) For all $\theta \in \Theta$ and $x \in \mathbb{X}$, $Q_\theta(\cdot|x) \ll \nu$, where ν is defined in A1 (an implication is the existence of densities $q_\theta(\cdot \mid x) \in L^1(\mathbb{Y}, \mathbb{R}, \nu)$ such that $\int_A q_\theta(y \mid x)\nu(dy) = Q_\theta(A|x)$ for any $A \subseteq \mathbb{Y}$ Borel); (ii) Θ is a compact subset of an Euclidean space and, for all $x \in \mathbb{X}$, $\theta \mapsto q_\theta(\cdot \mid x)$ is continuous $Q(\cdot \mid x)$-a.s.; ; (iii) For all $x \in \mathbb{X}$, there exists $g_x \in L^2(\mathbb{Y}, \mathbb{R}, Q(\cdot \mid x))$ such that, for all $\theta \in \Theta$, $|\ln(q(Y \mid x)/q_\theta(Y \mid x))| \leq g_x(y)$ a.s.-$Q(\cdot \mid x)$.*

Assumption 2(i) guarantees the existence of a density function, and 2(ii) is a standard parametric assumption on the subjective model. Assumption 2(iii) will be used to establish a uniform law of large numbers. This condition also implies that, for all θ and x, the support of $Q_\theta(\cdot \mid x)$ contains the support of $Q(\cdot \mid x)$; in particular, every observation can be generated by the agent's model.

Bayesian learning. The agent is Bayesian and starts with a prior μ_0 over the parameter space Θ. She observes past actions and consequences and uses this information to update her belief about Θ in every period. The timing is as follows: At each time t, the agent holds some belief μ_t. Given μ_t, she chooses an action x_t. Then the consequence y_t is drawn according to $Q(\cdot \mid x_t)$. The agent observes y_t, receives an immediate payoff of $\pi(x_t, y_t)$, and updates her belief to $\mu_{t+1} = B(x_t, y_t, \mu_t)$, where B is the Bayesian operator.[10] The next assumption guarantees that the prior has full support.

Assumption 3. $\mu_0(A) > 0$ *for any A open and non-empty.*

Policy and probability distribution over histories. A *policy* f is a function $f : \Delta(\Theta) \to \mathbb{X}$ specifying the action $f(\mu) \in \mathbb{X}$ that the agent takes at any moment in time in which her belief is μ. A *history* is a sequence $h = (x_0, y_0, ..., x_t, y_t, ...) \in \mathbb{H} \equiv (\mathbb{X} \times \mathbb{Y})^\infty$. Together with the

[10] The Bayesian operator $B : \mathbb{X} \times \mathbb{Y} \times \Delta(\Theta) \to \Delta(\Theta)$ satisfies, for all $A \subseteq \Theta$ Borel, for any $x \in \mathbb{X}$, and a.s.-$Q(\cdot \mid x)$, $B(x,y,\mu)(A) = \int_A q_\theta(y \mid x)\mu(d\theta) / \int_\Theta q_\theta(y \mid x)\mu(d\theta)$.

primitives of the problem, a policy f induces a probability distribution over the set of histories, which we will denote by $\boldsymbol{P}^f$.[11]

Policy correspondence. It will be convenient to characterize behavior for a family of policies, and not just for a single policy function. For this purpose, we define a *policy correspondence* to be a mapping $F : \Delta(\Theta) \rightrightarrows \mathbb{X}$, where $F(\mu) \subseteq \mathbb{X}$ denotes the set of actions that the agent might choose any time her belief is $\mu \in \Delta(\Theta)$. Let $Sel(F)$ denote the set of all policies f that constitute a selection from the correspondence F, i.e., with the property that $f(\mu) \in F(\mu)$ for all μ.

An important special case is one where the agent maximizes discounted expected utility with discount factor $\beta \in [0,1)$. This problem can be cast recursively as

$$W(\mu) = \max_{x \in \mathbb{X}} \int_{\mathbb{Y}} \left\{ \pi(x,y) + \beta W(\mu') \right\} \bar{Q}_\mu(dy|x) \tag{1}$$

where $W : \Delta(\Theta) \to \mathbb{R}$ is the (unique) solution to the Bellman equation (1), $\mu' = B(x,y,\mu)$ is the Bayesian posterior, and $\bar{Q}_\mu \equiv \int_\Theta Q_\theta \mu(d\theta)$. In this case, the optimal correspondence, which we will denote by F_β, is such that $F_\beta(\mu)$ is the set of actions that solve the optimization problem in (1) given belief μ and discount factor β. An important, property of the optimal correspondence F_β is that it is upper hemicontinuous (henceforth, uhc). We will be explicit in the statement of our results about any restrictions to the agent's policy. In particular, our main results will only require policies to be a selection from *some* uhc correspondence, not necessarily the optimal one.

The object of interest. Our main objective is to study regularities in asymptotic behavior. Previous work has focused on the sequence of actions. In cases where actions converge, previous work attempts to characterize the limiting action. But there are cases where actions do not converge (e.g., Nyarko (1991)), and in those cases previous work has not much else to say about asymptotic behavior. We make progress by studying the *frequency* of actions. We do so for two reasons. First, from a practical perspective, even if actions do not converge, it is possible for the frequency of actions to converge. Thus, studying frequencies can help uncover additional regularities in behavior, with important implications regarding, for example, average payoffs. Second, as we will show, asymptotic beliefs depend crucially on the frequency of actions. Because actions in turn depend on beliefs, future actions depend crucially on the

[11] Our results also hold if we allow for mixed actions, i.e., f mapping into $\Delta(\mathbb{X})$ instead of $\mathbb{X}$. Not allowing the agent to mix simplifies the exposition and also highlights that a mixed distribution may describe asymptotic behavior even if the agent does not explicitly mix.

frequency of past actions.

For every t, we define the *frequency of actions at time t* to be a function $\sigma_t : \mathbb{H} \to \Delta(\mathbb{X})$ defined such that, for all $h \in \mathbb{H}$ and $x \in \mathbb{X}$,

$$\sigma_t(h)(x) = \frac{1}{t} \sum_{\tau=1}^{t} \mathbf{1}_{(x)}(x_\tau(h))$$

is the fraction of times that action x occurs in history h by time period t.

3 Characterization of asymptotic beliefs

In this section, we take as given the sequence of frequencies of actions, $(\sigma_t)_t$, and we characterize the agent's asymptotic beliefs. In subsequent sections, we will use the characterization of beliefs to characterize the sequence $(\sigma_t)_t$, which is ultimately an endogenous object. The key object in our characterization is the notion of Kullback-Leibler divergence.[12]

Definition 1. The **Kullback-Leibler divergence** (KLD) is a function $K\colon \Theta \times \Delta(\mathbb{X}) \to \mathbb{R}$ such that, for any $\sigma \in \Delta(\mathbb{X})$ and $\theta \in \Theta$,

$$\begin{aligned} K(\theta, \sigma) &= \sum_{x \in \mathbb{X}} E_{Q(\cdot|x)} \ln \left(\frac{q(Y \mid x)}{q_\theta(Y \mid x)} \right) \sigma(x) \\ &= \sum_{x \in \mathbb{X}} \int_{\mathbb{Y}} \ln \left(\frac{q(y \mid x)}{q_\theta(y \mid x)} \right) q(y \mid x) \nu(dy) \sigma(x). \end{aligned}$$

The **set of closest parameter values given** σ is the set $\Theta(\sigma) \equiv \arg\min_{\theta \in \Theta} K(\theta, \sigma)$ and the **minimized KLD given** σ is $K^*(\sigma) \equiv \min_{\theta \in \Theta} K(\theta, \sigma)$.

Lemma 1. *Under Assumptions 1-2: (i)* $(\theta, \sigma) \mapsto K(\theta, \sigma) - K^*(\sigma)$ *is continuous; (ii)* $\Theta(\cdot)$ *is upper hemicontinuous (uhc), nonempty-, and compact-valued.*

Proof. See Appendix A.1. □

If the actions were drawn from an i.i.d. distribution $\sigma \in \Delta(\mathbb{X})$, we could directly apply Berk's (1966) result to conclude that $\lim_{t\to\infty} \mu_t(\Theta(\sigma)) = 1$, i.e., the posterior eventually

[12] Formally, what we call KLD is the Kullback-Leibler divergence between the distributions $q \cdot \sigma$ and $q_\theta \cdot \sigma$ defined over the space $\mathbb{X} \times \mathbb{Y}$.

concentrates on the set of closest parameter values given σ.[13] EP2016 showed that Berk's conclusion extends even if actions are not i.i.d., provided that the distribution over actions at time t converges to a distribution σ. This type of result is useful to characterize behavior under the assumption that it stabilizes, but it is insufficient to determine whether or not behavior stabilizes.

In the current section, we provide a characterization of beliefs that does not rely on the assumption that behavior stabilizes. Roughly speaking, we will show that the distance between the agent's belief at time t, μ_t, and the set of probability measures with support in $\Theta(\sigma_t)$ goes to zero as time goes to infinity, irrespective of whether or not $(\sigma_t)_t$ converges. We will establish this result in several steps, which we now discuss informally and then address formally in the proofs. First, we note that for any Borel set $A \subseteq \Theta$, the posterior belief over A can be written as

$$\begin{aligned}\mu_{t+1}(A) &= \frac{\int_A \prod_{\tau=1}^t q_\theta(y_\tau \mid x_\tau)\mu_0(d\theta)}{\int_\Theta \prod_{\tau=1}^t q_\theta(y_\tau \mid x_\tau)\mu_0(d\theta)} \\ &= \frac{\int_A e^{-tL_t(\theta)}\mu_0(d\theta)}{\int_\Theta e^{-tL_t(\theta)}\mu_0(d\theta)},\end{aligned} \tag{2}$$

where $L_t(\theta) \equiv t^{-1}\sum_{\tau=1}^t \ln \frac{q(y_\tau|x_\tau)}{q_\theta(y_\tau|x_\tau)}$ is the sample average of the log-likelihood ratios, and where we have omitted the history for simplicity. Naturally, we might expect the sample average to converge to its expectation for each θ. The next result strengthens this intuition and establishes that the difference between $L_t(\cdot)$ and $K(\cdot,\sigma_t)$ converges *uniformly* to zero as $t \to \infty$.

Lemma 2. *Under Assumptions 1-2, for any policy f,* $\lim_{t\to\infty}\sup_{\theta\in\Theta}|L_t(\theta) - K(\theta,\sigma_t)| = 0$ $\mathbf{P}^f$-a.s.

Proof. See Appendix A.2. □

The next step is to replace $L_t(\cdot)$ in (2) with $K(\cdot,\sigma_t)$. By Lemma 2, for sufficiently large t, we obtain

$$\mu_{t+1}(A) \approx \frac{\int_A e^{-tK(\theta,\sigma_t)}\mu_0(d\theta)}{\int_\Theta e^{-tK(\theta,\sigma_t)}\mu_0(d\theta)}. \tag{3}$$

As $t \to \infty$, the posterior concentrates on parameter values where $K(\theta,\sigma_t)$ is close to its minimized value, $K^*(\sigma_t)$. This statement is seen most easily for the case where Θ has only two

[13] See also Bunke and Milhaud (1998).

elements, θ_1 and θ_2. In this case, (3) becomes

$$\mu_{t+1}(\theta_1) \approx 1/(1+\frac{\mu_0(\theta_2)e^{-tK(\theta_2,\sigma_t)}}{\mu_0(\theta_1)e^{-tK(\theta_1,\sigma_t)}}). \tag{4}$$

Suppose, for example, that $(\sigma_t)_t$ converges to σ and that KLD is minimized at θ_1 given σ. Then there exists $\varepsilon > 0$ such that, for all sufficiently large t, $K(\theta_2,\sigma_t) - K(\theta_1,\sigma_t) > \varepsilon$. It follows from (4) that $\mu_{t+1}(\theta_1)$ converges to 1, so the posterior concentrates on the parameter value that minimizes KLD given σ. When $(\sigma_t)_t$ does not converge, however, we have to account for the possibility that $K(\theta_2,\sigma_t) - K(\theta_1,\sigma_t) > 0$ for all t but $K(\theta_2,\sigma_t) - K(\theta_1,\sigma_t) \to 0$ as $t \to 0$. In this case, we cannot say that the posterior eventually puts probability 1 on θ_1, even though θ_1 always minimizes KLD. This is why the next result says that the posterior concentrates on parameter values where $K(\theta,\sigma_t)$ is close to its minimized value, $K^*(\sigma_t)$, as opposed to saying that the posterior asymptotically concentrates on the minimizers of KLD given σ_t.[14] We now state the result formally and provide a proof.

Theorem 1. *Under Assumptions 1-3, for any policy f,*

$$\lim_{t\to\infty} \int_{\Theta} (K(\theta,\sigma_t) - K^*(\sigma_t))\mu_{t+1}(d\theta) = 0 \quad \mathbf{P}^f\text{-a.s.} \tag{5}$$

Proof. Fix a history h such that the condition of uniform convergence in Lemma 2 holds, and note that the set of histories with this property has probability one (henceforth, we omit the history from the notation). In particular, for all $\eta > 0$, there exists t_η such that, for all $t \geq t_\eta$,

$$|L_t(\theta) - K(\theta,\sigma_t)| < \eta \tag{6}$$

for all $\theta \in \Theta$.

Let $\bar{K}(\theta,\sigma) \equiv K(\theta,\sigma) - K^*(\sigma)$. Fix any $\varepsilon > 0$. We can use (2) and the fact that $0 \leq$

[14] Formally, what we are saying is that it is not generally true that $\lim_{t\to\infty} \int_\Theta \inf_{\theta' \in \Theta(\sigma_t)} \|\theta - \theta'\| \, \mu_{t+1}(d\theta) = 0$. This type of statement is true in Berk's iid setup and, as the previous discussion suggests, it is also true in our environment under the additional assumption that $(\sigma_t)_t$ converges (see below for Corollary 1).

$K^*(\sigma) < \infty$ for all σ (Lemma 1) to obtain

$$\begin{aligned}\int \bar{K}(\theta,\sigma_t)\mu_{t+1}(d\theta) &= \frac{\int_\Theta \bar{K}(\theta,\sigma_t)e^{-tL_t(\theta)}\mu_0(d\theta)}{\int_\Theta e^{-tL_t(\theta)}\mu_0(d\theta)} \\ &= \frac{\int_\Theta \bar{K}(\theta,\sigma_t)e^{-t(L_t(\theta)-K^*(\sigma_t))}\mu_0(d\theta)}{\int_\Theta e^{-t(L_t(\theta)-K^*(\sigma_t))}\mu_0(d\theta)} \\ &\le \varepsilon + \frac{\int_{\{\theta:\bar{K}(\theta,\sigma_t)\ge\varepsilon\}} \bar{K}(\theta,\sigma_t)e^{-t(L_t(\theta)-K^*(\sigma_t))}\mu_0(d\theta)}{\int_{\{\theta:\bar{K}(\theta,\sigma_t)\le\varepsilon/2\}} e^{-t(L_t(\theta)-K^*(\sigma_t))}\mu_0(d\theta)} \\ &=: \varepsilon + \frac{A_t^\varepsilon}{B_t^\varepsilon}.\end{aligned}$$

The proof concludes by showing that $\lim_{t\to\infty} A_t^\varepsilon / B_t^\varepsilon = 0$.

By (6), there exists t_η such that, for all $t \ge t_\eta$,

$$\begin{aligned}\frac{A_t^\varepsilon}{B_t^\varepsilon} &\le \frac{\int_{\{\theta:\bar{K}(\theta,\sigma_t)\ge\varepsilon\}} \bar{K}(\theta,\sigma_t)e^{-t(\bar{K}(\theta,\sigma_t)-\eta)}\mu_0(d\theta)}{\int_{\{\theta:\bar{K}(\theta,\sigma_t)\le\varepsilon/2\}} e^{-t(\bar{K}(\theta,\sigma_t)+\eta)}\mu_0(d\theta)} \\ &= e^{2t\eta}\frac{\int_{\{\theta:\bar{K}(\theta,\sigma_t)\ge\varepsilon\}} \bar{K}(\theta,\sigma_t)e^{-t\bar{K}(\theta,\sigma_t)}\mu_0(d\theta)}{\int_{\{\theta:\bar{K}(\theta,\sigma_t)\le\varepsilon/2\}} e^{-t\bar{K}(\theta,\sigma_t)}\mu_0(d\theta)}.\end{aligned}$$

Observe that the function $x \mapsto x\exp\{-tx\}$ is decreasing for all $x > 1/t$. Thus, for any $t \ge \max\{t_\eta, 1/\varepsilon\}$ it follows that $\bar{K}(\theta,\sigma_t)e^{-t\bar{K}(\theta,\sigma_t)} \le \varepsilon e^{-t\varepsilon}$ over $\{\theta : \bar{K}(\theta,\sigma_t) \ge \varepsilon\}$. Thus for all $t \ge \max\{t_\eta, 1/\varepsilon\}$,

$$\frac{A_t^\varepsilon}{B_t^\varepsilon} \le e^{t2\eta}\frac{e^{-t\varepsilon/2}}{\mu_0\left(\{\theta : \bar{K}(\theta,\sigma_t) \le \varepsilon/2\}\right)}. \tag{7}$$

In Appendix A.3, we show that continuity of $\bar{K}$ and compactness of $\Delta(\mathbb{X})$ imply that

$$\kappa_\varepsilon \equiv \inf_{\sigma\in\Delta(\mathbb{X})} \mu_0\left(\{\theta : \bar{K}(\theta,\sigma) \le \varepsilon/2\}\right) > 0 \tag{8}$$

for all $\varepsilon > 0$. Thus, setting $\eta = \varepsilon/8 > 0$, (7) implies that, for all $t \ge \max\{t_\eta, 1/\varepsilon\}$,

$$\frac{A_t^\varepsilon}{B_t^\varepsilon} \le \frac{e^{-t\varepsilon/4}}{\kappa_\varepsilon},$$

which goes to zero as $t \to \infty$. □

In Section 5, we use Theorem 1 to approximate the agent's belief, μ_t, with the set of probability measures with support in $\{\theta \in \Theta : K(\theta, \sigma_t) - K^*(\sigma_t) \leq \delta_t\}$, where $\delta_t \to 0$. Therefore, we will be able to study the asymptotic behavior of $(\sigma_t)_t$ via a stochastic difference equation that only depends on σ_t and an approximation error, and not on μ_t.

We can also use Theorem 1 to obtain the following corollary, which says that Berk's conclusion extends to the case where the frequency of actions converges. We will use this corollary in the next section to characterize the set of steady states.

Corollary 1. *Under Assumptions 1-3, for any policy f, if $(\sigma_t)_t$ converges to σ, then* $\lim_{k\to\infty} \mu_t(\Theta(\sigma)) = 1$ $\mathbf{P}^f$*-a.s.*

Proof. See Appendix A.4. □

4 Analysis of steady states

In this section, we characterize limiting behavior under the assumption that the sequence of frequencies of actions, $(\sigma_t)_t$, converges to some probability distribution $\sigma \in \Delta(\mathbb{X})$. This type of inquiry has a long tradition in economics, where the focus is often on characterizing steady-state behavior (often referred to as equilibrium behavior), abstracting away from the dynamics that may (or not) lead to such behavior.

The following theorem characterizes steady-state behavior.

Theorem 2. *Suppose that Assumptions 1-3 hold and let F be an uhc policy correspondence. For any policy $f \in Sel(F)$, the following holds P^f-a.s.: If $(\sigma_t)_t$ converges to σ, then*

$$\sigma \in \Gamma_F(\sigma), \tag{9}$$

where $\Gamma_F(\sigma) \equiv \Delta\left(\cup_{\mu \in \Delta(\Theta(\sigma))} F(\mu)\right)$.

For concreteness, we refer to any distribution that satisfies expression (9) as an **equilibrium** for policy correspondence F. It is convenient to express the fixed-point property in (9) in the following equivalent way (the proof is straightforward; hence, omitted).

Proposition 1. σ *is an equilibrium if and only if for all x such that $\sigma(x) > 0$, there exists $\mu_x \in \Delta(\Theta(\sigma))$ such that $x \in F(\mu_x)$.*

Proof of Theorem 2. We omit the history h in the notation. Suppose that x is such that $\sigma(x) > 0$. Since σ_t converges to σ, then x occurs infinitely often along the history, and so we can find a subsequence along which x occurs along the entire subsequence: $x_{t(j)} = x$ for all j. By compactness of $\Delta(\Theta)$, we can take a further subsequence $t(j(k))$ such that $\mu_x = lim_{k\to\infty}\mu_{t(j(k))}$ exists. Corollary 1 and the assumption that $(\sigma_t)_t$ converges to σ imply that $\mu_x \in \Delta(\Theta(\sigma))$. By definition of policy f, the action $x = f(\mu_{t(j(k))}) \in F(\mu_{t(j(k))})$ for all k. Finally, the assumption that F is uhc and the fact that $\mu_{t(j(k))} \to \mu_x$ imply that $x \in F(\mu_x)$. Therefore, we have established the claim that, for all x such that $\sigma(x) > 0$, there exists $\mu_x \in \Delta(\Theta(\sigma))$ such that $x \in F(\mu_x)$. By Proposition 1, this claim implies $\sigma \in \Delta\left(\cup_{\mu\in\Delta(\Theta(\sigma))}F(\mu)\right)$.
□

By showing that $\sigma \mapsto \Gamma_F(\sigma)$ satisfies the conditions of Kakutani's fixed-point theorem, we can establish that equilibrium always exists for any uhc policy correspondence F.

Proposition 2. *Under Assumptions 1-2, an equilibrium exists for any uhc policy correspondence F.*

Proof. See the Appendix A.5. □

One feature of Theorem 2 that may appear surprising is that it allows each action in the support of an equilibrium to be justified by a different belief. This feature is absent from previous analyses of equilibrium in the types of environments that we study.[15] Adopting Fudenberg and Levine's (1993) terminology from a different context, we call σ a **unitary equilibrium** for policy correspondence F if there exists $\mu \in \Delta(\Theta(\sigma))$ such that, for all x such that $\sigma(x) > 0$, $x \in F(\mu)$. A unitary equilibrium is therefore an equilibrium where every action in the support is justified by the same belief.

The next example shows why it is important to allow for non-unitary equilibria.

Example 1. Setup. The agent faces a matching problem. There are two states, A and B, two actions, x_A and x_B, and the objective of the agent is to match the action with the state. The payoff is 1 from a successful match (either x_A in state A or x_B in state B) and zero from an unsuccessful match, and the agent does not know which state will realize before making a choice. There is also safe action, x_S, that yields a payoff of .55 irrespective of the state. Importantly, the realized state is observed after the agent chooses x_A or x_B, but it is

[15] As formalized by Fudenberg and Levine (1993), this feature may be present in a type of environment, different from the one in this paper, where there is a population of agents in the role of one player, and different agents may have different experiences (hence, beliefs) about other players.

not observed if the agent chooses x_S. Thus, the set of actions is $\mathbb{X} = \{x_A, x_B, x_S\}$ and the set of (observable) consequences is $\mathbb{Y} = \{A, B, \diamondsuit\}$, where $\diamondsuit$ stands for the case where the agent chooses x_S and does not observe the state; i.e., $Q(\diamondsuit \mid x_S) = 1$. Also, the payoffs that represent this problem are $\pi(x_A, A) = \pi(x_B, B) = 1$, $\pi(x_A, B) = \pi(x_B, A) = 0$, and $\pi(x_S, \diamondsuit) = .55$.[16]

The agent knows the structure for the problem, except that she does not know the probability distribution over the states $\{A, B\}$. The agent believes that she has no control over the state, and so her model is given by $Q_\theta(B \mid x_A) = Q_\theta(B \mid x_B) = \theta$. We assume that $\theta \in \Theta = \{1/4, 3/4\}$, so that the agent believes that the probability of $y = B$ is either $1/4$ or $3/4$. We will later consider the case with more than two models of the world. Let $\mu \in [0,1]$ denote the agent's subjective probability that $\theta = 3/4$.

For simplicity, we assume that the agent follows a policy that is myopically optimal given her model of the world. In his case, the agent's policy is $F_0(\mu) = \{x_A\}$ if $\mu < .4$, $F_0(\mu) = \{x_S\}$ if $\mu \in (.4, .6)$, $F_0(\mu) = \{x_B\}$ if $\mu > .6$, with the agent being indifferent between $\{x_A, x_S\}$ at $\mu = .4$ and between $\{x_S, x_B\}$ at $\mu = .6$. The example easily extends to the case of a more patient agent.

The reality is a bit more complicated, because, in fact, the agent does affect the state with her action. In particular, we assume that $Q(A \mid x_B) = Q(B \mid x_A) = 1$, meaning that action x_B leads to $y = A$ for sure and action x_A leads to $y = B$ for sure. If the agent were aware of these facts, she would obviously prefer to choose x_S.

Equilibrium. We use Theorem 2 to solve for the set of equilibria. The first step is to find the set of closest parameter values $\Theta(\sigma)$ for all $\sigma = (\sigma(x_A), \sigma(x_B), \sigma(x_S))$. The KLD function is

$$K(\theta, \sigma) = \sigma(x_A) \ln \frac{1}{\theta} + \sigma(x_B) \ln \frac{1}{1-\theta} + \sigma(x_S) \ln \frac{1}{1}. \tag{10}$$

Since $K(1/4, \sigma) < (=) > K(3/4, \sigma)$ for $\sigma(x_A) < (=) > \sigma(x_B)$, it follows that

$$\Theta(\sigma) = \begin{cases} \{1/4\} & \text{if } \sigma(x_A) < \sigma(x_B) \\ \{1/4, 3/4\} & \text{if } \sigma(x_A) = \sigma(x_B) \\ \{3/4\} & \text{if } \sigma(x_A) > \sigma(x_B) \end{cases}.$$

We now apply the characterization of equilibrium from Proposition 1. Consider first the case $\sigma(x_A) < \sigma(x_B)$. If this case occurred in equilibrium, then Proposition 1 says that, because x_B is in the support of σ, there must exist $\mu_{x_B} \in \Delta(\Theta(\sigma))$ such that $x_B \in F_0(\mu_{x_B})$. In this

[16] The payoffs $\pi(x_A, \diamondsuit)$, $\pi(x_B, \diamondsuit)$, $\pi(x_S, A)$, $\pi(x_S, B)$ do not need to be specified because we will shortly assume that the agent believes that these action-consequence pairs have zero probability.

case, $\Theta(\sigma) = \{1/4\}$, and, therefore $\mu = 0$ is the only belief in $\Delta(\Theta(\sigma))$; recall that μ is the probability that $\theta = 3/4$. Under the belief $\mu = 0$, however, $x_B \notin F_0(0) = \{x_A\}$, and so we cannot have $\sigma(x_A) < \sigma(x_B)$ in equilibrium. A similar argument shows we cannot have $\sigma(x_A) > \sigma(x_B)$ in equilibrium, since then $\Theta(\sigma) = \{3/4\}$, $\mu = 1$, but $x_A \notin F_0(1) = \{x_B\}$.

Finally, consider the case $\sigma(x_A) = \sigma(x_B)$. In this case, $\Theta(\sigma) = \{1/4, 3/4\}$, and so any belief $\mu \in [0,1]$ is possible. In particular, there exists a belief (any in the interval $[.4,.6]$) that makes x_S optimal, and so $\sigma(x_S) = 1$ is an equilibrium. This equilibrium corresponds to the case where the agent settles down for the safe alternative and stops (or never begins) experimentation with the risky actions. If we restrict attention to unitary equilibria, then this is the only one, because there is no belief that makes the agent indifferent between x_A and x_B, and, therefore, we cannot have $\sigma(x_A) = \sigma(x_B) > 0$ in a unitary equilibrium. But there is in fact a continuum of non-unitary equilibria with $\sigma(x_A) = \sigma(x_B) \in (0,1)$: In particular, we can choose $\mu_{x_A} \leq .4$, $\mu_{x_B} \geq .6$, and $.4 \leq \mu_{x_S} \leq .6$ to support actions x_A, x_B, and x_S, respectively.

Dynamics. Let ρ be the ratio of priors for $\theta = 3/4$ vs. $\theta = 1/4$. Let R_t denote the ratio of posteriors at time t. Bayes' rule implies that $R_{t+1} = 3R_t$ if $y_t = 1$, $R_{t+1} = (1/3)R_t$ if $y_t = 0$, and $R_{t+1} = R_t$ if $y_t = \diamondsuit$. Also, using the fact that $\mu_t = R_t/(1+R_t)$, it follows that the agent chooses x_A if $R_t < 2/3$, x_S if $R_t \in (2/3, 3/2)$, and x_B if $R_t > 3/2$.

If the prior satisfies the case where $2/3 < \rho < 3/2$, the agent chooses x_S in the first period and forever after. Suppose instead that $\rho < 2/3$ (we omit the analogous case $\rho > 3/2$). The agent then chooses x_A in the first period. Because feedback is deterministic, the agent observes $y_1 = 1$ for sure. Therefore, the ratio increases from $R_0 = \rho$ to $R_1 = 3R_0$. If R_1 remains below $2/3$, the situation repeats itself. The process continues until the ratio of posteriors surpasses $2/3$. There are two possibilities: Either the ratio falls in the interval $(2/3, 3/2)$ and the agent chooses x_S forever after, or the ratio jumps above $3/2$. The first case occurs if the ratio of priors, ρ, is such that $\rho 3^{\hat{t}} \in (2/3, 3/2)$ for some $\hat{t}$ and the second case occurs if no such $\hat{t}$ exists. Clearly, there are values of ρ such that the second case occurs.[17] In this second case, the ratio of posteriors jumps above $3/2$. Subsequently, the agent takes action x_B, observes $y = 0$, and the ratio of posteriors goes back to the level of the previous period, before the jump, where it is below $2/3$ and the agent chooses x_A. The situation repeats itself indefinitely, and so the actions cycle deterministically and the frequency of actions converges to $\sigma(x_A) = \sigma(x_B) = 1/2$.

By the previous discussion, the frequency of actions either converges to $\sigma(x_S) = 1$ or it converges to $\sigma(x_A) = \sigma(x_B) = 1/2$; moreover, we can tell which is the case from knowledge of the ratio of priors, ρ. The latter of these two equilibria, however, is not unitary. Thus, in

[17] To see this claim, note that if the ratio of posteriors is at the boundary $2/3$ and the agent chooses x_0, then the ratio would jump to $3(2/3) = 2 > 3/2$.

this example, it is inappropriate to restrict attention to unitary equilibria if one cares about characterizing steady states.[18] □

4.1 Optimal policy and comparison to previous results

Theorem 2 differs from the typical characterization of equilibrium or steady state in the literature in three aspects. The first difference is technically trivial but expands the range of applications in a way that is currently unexplored: We do not necessarily assume that the agent follows an optimal policy, thus allowing for heuristics or other type of suboptimal behavior. Typically, however, equilibrium concepts require optimality, and so in this section we study the special case of our environment where the agent is assumed to follow an optimal policy (defined in Section 2 and denoted by F_β, where β is the discount factor). Theorem 2 is applicable to this special case because F_β is uhc.

Even after restricting Theorem 2 to apply to optimal policies, there remain two important differences with other equilibrium concepts. The first difference is that, in contrast to the definitions of Nash and Berk-Nash equilibrium, non-unitary equilibria are not ruled out by Theorem 2. The reason is that previous work has focused on environments where the unitary restriction is indeed appropriate.

There are two known situations where the unitary restriction is appropriate. The first case is one where the action itself converges. In this case, the equilibrium distribution is degenerate and, trivially, equilibrium is unitary. In other words, the unitary restriction only has bite for mixed-strategy equilibria. As pointed out by Fudenberg and Kreps (1993) in the context of Nash equilibria, play is unlikely to converge to mixed strategy equilibria in the standard model, but such convergence is reasonable when the agents' payoffs are randomly perturbed. With perturbations, one can study the sequence of distributions over actions, and the limit of this sequence may be interpreted as a mixed equilibrium; this approach is also followed by EP2016 to justify mixed-strategy Berk-Nash equilibrium. The second case where the unitary restriction is appropriate is one where payoffs are randomly perturbed and the resulting distribution over actions converges to a mixed (i.e., nondegenerate) distribution. The reason is that, far enough into the future, every action in the support of the convergent distribution is played

[18] Of course, the agent is being a bit too silly in this example. Because feedback is deterministic, she always keeps getting bad outcomes, but perpetually expects to get good outcomes in the future, and, in addition, her behavior cycles in a deterministic manner. We chose this extreme example with deterministic feedback to make our point because it is straightforward to solve for the dynamics. The example generalizes to random realizations of y, in which the cyclical behavior of the agent also becomes random. See Chapter 8 in Fudenberg and Levine (1998) for further discussion about an agent's ability to detect (or not) random or deterministic cycles in behavior.

with positive probability at *each point in time*. Thus, there is a common belief that justifies all these actions at each of these points in time.

In our environment, we allow for the possibility that the actions (or distributions over actions if we were to include payoff perturbations) do not converge, and so we must allow for non-unitary equilibria. The convenience and ubiquity of the unitary restriction, however, raises the question of finding general conditions under which all equilibria are unitary in our environment.

The other important difference is that, in order to apply Theorem 2, one needs to find the optimal correspondence F_β. When the agent is nonmyopic, i.e., $\beta > 0$, finding this correspondence requires solving a dynamic optimization problem that is often nontrivial. In contrast, equilibrium concepts in the literature, such as Nash or Berk-Nash, define optimality with respect to a myopic problem; in our case, this would be the correspondence F_0. The property of myopic optimality is very convenient because it entails solving a simpler optimization problem. This issue raises the question of the existence of some general condition (other than simply assuming that the agent is myopic, as is often done in the literature) under which we can replace the optimality condition with the simpler condition of myopic optimality.

Our next result speaks to both of the previous issues. We establish that, under a specific identification condition, equilibria corresponding to the optimal policy are unitary and can also be conveniently characterized with respect to myopic optimality.

Definition 2. The subjective model is **weakly identified given** $\sigma \in \Delta(\mathbb{X})$ if $\theta, \theta' \in \Theta(\sigma)$ implies that $Q_\theta(\cdot \mid x) = Q_{\theta'}(\cdot \mid x)$ for all x such that $\sigma(x) > 0$.

The definition of weak identification was introduced by EP2016. It says that the belief is uniquely determined along the equilibrium path, but leaves open the possibility of multiple beliefs for actions that are not in the support of σ. Weak identification is immediately satisfied if the agent's model is correctly specified, but it is also satisfied in many of the applications of misspecified learning in the literature; see EP2016 for further discussion.

Proposition 3. *Let F_β be the optimal policy correspondence with discount factor $\beta \in [0,1)$ and let σ be an equilibrium for policy correspondence F_β, i.e., $\sigma \in \Delta\left(\cup_{\mu \in \Delta(\Theta(\sigma))} F_\beta(\mu)\right)$. If the subjective model is weakly identified given σ, then*

$$\sigma \in \cup_{\mu \in \Delta(\Theta(\sigma))} \Delta(F_0(\mu)), \tag{11}$$

i.e., there exists $\mu \in \Delta(\Theta(\sigma))$ such that, for all x such that $\sigma(x) > 0$, $x \in F_0(\mu)$.[19]

Proof. See the Appendix A.6. □

EP2016 call a point satisfying equation (11) a Berk-Nash equilibrium. They show that, if the sequence of distributions over actions converges, then it converges to a Berk-Nash equilibrium. In our environment there is no motive for mixing, so convergence of the sequences of distributions over actions implies that the sequence of actions must converge. Theorem 2 and Proposition 3 strengthen their conclusion by showing that, under weak identification, even though the sequence of actions may not converge, if the sequence of *frequencies* of actions converges, then it converges to a Berk-Nash equilibrium.

The intuition behind Proposition 3 is as follows. First, weak identification implies that actions in the support of a limit frequency σ do not provide any new information about the parameter space. By the convexity of the value function and the martingale property of Bayesian beliefs, the value of experimenting with an off-equilibrium action is nonnegative. Therefore, if an action x in the support of σ is better than an action x' in the problem where $\beta > 0$ (so that the value of experimentation provided by x' is taken into account), then x must be better than x' in the myopic problem where $\beta = 0$ (so that the potential informational value of x' is not taken into account).[20]

Second, because optimization reduces to myopic optimization, the distribution over consequences is sufficient to evaluate the optimality of an action. By weak identification, for every action in the support of an equilibrium σ, the distribution over consequences conditional on that action is the same for all beliefs in $\Delta(\Theta(\sigma))$. Therefore, we can take a single belief to justify all actions in the support of σ.

Example 1, modified. The subjective model in this example is not weakly identified given σ satisfying $\sigma(x_A) = \sigma(x_B)$, since in this case $\Theta(\sigma) = \{1/4, 3/4\}$, and these two parameter values imply different distributions over consequences: $Q_{1/4}(1 \mid x) = 1/4 \neq 3/4 = Q_{3/4}(1 \mid x)$ for $x \in \{x_A, x_B\}$.

Consider next a modified example, where $\Theta = [1/4, 3/4]$ now includes all parameter values in between $1/4$ and $3/4$. The KLD function is the same as before (see equation (10)), but its

[19] Recall that, by definition, $F_0(\mu) = \arg\max_{x \in \mathbb{X}} E_{\bar{Q}_\mu(\cdot|x)}[\pi(x, Y)]$.

[20] This argument was previously made in the online appendix of EP2016 for the case of nonmyopic agents.

minimization now yields

$$\Theta(\sigma)=\begin{cases}\{1/4\} & \text{if } \sigma(x_A)/(\sigma(x_A)+\sigma(x_B))\leq 1/4\\ \{\sigma(x_A)/(\sigma(x_A)+\sigma(x_B))\} & \text{if } \sigma(x_A)/(\sigma(x_A)+\sigma(x_B))\in(1/4,3/4)\\ \{3/4\} & \text{if } \sigma(x_A)/(\sigma(x_A)+\sigma(x_B))\geq 3/4\end{cases}$$

provided that $\sigma(x_A)+\sigma(x_B)>0$, and $\Theta(\sigma)=\Theta$ otherwise. In particular, the subjective model is now weakly identified: There is a singleton minimizer in all cases except $\sigma(x_S)=1$, and, in that case, all parameter values in Θ yield the same distribution over consequences given x_S, mainly the consequence that nothing is observed. Theorem 2 and Proposition 3 imply that we can restrict attention to unitary equilibria.

Consider first the case $\sigma(x_A)+\sigma(x_B)>0$. As shown above, the belief in this case is degenerate, and we denote by δ_θ the Dirac measure on θ. Optimality implies that

$$F_0(\delta_\theta)=\begin{cases}\{x_A\} & \text{if } \theta<.45\\ \{x_S\} & \text{if } .45<\theta<.55.\\ \{x_B\} & \text{if } \theta>.55\end{cases}$$

We also have $F_0(\delta_\theta)=\{x_A,x_S\}$ if $\theta=.45$ and $F_0(\delta_\theta)=\{x_S,x_B\}$ if $\theta=.55$. In particular, there is no belief under which both x_A and x_B are optimal, and so the only possible equilibria are the two pure ones, $\sigma(x_A)=1$ and $\sigma(x_A)=1$. But we can rule out both of these cases: If $\sigma(x_A)=1$, then $\Theta(\sigma)=\{3/4\}$, but $x_A\notin F_0(\delta_{3/4})$; similarly, we can rule out $\sigma(x_B)=1$ in equilibrium.

Finally, consider the case $\sigma(x_S)=1$. In this case, $\Theta(\sigma)=\Theta$ and, in particular, there exists a belief with support in Θ (e.g., δ_θ, for any $.45\leq\theta\leq.55$) that makes x_S optimal. Thus, there is only one equilibrium and it satisfies $\sigma(x_S)=1$.

This analysis shows that, when the set of parameter values is enlarged to $\Theta=[1/3,3/4]$, there is a unique equilibrium, and it involves always choosing the safe action. In particular, there is no longer an equilibrium where actions x_A and x_B are taken. In addition, Proposition 3 implies that $\sigma(x_S)=1$ is also the unique equilibrium for *any* discount factor $\beta>0$, a statement that is not a priori obvious. □

5 Characterization of asymptotic outcomes

In this section, we propose a method to study the asymptotic behavior of the frequencies of actions, and not just steady states. Among other benefits, one can use the method to determine if behavior converges or not. The key departure from previous approaches in the literature is to focus on the evolution of frequencies of actions. Using the characterization of beliefs in Theorem 1, we write this evolution as a stochastic difference equation expressed exclusively in terms of the frequencies of actions. We then use tools from stochastic approximation developed by Benaïm, Hofbauer and Sorin (2005) (henceforth, BHS2015) to characterize the solutions of this difference equation in terms of the solution to a generalization of a differential equation. In Section 5.2, we apply our method to two common environments in the literature.

5.1 Characterization of asymptotic frequencies

We first provide a heuristic description of our approach. The sequence of frequencies of actions, $(\sigma_t)_t$, can be written recursively as follows:

$$\sigma_{t+1} = \sigma_t + \frac{1}{t+1}\left(\mathbf{1}(x_{t+1}) - \sigma_t\right), \tag{12}$$

where $\mathbf{1}(x_{t+1}) = (\mathbf{1}_x(x_{t+1}))_{x\in\mathbb{X}}$ and $\mathbf{1}_x(x_{t+1})$ is the indicator function that takes the value 1 if $x_{t+1} = x$ and 0 otherwise.

By adding and subtracting the conditional expectation of $\mathbf{1}(x_{t+1})$ (i.e., the probability that each action is played at time $t+1$ given the belief at time $t+1$), we obtain

$$\sigma_{t+1} = \sigma_t + \frac{1}{t+1}\left(E\left[\mathbf{1}(x_{t+1}) \mid \mu_{t+1}\right] - \sigma_t\right) + \frac{1}{t+1}\left(\mathbf{1}(x_{t+1}) - E\left[\mathbf{1}(x_{t+1}) \mid \mu_{t+1}\right]\right). \tag{13}$$

The last term in equation (13) is a Martingale difference sequence and essentially adds a noise term to the equation that can be controlled asymptotically. This is true for the general case where we allow the agent to play a mixed action for a given belief. In our case, we have simplified matters and do not allow the agent to randomize among actions, and so this third term is exactly equal to zero.

The reason it is hard to characterize $(\sigma_t)_t$ using (13) is that its evolution depends on the agent's belief. If we could somehow write the belief μ_{t+1} as a function of σ_t, then we would have a recursion where σ_{t+1} depends only on σ_t. This is where Theorem 1 from Section 3 is useful. This theorem will allow us to approximate μ_{t+1} with a set of probability measures that depends on σ_t.

The ultimate objective is not really to approximate μ_{t+1} but rather the conditional expectation $E[\mathbf{1}(x_{t+1}) \mid \mu_{t+1}]$ in equation (13). The conditional expectation, however, is typically discontinuous in the belief (this is particularly so for a belief under which the agent is indifferent between two actions). Thus, replacing μ_{t+1} with a good approximation does not necessarily yield a good approximation for the conditional expectation. There are two ways to tackle this discontinuity issue. One way is to replace $\mu \mapsto E[\mathbf{1}(x_{t+1}) \mid \mu]$ with a smooth, continuous version, that is very close to the original conditional expectation. This approach comes at the cost of having to approximate the agent's behavior.[21] An alternative approach, that we follow in this paper, is to replace the function $\mu \mapsto E[\mathbf{1}(x_{t+1}) \mid \mu]$ with a correspondence that contains this function and is well behaved.

To see how this approach works, note that $E[\mathbf{1}(x_{t+1}) \mid \mu] \in \Delta(F(\mu))$ for all μ. Therefore, we can view equation (13) as a particular case of the following stochastic difference *inclusion*:

$$\sigma_{t+1} = \sigma_t + \frac{1}{t+1}(r_{t+1} - \sigma_t) + \frac{1}{t+1}\xi_{t+1}, \tag{14}$$

where $r_{t+1} \in \Delta(F(\mu_{t+1}))$ and $E_{r_{t+1}}[\xi_{t+1}] = 0$. It is called a difference inclusion because r_{t+1} can take multiple values.[22] Importantly, we use Theorem 1 to approximate μ_{t+1} with the set of probability measures μ satisfying $\int_\Theta (K(\theta, \sigma_t) - K^*(\sigma_t))\mu(d\theta) \leq \delta_t$, where $\delta_t \to 0$ is a vanishing approximation error. In particular, if the error were exactly zero, the set would be equal to $\Delta(\Theta(\sigma_t))$. More generally, the difference equation (14) can be written entirely in terms of $(\sigma_t)_t$ and approximation errors.

A key insight from the theory of stochastic approximation is that, in order to characterize a discrete-time process such as $(\sigma_t)_t$, it is is convenient to work with its continuous-time interpolation. Because of the multiplicity inherent in equation (14), we apply the specific methods developed by BHS2015, who extend Benaim's (1996) ordinary-differential equation method to the case of differential inclusions.[23]

Set $\tau_0 = 0$ and $\tau_t = \sum_{i=1}^{t} 1/i$ for $t \geq 1$. The continuous-time interpolation of $(\sigma_t)_t$ is the

[21] The arguments below can be easily modified to handle this continuous case. The relevant object in the analog of Theorem 3 would be a differential equation, and not a differential inclusion.

[22] As explained earlier, in the case where the agent is not allowed to mix, $\xi_{t+1} = 0$ for all t.

[23] See Borkar (2009) for a textbook treatment of the ordinary-differential equation method in stochastic approximation.

function $\mathbf{w} : \mathbb{R}_+ \to \Delta(\mathbb{X})$ defined as

$$\mathbf{w}(\tau_t + s) = \sigma_t + s\frac{\sigma_{t+1} - \sigma_t}{\tau_{t+1} - \tau_t}, \quad s \in [0, \frac{1}{t+1}). \tag{15}$$

Figure X illustrates this simple interpolation for a specific value of $x \in \mathbb{X}$. A convenient property of the interpolation is that it preserves the accumulation points of the discrete process.

Equations (14) and (14) can be combined to show that the derivate of $\mathbf{w}$ with respect to (a re-indexing of) time, which we denote by $\dot{\mathbf{w}}$, is approximately given by $r_{t+1} - \sigma_t$. As argued earlier, r_{t+1} belongs to a set that depends on σ_t and an approximation error, and this set is equal to $\Delta(F(\Theta(\sigma_t)))$. Thus, the derivate is roughly equal to $\Gamma_F(\sigma_t) \equiv \Delta(F(\Theta(\sigma_t))) - \sigma_t$. The next step is to replace σ_t in this last expression by its interpolation $\mathbf{w}(t)$. This replacement adds yet another vanishing approximation error, and we therefore obtain, ignoring the approximation error, that $\dot{\mathbf{w}}(t) \in \Delta(F(\Theta(\mathbf{w}(t)))) - \mathbf{w}(t)$. Thus, we can show that the continuous-time interpolation of $(\sigma_t)_t$ is well approximated by solutions of the following differential inclusion:

$$\dot{\boldsymbol{\sigma}}(t) \in \Gamma_F(\boldsymbol{\sigma}(t)) - \boldsymbol{\sigma}(t). \tag{16}$$

Formally, a solution to the differential inclusion (16) over an interval $[0,T]$ with initial point $\sigma \in \Delta(\mathbb{X})$ is an absolutely continuous mapping $\boldsymbol{\sigma} : [0,T] \to \Delta(\mathbb{X})$ such that $\boldsymbol{\sigma}(0) = \sigma$ and (16) is satisfied for almost every $t \in [0,T]$. Let S^T_σ denote the set of solutions to (16) over $[0,T]$ with initial point σ. The assumption that F is uhc implies that, for every initial point, there exists a (possibly nonunique) solution to (16); see, e.g., Aubin and Cellina (2012).

We now state the main result.

Theorem 3. *Suppose that Assumptions 1-3 hold and let F be an uhc policy correspondence. For any policy $f \in Sel(F)$, the following holds P^f-a.s.: For all $T > 0$,*

$$\lim_{t\to\infty} \inf_{\boldsymbol{\sigma} \in S^T_{\mathbf{w}(t)}} \sup_{0 \le s \le T} \|\mathbf{w}(t+s) - \boldsymbol{\sigma}(s)\| = 0. \tag{17}$$

Proof. See Appendix A.7. □

Theorem 3 says that, for any $T > 0$, the curve $\mathbf{w}(t + \cdot) : [0,T] \to \Delta(\mathbb{X})$ defined by the continuous-time interpolation of $(\sigma_t)_t$ approximates some solution to the differential inclusion (16) with initial condition $\mathbf{w}(t)$ over the interval $[0,T]$ with arbitrary accuracy for sufficiently large t. This result is convenient because it allows us to characterize asymptotic properties of $(\sigma_t)_t$ by solving the differential inclusion in (16).

BHS2005 refer to a function $\mathbf{w}$ satisfying (17) as an asymptotic pseudotrajectory of the differential inclusion. For many practical applications (including the examples we study in this paper), Theorem 3 implies that we can find the limit points of $(\sigma_t)_t$ by studying the limits of the trajectories that solve the differential inclusion (16). This is not true in all cases, as originally explained by Benaim (1996). Indeed, BHS2005 show that the limit set of a (bounded) asymptotic pseudotrajectory is internally chain transitive (this is an extension of Benaim's original result for the special case of differential equations). Thus, a corollary of Theorem 3 is that the frequency of actions converges almost surely to an internally chain transitive set of the differential inclusion. For many examples in economics, it is not necessary to go to this further characterization, and so we refer the reader to BHS2005 for this more refined characterization and the definition of an internally chain transitive set, as well as for additional implications of Theorem 3.

5.2 Applications

Theorem 3 can be applied to any model of misspecified learning that fits our environment. For concreteness, we now show how to apply this theorem to two classes of environments considered in the literature. It is convenient to divide these environments into what we will call environments with negative or positive reinforcement.

5.2.1 Negative reinforcement

In one class of problems, when the agent takes an action, then the agent's resulting belief (as computed by minimizing KLD conditional on the action taken) is such that the agent prefers to take a different action. We say that such an action is negatively reinforcing. Nyarko (1991) considers a monopolist who views one of two possible demand models to be true (i.e., Θ contains two elements), but neither of these models is the correct model. He shows that prices do not converge. Fudenberg, Romanyuk, and Strack (2017; henceforth FRS2017) study a related problem and give the following story: A seller thinks that she faces a constant-elasticity demand function, but does not know this elasticity. In reality, the elasticity is not constant but is high for low prices and low for high prices. A seller who chooses low prices will estimate a low elasticity. But then she will prefer to set a high price. Similarly, a seller who chooses high prices will estimate a high elasticity, but the she will prefer to set a low price. Prices are therefore negatively reinforcing.

Formally, let x_A and x_B be two actions (say, low and high prices), and, let $\sigma = (\sigma(x_A), \sigma(x_B))$ denote a probability distribution over actions. Suppose that both actions are negatively reinforcing: for any belief with support in $\Theta((1,0))$ the optimal action is x_B and for any belief with support in $\Theta((0,1))$ the optimal action is x_A. While Nyarko and FRS2017 restrict attention to the case where Θ has two elements, the following discussion holds for an arbitrary number of elements.

To apply Theorem 3, we first write down the differential inclusion (16) for this example. The first step is to find the mapping $\sigma \mapsto \Gamma_F(\sigma) \equiv \Delta(\cup_{\mu \in \Delta(\Theta(\sigma))} F(\mu))$. By the assumption that the actions are negatively reinforcing, it follows that never choosing x_B leads the agent to want to do x_B, i.e., $\Gamma_F((1,0)) = \delta_{x_B}$, and, similarly, $\Gamma_F((0,1)) = \delta_{x_A}$, where δ_x stands for the Dirac measure on action x.

Suppose that σ^* is the unique equilibrium, i.e., $\sigma^* \in \Gamma_F(\sigma^*)$. Figure X provides an example. The assumption of a unique equilibrium is satisfied in the examples studied in the literature, so we will make it here, but it would be straightforward to extend the discussion to the case of multiple equilibria.

The negatively reinforcing and unique equilibrium properties capture the above monopoly story as well as others.[24] If the agent chooses x_A, then her belief will be such that she will be less confident that x_A is best. Eventually, once the frequency of x_A is above $\sigma^*(x_A)$, she will believe that x_B is best and will start doing x_B. But this will bring the frequency of x_A back towards $\sigma^*(x_A)$. At the threshold σ^*, the agent has a belief that makes her indifferent between x_A and x_B.

According to equation (16) and Figure X, $\dot{\boldsymbol{\sigma}}(x_A) > 0$ for $\boldsymbol{\sigma}(x_0) < \sigma^*(x_0)$ and $\dot{\boldsymbol{\sigma}}(x_0) < 0$ for $\boldsymbol{\sigma}(x_0) > \sigma^*(x_0)$. At σ^*, the derivate can take multiple values, because $\cup_{\mu \in \Delta(\Theta(\sigma^*))} F(\mu) = \{x_0, x_1\}$ and so $\Gamma(\sigma^*) = \Delta(\{x_0, x_1\})$, but this multiplicity is irrelevant for determining the limit of solutions to the differential inclusion: For any initial condition, a solution converges to σ^*. Therefore, by Theorem 3, the underlying frequency of actions, $(\sigma_t)_t$, converges to σ^*, even though the action itself does not converge.

Ours is the first result in the literature that shows explicit convergence of the frequency of actions in a misspecified-learning problem. EP2016 (Online Appendix) considered Nyarko's example and obtained convergence to a mixed strategy by adding payoff perturbations and considering very specific assumptions (e.g., normal distribution, specific functional forms,

[24]Negative reinforcement is also present in some of the examples in Spiegler (2016) as well as in the voting environments of Esponda and Pouzo (2017, 2019) and Esponda and Vespa (2018) and in the investment environment of Jehiel (2018).

etc.). The current analysis does away with payoff perturbations and also holds under very general conditions (only negative reinforcement is needed; the uniqueness of the fixed point σ^* is used just for convenience in the exposition).

We now turn to an extension of Nyarko's model introduced by FRS2017. They make two assumptions that lend considerable tractability: Time is continuous, with the flow payoff being the sum of an unknown drift plus a Brownian option, and the agent considers the unknown drift to be one of two possible kinds, i.e., $\Theta = \{\theta_L, \theta_H\}$ has two elements. They show that the belief can be characterized as a solution to a single-dimensional stochastic differential equation, and they can rely on existing mathematical results regarding such equations. Single-dimensionality is important here, and it follows from the assumption that Θ only has two elements.

While FRS2017 consider all possible cases of their model, we will focus on the case where the agent continues to have two negatively reinforcing actions but now also has a third, uninformative action, that does not allow her to distinguish between the two possible parameter values in Θ. We denote this third action by x_S. We will, however, work in discrete time, and we will not make distributional assumptions.

With three actions, we abuse notation and let $\tilde{\sigma} = \sigma(x_0)/(\sigma(x_0) + \sigma(x_1))$ denote the frequency of x_0 relative to x_0 and x_1, excluding x_S. For simplicity, we assume that this relative frequency is a sufficient statistic for determining the minimizers of KLD and that there is a threshold $\tilde{\sigma}^*$ such that one of the parameter values, say θ_L, minimizes KLD for $\tilde{\sigma} < \tilde{\sigma}^*$ and the other parameter value, θ_H, minimizes KLD for $\tilde{\sigma} > \tilde{\sigma}^*$; at $\tilde{\sigma}^*$, both parameter values minimize KLD. Suppose also that x_0 is optimal under θ_L and x_1 is optimal under θ_H. Under these assumptions, each of the actions x_0 and x_1 are negatively reinforcing (for example, if x_1 is chosen with probability one, then θ_L minimizes KLD, but then x_0 is the optimal action).

Action x_S, on the other hand, is completely uninformative, and the willingness to choose this action depends on the agent's discount factor. Consider first the case where the agent is sufficiently patient in the sense that there is no belief under which she wants to choose the uninformative action, x_S. Figure X depicts the solutions to the differential inclusion corresponding to this case. Consider a point in the upper triangle, such as A. At this point, $\tilde{\sigma} < \tilde{\sigma}^*$, and the parameter value that minimizes KLD is θ_L. It is then optimal for the agent to choose x_0, as shown by the direction of the arrows, pointing towards the point $(1,0,0)$. The opposite is true for points in the lower triangle. Along the diagonal, where σ is such that $\tilde{\sigma} = \tilde{\sigma}^*$, $\cup_{\mu \in \Delta(\Theta(\sigma))} F(\mu) = \{x_0, x_1\}$, and so any convex combination between x_0 and x_1 is possible. No matter which combination we pick, all trajectories converge to the point $\sigma^* = (\tilde{\sigma}^*, 1 - \tilde{\sigma}^*, 0)$.

FRS2017 show that the action does not converge if the agent is sufficiently patient. Our analysis corroborates this finding, but also shows that the frequency of actions does converge to σ^*.

Consider next the case where the agent is not sufficiently patient, meaning that there are intermediate beliefs for which x_S is optimal. The dynamics starting from the upper or lower triangle remain the same as in Figure X. Also as before, any belief is possible along the diagonal. The difference is that now action x_S can be rationalized by one of these beliefs, and so $\cup_{\mu \in \Delta(\Theta(\sigma))} F(\mu) = \{x_0, x_1, x_S\}$ for any σ such that $\tilde{\sigma} = \tilde{\sigma}^*$. Thus, any convex combination between x_0, x_1, *and* x_S is possible along the diagonal. Figure X illustrates this new situation. The limit set of the trajectories of the differential inclusion is now the entire segment connecting the origin to the point $(\tilde{\sigma}^*, 1-\tilde{\sigma}^*, 0)$. In particular, it is possible that the frequency converges to $(0,0,1)$, or to $(\tilde{\sigma}^*, 1-\tilde{\sigma}^*, 0)$, or to any other point in between, or to even diverge (restricted to the diagonal). This example illustrates a limitation of our approach. Theorem 3 only says that the frequency eventually stays on the diagonal, but does not further specify where in the diagonal.

To see why a result as general as Theorem 3 could not possibly provide a full answer in this example, it helps to go back to Example 1 in Section 4, which is a special case of the example we are now analyzing. In Example 1, we found that we could partition the set of priors into two sets: For priors in one set, the frequency converges to $(0,0,1)$ and for priors in the other set, it converges to $(\tilde{\sigma}^*, 1-\tilde{\sigma}^*, 0)$, where $\tilde{\sigma}^* = 1/2$. A result such as Theorem 3 exploits properties of asymptotic beliefs and behavior, and, on its own, it can say nothing about how different priors can lead to different outcomes.

Interestingly, FRS2017 find that convergence is always to $(0,0,1)$ in their example. The reason is that their example contains an infinite number of consequences, and there are values for those consequences that result in an infinitesimal move of the posterior. Thus, the posterior eventually enters into the attracting region where it is optimal to choose x_S. But, more generally, in order to fully determine the asymptotic behavior, one would have to specifically study the prior and relate it to the informativeness of the consequences, as we did when solving Example 1.

Finally, let's continue to assume that the agent is impatient, so that there is a set of beliefs such that x_S is optimal, but let's now assume that Θ has more than two elements, a case which cannot be tackled with the techniques of FRS2017. In particular, suppose that we convexify the set of parameter values to be $\Theta = [\theta_L, \theta_H]$. A special case is our modified version of Example 1 in Section 4.1, where $\Theta = [1/4, 3/4]$. Figure X plots the trajectories of the corresponding differential inclusion. There are now three regions in the figure. The upper and lower regions are exactly as before: Frequencies leads to extreme beliefs which in turn lead to actions in the opposite direction. But, in the middle region, frequencies now lead to intermediate beliefs under which x_S is optimal. Along the boundaries of these two regions, there is multiplicity in actions (between x_0 and x_S in one case and x_1 and x_S in the other). But, as shown in Figure X, all trajectories converge to the origin, $(0,0,1)$. Thus, Theorem 3 implies that the action converges to the safe action, x_S. Interestingly, irrespective of the informativeness of consequences, we obtain FRS2017's prediction for a richer parameter space.

5.2.2 Positive reinforcement

A different kind of environment studied in the literature is one where actions are positively reinforcing. Esponda (2008) studies a class of adverse selection environments with two properties. The first property, which he calls the monotone selection property, says that lower actions result in a worse selection of outcomes. The second property postulates complementarity between beliefs and actions, so that a worse selection of outcomes encourages lower actions. An example is the acquire-the-company game, where a lower price offer by the buyer results in a worse selection of companies, which in turn leads the buyer to believe that the company is worth less and to consequently offer even lower prices. Esponda (2008) discusses additional applications, including insurance markets, auctions, and performance pay. He focuses, however, on defining and characterizing equilibrium, except for one simple acquire-the-company game where, under fairly restrictive parametric assumptions, he shows convergence to the (unique) equilibrium.

More recently, Heidhues, Kőszegi, and Strack (2018; henceforth, HKS2018) study an environment where the agent's performance is determined by his ability, his action (effort), and an unknown state of the world. They assume that the agent is overconfident about his own ability, leading to biased learning about the state of the world. Due to overconfidence, the agent becomes pessimistic about the state as he obtains feedback about his performance. This pessimism leads him to choose lower effort, which in turn makes him even more pessimistic, inducing him to choose even lower effort, and so on. In contrast to Esponda (2008), HKS2018 completely characterize asymptotic behavior under fairly general conditions. In particular, they show that the agent's action converges to the (unique) equilibrium. The conditions they assume include the uniqueness of the equilibrium as well as some additional parametric restrictions for the case where the agent is nonmyopic. They discuss applications to delegation, control in organizations, and public policy choices.

We will revisit these applications using the methods developed in this paper; as a side product, we will illustrate that in many instances it is possible to reduce the dimensionality of the differential inclusion in (16) to make the analysis more tractable. Suppose that $Y = h(x) + \varepsilon$ is the observed consequence, where ε is a random error term, x is the agent's action, and $h(\cdot)$ is a strictly positive function. The agent's model is $Y = \theta g(x) + \varepsilon$, where $\theta \in \Theta = [\min_x h(x)/g(x), \max_x h(x)/g(x)] \subset \mathbb{R}$ is a single-dimensional parameter and $g(\cdot)$ is a strictly positive function. In Esponda's (2008) acquire-the-company example, x is the price offer and Y is the value of the firm. The agent does not realize that the price affects the value, and so $g(\cdot)$ is a constant function. In HKS2018's overconfidence example, Y is production, and it depends on an unknown state θ and the agent's effort x, as mediated by a function g that may be the wrong function (exhibiting, say, overconfidence). We assume for simplicity that ε is normally distributed.[25] In this case, there is a unique parameter value that minimizes KLD and it is a convex combination of the parameter values that minimize KLD for the degenerate actions,

$$\theta^*(\sigma) = \sum_{x \in \mathbb{X}} \theta^*(\delta_x) \rho(x, \sigma),$$

where $\rho(x,\sigma) = g^2(x)\sigma(x) / \sum_{\tilde{x}} g^2(\tilde{x})\sigma(\tilde{x})$ and $\theta^*(\delta_x) = h(x)/g(x)$. For the special case where $g(\cdot)$ is constant, each action is weighted according to its frequency in the data, $\rho(x,\sigma) = \sigma(x)$. For the more general case, the weights simply reflect the fact that actions that yield lower variance receive higher weight.[26]

Taking the derivative of $\boldsymbol{\theta}^*(\boldsymbol{\sigma}(t))$ with respect to time and using (16), we obtain

$$\begin{aligned} \dot{\boldsymbol{\theta}}^*(\boldsymbol{\sigma}(t)) &= D\rho(\boldsymbol{\sigma}(t))\theta^*(\delta_\cdot) \cdot \dot{\boldsymbol{\sigma}}(t) \\ &\in D\rho(\boldsymbol{\sigma}(t))\theta^*(\delta_\cdot) \cdot \Delta(F(\delta_{\boldsymbol{\theta}^*(\boldsymbol{\sigma}(t))})), \end{aligned} \tag{18}$$

where $D\rho(\sigma) \equiv (\partial\rho(x,\sigma)/\partial\sigma(\tilde{x}))_{x,\tilde{x}}$ is an $|\mathbb{X}| \times |\mathbb{X}|$ matrix, $\theta^*_{\delta_\cdot} \equiv (\theta^*(\delta_x))_x$ is an $|\mathbb{X}| \times 1$ vector, and where the last line uses the fact that $D\rho(\sigma)\theta^*(\delta_\cdot) \cdot \sigma = \mathbf{0}$ for all σ, which follows from homogeneity of degree 0 of $\rho(x,\cdot)$.

Consider the case where the parameter value is such that there is a unique optimal action,

[25] The results can be extended to more general distributions and functional form assumptions.

[26] If we think of the observable as $Y/g(x)$, then the mean of this random variable is θ and the variance depends negatively on $g(x)$, so that actions with higher values of g should have higher weight.

say $\hat{x}$, i.e., $F(\delta_{\boldsymbol{\theta}^*(\boldsymbol{\sigma}(t))}) = \{\hat{x}\}$. Using (18) and some algebra, it follows that

$$\dot{\boldsymbol{\theta}}^*(\boldsymbol{\sigma}(t)) = \frac{(g(x^*))^2}{\sum_{x \in \mathbb{X}}(g(x))^2 \boldsymbol{\sigma}(t)} \left(\theta^*(\delta_{\hat{x}}) - \boldsymbol{\theta}^*(\boldsymbol{\sigma}(t))\right). \tag{19}$$

We can track the evolution of $\boldsymbol{\theta}^*(\sigma(t))$ using equation (19) in general, but, for concreteness, we now introduce two additional assumptions that are key features of the types of environments with positively reinforcing actions. To facilitate the definitions, suppose that there is a complete order on the action space: $x_1 < ... < x_{|\mathbb{X}|}$. First, we assume that the mapping $\theta \mapsto F(\delta_\theta)$ is increasing, in the following sense. There is an increasing function $\theta \mapsto x^*(\theta)$ and a set of at most $|\mathbb{X}| - 1$ thresholds, $\bar{\Theta}$, such that, for all $\theta \in \Theta \backslash \bar{\Theta}$, $F(\delta_\theta) = \{x^*(\theta)\}$; moreover, at each threshold $\bar{\theta} \in \bar{\Theta}$, the agent is indifferent between the two adjacent actions, $\lim_{\theta \uparrow \bar{\theta}} x^*(\theta)$ and $\lim_{\theta \downarrow \bar{\theta}} x^*(\theta)$. Figure X depicts an example of this standard property, which simply says that higher beliefs lead to higher actions. The second assumption is that higher actions in turn lead to higher beliefs: $x \mapsto \theta^*(\delta_x)$ is an increasing function.

With the two assumptions above and the help of (19), we can characterize the evolution of $\boldsymbol{\theta}^*(\boldsymbol{\sigma}(t))$ starting from any $\theta \in \Theta \backslash \bar{\Theta}$. Since the uniquely optimal action is $x^*(\theta)$, equation (19) implies that $\boldsymbol{\theta}^*(\boldsymbol{\sigma}(t))$ increases (decreases, stays the same) if $\theta^*(\delta_{x^*(\theta)})$ is higher (lower, equal) than θ. Suppose, for example, that $\theta^*(\delta_{x^*(\theta)}) > \theta$, so that $\boldsymbol{\theta}(\boldsymbol{\sigma}(t))$ increases. If the agent continued to choose $x^*(\theta)$ indefinitely, then the belief would converge to $\theta^*(\delta_{x^*(\theta)})$. If the agent were to choose a different action, because the belief is increasing, then the new action would be higher. So $\theta^*(\delta_{x^*(\theta)})$ is a lower bound to the agent's eventual belief. On the other hand, if $\theta^*(\delta_{x^*(\theta)}) < \theta$, then $\boldsymbol{\theta}(\boldsymbol{\sigma}(t))$ decreases and $\theta^*(\delta_{x^*(\theta)})$ is an upper bound to the agent's eventual belief. Thus, we can track the evolution of $\boldsymbol{\theta}(\boldsymbol{\sigma}(t))$ by studying the mapping $\theta \mapsto \theta^*(\delta_{x^*(\theta)})$, which, given our assumptions, is always increasing.[27]

Figure X shows an example of the mapping $\theta \mapsto \theta^*(\delta_{x^*(\theta)})$. There are three equilibrium beliefs in this example. Starting from any initial condition, we converge to either θ_1^* or θ_3^*, and so an implication of Theorem 3 is that there is almost sure convergence of the belief to either θ_1^* or θ_3^* (which implies convergence of the action as well, because these are beliefs where the optimal action is unique). In addition, there is zero probability of convergence to θ_2^*. Moreover, as Figure X illustrates (see also footnote 27), there can be no convergence to

[27] We are ignoring the cases starting from $\theta \in \bar{\Theta}$. Here, there are two optimal actions, say $x_1^*(\theta)$ and $x_2^*(\theta)$, and $\theta^*(\delta_{x^*(\theta)})$ should be replaced by a weighted average of $\theta^*(\delta_{x_1^*(\theta)})$ and $\theta^*(\delta_{x_2^*(\theta)})$. If both of these values are lower (or higher) than θ, then the analysis proceeds as before. If $\theta^*(\delta_{x_1^*(\theta)}) < \theta < \theta^*(\delta_{x_2^*(\theta)})$, then the agent either chooses action $x_1^*(\theta)$ and $\boldsymbol{\theta}(\boldsymbol{\sigma}(t))$ decreases and stays below θ forever, or she chooses $x_2^*(\theta)$ and $\boldsymbol{\theta}(\boldsymbol{\sigma}(t))$ increases and stays above θ forever (recall that the agent is not allowed to explicitly mix). In particular, there can be no convergence to a nondegenerate (i.e., mixed) equilibrium.

a nondegenerate (i.e., mixed) equilibrium in this problem. These features do not rely on the specifics of the figure: The analysis remains generally true due to the monotonicity of the mapping $\theta \mapsto \theta^*(\delta_{x^*(\theta)})$, which captures the economic idea of positively reinforcing actions.

6 Conclusion

Settings where agents have a misspecified model of their environment are becoming increasingly common in economics. One reason is the growing recognition that it is natural for people to be uncertain of their environment and to represent this uncertainty via parsimonious models. Another reason is that this literature can explain how biases in beliefs and behavior arise endogenously as a function of the agent's model, as opposed to simply postulating that agents have a specific, exogenous, incorrect belief. The literature, however, has mostly proceeded by studying different examples or applications in isolation.

In this paper, we developed general techniques that can be used to characterize asymptotic beliefs and actions in a large class of settings. The starting insight is that beliefs can be asymptotically characterized as a function of the frequency of actions. We can then use this result to characterize asymptotic frequencies. Even if actions do not converge, it is possible that frequencies converge, thus providing useful information about the regularity of asymptotic behavior. Our focus on frequencies also leads to new insights regarding the nature of equilibrium analysis. Key features of previous equilibrium concepts, such as the existence of a unique belief rationalizing behavior and the condition of myopic, not dynamic, optimization, are well-justified provided that the subjective model is weakly identified. Finally, we show that a differential inclusion can be used to study the asymptotic dynamics of behavior, including whether or not the action or frequency of actions converges.

Following most of the literature, we have taken the agent's misspecified model as a primitive of the environment. But, ideally, one would want to know which misspecified models people use in different circumstances. This is a question that could benefit from both theoretical and empirical analysis. We should explore how people choose which models or paradigms to adopt and how these paradigms are updated.[28] On the theory side, it seems important to formalize why people use parsimonious models in the first place.[29] On the experimental and empirical sides, we can test to what extent different models explain behavior and we can ex-

[28] For example, Esponda (2008) further restricts the agent's model to satisfy a payoff consistency requirement and Cho and Kasa (2014) assume that the agent tries to detect misspecification by running certain tests.

[29] See, for example, Aragones, Gilboa, Postlewaite and Schmeidler (2005), Al-Najjar (2009), Al-Najjar and Pai (2013), and Schwartzstein (2014).

amine robustness to non-traditional identifying assumptions (such as relaxing rational expectations) when estimating the primitives of an empirical model.[30] Many of these advances are already taking place. By developing tools that make environments with misspecified learning more tractable and easier to analyze, we hope to encourage even further work.

[30]For experimental work where subjects must learn the primitives and different subjective models are evaluated, see Esponda and Vespa (2018). For a review of empirical work relaxing the Nash identifying assumption in industrial organization, see Aguirregabiria and Jeon (2018).

References

Aghion, P., P. Bolton, C. Harris, and B. Jullien, "Optimal learning by experimentation," *The review of economic studies*, 1991, *58* (4), 621–654.

Aguirregabiria, Victor and Jihye Jeon, "Firms' Beliefs and Learning: Models, Identification, and Empirical Evidence," 2018.

Al-Najjar, N., "Decision Makers as Statisticians: Diversity, Ambiguity and Learning," *Econometrica*, 2009, *77* (5), 1371–1401.

_ **and M. Pai**, "Coarse decision making and overfitting," *Journal of Economic Theory, forthcoming*, 2013.

Aliprantis, C.D. and K.C. Border, *Infinite dimensional analysis: a hitchhiker's guide*, Springer Verlag, 2006.

Aragones, E., I. Gilboa, A. Postlewaite, and D. Schmeidler, "Fact-Free Learning," *American Economic Review*, 2005, *95* (5), 1355–1368.

Arrow, K. and J. Green, "Notes on Expectations Equilibria in Bayesian Settings," *Institute for Mathematical Studies in the Social Sciences Working Paper No. 33*, 1973.

Aubin, J-P and Arrigo Cellina, *Differential inclusions: set-valued maps and viability theory*, Vol. 264, Springer Science & Business Media, 2012.

Benaim, M. and M.W. Hirsch, "Mixed equilibria and dynamical systems arising from fictitious play in perturbed games," *Games and Economic Behavior*, 1999, *29* (1-2), 36–72.

Benaim, Michel, "A dynamical system approach to stochastic approximations," *SIAM Journal on Control and Optimization*, 1996, *34* (2), 437–472.

Benaïm, Michel, Josef Hofbauer, and Sylvain Sorin, "Stochastic approximations and differential inclusions," *SIAM Journal on Control and Optimization*, 2005, *44* (1), 328–348.

Berk, R.H., "Limiting behavior of posterior distributions when the model is incorrect," *The Annals of Mathematical Statistics*, 1966, *37* (1), 51–58.

Bohren, J Aislinn, "Informational herding with model misspecification," *Journal of Economic Theory*, 2016, *163*, 222–247.

__ **and Daniel N Hauser**, "Social Learning with Model Misspecification: A Framework and a Robustness Result," 2018.

Borkar, Vivek S, *Stochastic approximation: a dynamical systems viewpoint*, Vol. 48, Springer, 2009.

Bunke, O. and X. Milhaud, "Asymptotic behavior of Bayes estimates under possibly incorrect models," *The Annals of Statistics*, 1998, *26* (2), 617–644.

Cho, In-Koo and Kenneth Kasa, "Learning and model validation," *The Review of Economic Studies*, 2014, *82* (1), 45–82.

Easley, D. and N.M. Kiefer, "Controlling a stochastic process with unknown parameters," *Econometrica*, 1988, pp. 1045–1064.

Esponda, I., "Behavioral equilibrium in economies with adverse selection," *The American Economic Review*, 2008, *98* (4), 1269–1291.

__ **and D. Pouzo**, "Berk–Nash Equilibrium: A Framework for Modeling Agents With Misspecified Models," *Econometrica*, 2016, *84* (3), 1093–1130.

__ **and** __ , "Conditional Retrospective Voting in Large Elections," *American Economic Journal: Microeconomics*, 2017, *9* (2), 54–75.

__ **and** __ , "Eqilibrium in Misspecified Markov Decision Processes," *working paper*, 2017.

__ **and** __ , "Retrospective voting and party polarization," *International Economic Review*, 2019, *60* (1), 157–186.

__ **and E. I. Vespa**, "Endogenous sample selection: A laboratory study," *Quantitative Economics*, 2018, *9* (1), 183–216.

Eyster, E. and M. Rabin, "Cursed equilibrium," *Econometrica*, 2005, *73* (5), 1623–1672.

Eyster, Erik and Matthew Rabin, "Naive herding in rich-information settings," *American economic journal: microeconomics*, 2010, *2* (4), 221–43.

Frick, Mira, Ryota Iijima, and Yuhta Ishii, "Misinterpreting Others and the Fragility of Social Learning," 2019.

Fudenberg, D. and D. Kreps, "Learning Mixed Equilibria," *Games and Economic Behavior*, 1993, *5*, 320–367.

_ **and D.K. Levine**, "Self-confirming equilibrium," *Econometrica*, 1993, pp. 523–545.

_ **and** _ , *The theory of learning in games*, Vol. 2, The MIT press, 1998.

Fudenberg, Drew, Gleb Romanyuk, and Philipp Strack, "Active learning with a misspecified prior," *Theoretical Economics*, 2017, *12* (3), 1155–1189.

Gagnon-Bartsch, Tristan and Matthew Rabin, "Naive social learning, mislearning, and unlearning," *work*, 2017.

Harsanyi, J.C., "Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points," *International Journal of Game Theory*, 1973, *2* (1), 1–23.

Heidhues, Paul, Botond Kőszegi, and Philipp Strack, "Unrealistic expectations and misguided learning," *Econometrica*, August 2018, *86* (4), 1159–1214.

Hofbauer, J. and W.H. Sandholm, "On the global convergence of stochastic fictitious play," *Econometrica*, 2002, *70* (6), 2265–2294.

Jehiel, P., "Limited horizon forecast in repeated alternate games," *Journal of Economic Theory*, 1995, *67* (2), 497–519.

_ , "Analogy-based expectation equilibrium," *Journal of Economic theory*, 2005, *123* (2), 81–104.

_ **and F. Koessler**, "Revisiting games of incomplete information with analogy-based expectations," *Games and Economic Behavior*, 2008, *62* (2), 533–557.

Jehiel, Philippe, "Investment strategy and selection bias: An equilibrium perspective on overoptimism," *American Economic Review*, 2018, *108* (6), 1582–97.

Kirman, A. P., "Learning by firms about demand conditions," in R. H. Day and T. Groves, eds., *Adaptive economic models*, Academic Press 1975, pp. 137–156.

Lazear, Edward P, "Performance pay and productivity," *American Economic Review*, 2000, *90* (5), 1346–1361.

Liebman, Jeffrey B and Richard J Zeckhauser, "Schmeduling," 2004.

Molavi, Pooya, "Macroeconomics with Learning and Misspecification: A General Theory and Applications," 2018.

Nyarko, Y., "Learning in mis-specified models and the possibility of cycles," *Journal of Economic Theory*, 1991, *55* (2), 416–427.

Osborne, M.J. and A. Rubinstein, "Games with procedurally rational players," *American Economic Review*, 1998, *88*, 834–849.

Rabin, M. and D. Vayanos, "The gambler's and hot-hand fallacies: Theory and applications," *The Review of Economic Studies*, 2010, *77* (2), 730–778.

Sargent, T. J., *Bounded rationality in macroeconomics*, Oxford University Press, 1993.

Schwartzstein, Joshua, "Selective attention and learning," *Journal of the European Economic Association*, 2014, *12* (6), 1423–1452.

Sobel, J., "Non-linear prices and price-taking behavior," *Journal of Economic Behavior & Organization*, 1984, *5* (3), 387–396.

Spiegler, Ran, "Bayesian networks and boundedly rational expectations," *The Quarterly Journal of Economics*, 2016, *131* (3), 1243–1290.

_ , "Data Monkeys: A Procedural Model of Extrapolation from Partial Statistics," *The Review of Economic Studies*, 2017, *84* (4), 1818–1841.

Tversky, T. and D. Kahneman, "Availability: A heuristic for judging frequency and probability," *Cognitive Psychology*, 1973, *5*, 207–232.

A Appendix

In this appendix, we present the proofs omitted from the text. In some places, we use the fact that $\theta \mapsto \log \frac{q(Y|x)}{q_\theta(Y|x)}$ is finite and continuous $Q(\cdot|x) - a.s.$ for all $x \in \mathbb{X}$. This fact follows from Assumptions 1-2.

A.1 Proof of Lemma 1

Continuity of K: For any $(\theta,\sigma)\in\Theta\times\Delta(\mathbb{X})$ take a sequence $(\theta_n,\sigma_n)_n$ in $\Theta\times\Delta(\mathbb{X})$ that converges to this point. By the triangle inequality and the fact that K is finite under Assumption 2(iii) it follows that

$$|K(\theta_n,\sigma_n)-K(\theta,\sigma)|\leq|K(\theta_n,\sigma)-K(\theta,\sigma)|+|K(\theta_n,\sigma_n)-K(\theta_n,\sigma)|.$$

It suffices to show that both terms on the RHS vanish as $n\to\infty$. Regarding the first term in the RHS, observe that for any $\sigma\in\Delta(\mathbb{X})$, $\theta\mapsto\log\frac{q(Y|X)}{q_\theta(Y|X)}$ is finite and continuous $Q\cdot\sigma-a.s.$ Under Assumption 2(iii), by the DCT this implies that $\theta\mapsto K(\theta,\sigma)$ is continuous for any $\sigma\in\Delta(\mathbb{X})$. Thus $\lim_{n\to\infty}|K(\theta_n,\sigma)-K(\theta,\sigma)|=0$. Regarding the other term in the RHS of the display, observe that under Assumption 2(iii)

$$|K(\theta_n,\sigma_n)-K(\theta_n,\sigma)|\leq\sum_{x\in\mathbb{X}}\int g_x(y)Q(dy\mid x)|\sigma_n(x)-\sigma(x)|$$

and the RHS vanishes as $\int g_x(y)Q(dy\mid x)<\infty$ for all $x\in\mathbb{X}$.

Finally, continuity of K, compactness of Θ (by Assumption 2(ii)) and the Theorem of the Maximum imply that $\sigma\mapsto\Theta(\sigma)$ is compact-valued, uhc, and that $\sigma\mapsto K^*(\sigma)$ is continuous.

A.2 Proof of Lemma 2

Let $(\theta,z)\mapsto g(\theta,z)\equiv\log\frac{q(y|x)}{q_\theta(y|x)}$, where $z=(y,x)\in\mathbb{Y}\times\mathbb{X}$. For any $\theta\in\Theta$ and any $\varepsilon>0$, let $O(\theta,\varepsilon)\equiv\{\theta'\colon||\theta'-\theta||<\varepsilon\}$.

STEP 1. Pointwise convergence. Fix any $\varepsilon>0$ and any $\theta\in\Theta$. For any $\tau\geq0$ and history h, let

$$\zeta_\tau(h)\equiv\sup_{\theta'\in O(\theta,\varepsilon)}g(\theta',z_\tau(h))-E_{Q(\cdot|x_\tau(h))}\left[\sup_{\theta'\in O(\theta,\varepsilon)}g(\theta',Y,x_\tau(h))\right].$$

The process $(\zeta_t)_t$ is a Martingale difference under $\mathbf{P}^f$ and the filtration generated by $\{h^t\equiv(x_0(h),y_0(h),x_1(h),y_1(h),...,x_t(h))\colon t\geq0\}$, because $E_{\mathbf{P}^f(\cdot|h^t)}[\zeta_t(h)]=0$ for all t. Define $h\mapsto\zeta^t(h)\equiv\sum_{\tau=0}^t(1+\tau)^{-1}\zeta_\tau(h)$ for any $t\geq0$. Since $(\zeta_t)_t$ is a Martingale difference sequence, then $(\zeta^t)_t$ is also a Martingale difference.

By the Martingale Convergence Theorem, there exist a $\mathscr{H}\subseteq\mathbb{H}$ (potentially depending on $\theta\in\Theta$) and $\zeta\in L^2(\mathbb{H},\mathbb{R},\mathbf{P}^f)$ such that $\mathbf{P}^f(\mathscr{H})=1$ and, for any $h\in\mathscr{H}$, $\zeta^t(h)\to\zeta(h)$, provided $\sup_t E_{\mathbf{P}^f}\left[(\zeta^t)^2\right] < \infty$. This condition is satisfied because

$$\begin{aligned}
E_{\mathbf{P}^f}\left[(\zeta^t)^2\right] &= E_{\mathbf{P}^f}\left[\sum_{\tau=0}^{t}(1+\tau)^{-2}(\zeta_\tau)^2\right] + 2E_{\mathbf{P}^f}\left[\sum_{\tau>\tau'}(1+\tau)^{-1}\left(1+\tau'\right)^{-1}\zeta_\tau\zeta_{\tau'}\right] \\
&= \sum_{\tau=0}^{t}(1+\tau)^{-2}E_{\mathbf{P}^f}\left[(\zeta_\tau)^2\right] \\
&\leq \sum_{\tau=0}^{t}(1+\tau)^{-2}E_{\boldsymbol{P}^f}\left[\int\left(\sup_{\theta'\in O(\theta,\varepsilon)} g(\theta',y,X_\tau)\right)^2 Q(dy \mid X_\tau)\right] \\
&\leq C\max_{x\in\mathbb{X}}\int \sup_{\theta'\in O(\theta,\varepsilon)}\left(g(\theta',y,x)\right)^2 Q(dy \mid x),
\end{aligned}$$

where the second line follows from the fact that, for any $\tau > \tau'$, $E_{\mathbf{P}^f}[\zeta_\tau\zeta_{\tau'}] = E_{\mathbf{P}^f}\left[E_{\mathbf{P}^f(\cdot|h^\tau)}[\zeta_\tau]\,\zeta_{\tau'}\right] = 0$, and where the last line follows from the fact that $C \equiv \lim_{t\to\infty}\sum_{\tau=0}^{t}(1+\tau)^{-2} < \infty$. By Assumption 2(iii), for any $(x,y)\in\mathbb{X}\times\mathbb{Y}$, $\sup_{\theta'\in O(\theta,\varepsilon)}(g(\theta',y,x))^2 \leq (g_x(y))^2$ with $\int (g_x(y))^2 Q(dy \mid x) < \infty$. Thus, $\sup_t E_{\mathbf{P}^f}\left[(\zeta^t)^2\right] < \infty$. By invoking Kronecker Lemma it follows that

$$\lim_{t\to\infty}(1+t)^{-1}\sum_{\tau=0}^{t}\zeta^t = 0$$

$\boldsymbol{P}^f$-a.s. Therefore, we have established that, for all $\theta\in\Theta$,

$$\lim_{t\to\infty}(1+t)^{-1}\sum_{\tau=0}^{t}\left(\sup_{\theta'\in O(\theta,\varepsilon)} g(\theta',z_\tau) - E_{Q(\cdot|x_\tau)}\left[\sup_{\theta'\in O(\theta,\varepsilon)} g(\theta',Y,x_\tau)\right]\right) = 0$$

$\boldsymbol{P}^f$-a.s.

STEP 2. Uniform convergence. Observe that, for any $\varepsilon > 0$ and any $\theta\in\Theta$, there exists $\delta(\theta,\varepsilon)$ such that

$$E_{Q(\cdot|x)}\left[\sup_{\theta'\in O(\theta,\delta(\theta,\varepsilon))} g(\theta',Y,x) - g(\theta,Y,x)\right] < 0.25\varepsilon \tag{20}$$

for all $x\in\mathbb{X}$. To see this claim, note that, since $\theta\mapsto g(\theta,Y,x)$ is continuous $Q(\cdot|x)-a.s.$ for all $x\in\mathbb{X}$, $\lim_{\delta\to 0}\sup_{\theta'\in O(\theta,\delta)}|g(\theta',Y,x)-g(\theta,Y,x)| = 0$ a.s.$-Q(\cdot \mid x)$ for all $x\in\mathbb{X}$. Also, by Assumption 2(iii), $\sup_{\theta'\in O(\theta,\delta)}|g(\theta',y,x)-g(\theta,y,x)| \leq 2g_x(y)$ and $\int g_x(y)Q(dy|x) < \infty$, Thus, by the DCT, $\lim_{\delta\to 0} E_{Q(\cdot|x)}\left[\sup_{\theta'\in O(\theta,\delta)} |g(\theta',Y,x)-g(\theta,Y,x)|\right]=0$ for all $x\in\mathbb{X}$.

Observe that $(O(\theta,\delta(\theta,\varepsilon)))_{\theta\in\Theta}$ is an open cover of Θ. By compactness of Θ, there exists a finite sub-cover $(O(\theta_j,\delta(\theta_j,\varepsilon)))_{j=1,...J(\varepsilon)}$. Thus, for all $\varepsilon>0$,

$$\begin{aligned}
&\sup_{\theta\in\Theta}\left|(1+t)^{-1}\sum_{\tau=0}^{t}\left(g(\theta,z_\tau)-E_{Q(\cdot|x_\tau)}[g(\theta,Y,x_\tau)]\right)\right| \\
\leq&\max_j \sup_{\theta\in O(\theta_j,\delta(\theta_j,\varepsilon))}\left|(1+t)^{-1}\sum_{\tau=0}^{t}\left(g(\theta,z_\tau)-E_{Q(\cdot|x_\tau)}[g(\theta,Y,x_\tau)]\right)\right| \\
\leq&\max_j (1+t)^{-1}\sum_{\tau=0}^{t}\left(\sup_{\theta\in O(\theta_j,\delta(\theta_j,\varepsilon))}\left|g(\theta,z_\tau)-E_{Q(\cdot|x_\tau)}[g(\theta,Y,x_\tau)]\right|\right) \\
\leq&\max_j (1+t)^{-1}\sum_{\tau=0}^{t}\left(\left|\sup_{\theta\in O(\theta_j,\delta(\theta_j,\varepsilon))} g(\theta,z_\tau)-E_{Q(\cdot|x_\tau)}\left[\inf_{\theta\in O(\theta_j,\delta(\theta_j,\varepsilon))} g(\theta,Y,x_\tau)\right]\right|\right) \\
\leq&\max_j (1+t)^{-1}\sum_{\tau=0}^{t}\left(\left|\sup_{\theta\in O(\theta_j,\delta(\theta_j,\varepsilon))} g(\theta,z_\tau)-E_{Q(\cdot|x_\tau)}\left[\sup_{\theta\in O(\theta_j,\delta(\theta_j,\varepsilon))} g(\theta,Y,x_\tau)\right]\right|\right) \\
&+\max_j (1+t)^{-1}\sum_{\tau=0}^{t}\left(E_{Q(\cdot|x_\tau)}\left[\sup_{\theta\in O(\theta_j,\delta(\theta_j,\varepsilon))} g(\theta,Y,x_\tau)-\inf_{\theta\in O(\theta_j,\delta(\theta_j,\varepsilon))} g(\theta,Y,x_\tau)\right]\right) \\
=&I+II.
\end{aligned}$$

By Step 1 and the fact that we are adding over a finite number of θ_j's, the limit as $t\to\infty$ of the term I is equal to zero $\boldsymbol{P}^f$-a.s. For the second term, note that (20) implies that

$$II \leq 2\max_{x\in\mathbb{X}}\int \sup_{\theta\in O(\theta_j,\delta(\theta_j,\varepsilon))}\left|g(\theta,y,x)-g(\theta_j,y,x)\right| Q(dy\mid x)\leq 0.5\varepsilon.$$

Since $0\leq II\leq 0.5\varepsilon$ holds for all $\varepsilon>0$, it follows that $II=0$. Therefore, using the definition of g, we have established that

$$\lim_{t\to\infty}\sup_{\theta\in\Theta}(1+t)^{-1}\sum_{\tau=0}^{t}\left(\log\frac{q(y_\tau\mid x_\tau)}{q_\theta(y_\tau\mid x_\tau)}-E_{Q(\cdot|x_\tau)}\left[\log\frac{q(Y\mid x_\tau)}{q_\theta(Y\mid x_\tau)}\right]\right)=0$$

$\boldsymbol{P}^f$-a.s. The statement in the lemma then follows by noting that

$$K(\theta,\sigma_t)=\sum_{x\in\mathbb{X}}E_{Q(\cdot|x)}\left[\log\frac{q(Y\mid x)}{q_\theta(Y\mid x)}\right]\sigma_t(x)=(1+t)^{-1}\sum_{\tau=0}^{t}E_{Q(\cdot|x_\tau)}\left[\log\frac{q(Y\mid x_\tau)}{q_\theta(Y\mid x_\tau)}\right].$$

A.3 Proof of equation (8) in Theorem 1

For simplicity, set $k \equiv \varepsilon/2 > 0$. Continuity of $(\theta,\sigma) \mapsto \bar{K}(\theta,\sigma) \equiv K(\theta,\sigma) - K^*(\sigma)$ (see Lemma 1(i)) and compactness of $\Theta \times \Delta(\mathbb{X})$ imply that $\bar{K}$ is uniformly continuous. For any σ, take some $\theta_\sigma \in \Theta(\sigma)$ (this is possible because $\Theta(\sigma)$ is nonempty; see Lemma 1(ii)). By uniform continuity of $\bar{K}$, there exists $\delta_k > 0$ such that $\|\theta_\sigma - \theta'\| < \delta_k$ and $\|\sigma - \sigma'\| < \delta_k$ imply $\bar{K}(\theta',\sigma') < \bar{K}(\theta_\sigma,\sigma) + k = k$, where the last equality follows because $\bar{K}(\theta_\sigma,\sigma) = 0$. This implies that for all σ, $\{\theta' : \|\theta_\sigma - \theta'\| < \delta_k\} \subseteq \{\theta : \bar{K}(\theta,\sigma') \le k\}$ for all $\sigma' \in B(\sigma,\delta_k) \equiv \{\sigma' : \|\sigma - \sigma'\| < \delta_k\}$. Thus, for all σ, $\mu_0(\{\theta : \bar{K}(\theta,\sigma') \le k\}) \ge \mu_0(\{\theta' : \|\theta_\sigma - \theta'\| < \delta_k\})$ for all $\sigma' \in B(\sigma,\delta_k)$. The balls $\{B(\sigma,\delta_k)\}_\sigma$ form an open cover for $\Delta(\mathbb{X})$. Since $\Delta(\mathbb{X})$ is compact, there exists a finite subcover $\{B(\sigma^i,\delta_k)\}_{i=1}^n$. Let $r \equiv \min_{i\in\{1,\ldots,n\}} \mu_0(\{\theta' : \|\theta_\sigma - \theta'\| < \delta_k\})$ which is strictly positive by Assumption 3. Take any σ', there exists i such that $\sigma' \in B(\sigma^i,\delta_k)$; by the previous argument $\mu_0(\{\theta : \bar{K}(\theta,\sigma') \le k\}) \ge \mu_0(\{\theta' : \|\theta_{\sigma^i} - \theta'\| < \delta_k\}) \ge r > 0$.

A.4 Proof of Corollary 1

We will show that every (weakly) convergent subsequence $(\mu_{t(k)})_k$ (converging to μ) has the desired limiting property, which then implies, by compactness of $\Delta(\Theta)$, that the sequence has the desired limiting property. Observe that $\mu\{(\theta : \bar{K}(\theta,\sigma) = 0\}) = 1$ is equivalent to $\mu(\Theta(\sigma)) = 1$ where $(\theta,\sigma) \mapsto \bar{K}(\theta,\sigma) \equiv K(\theta,\sigma) - K^*(\sigma)$, so it suffices to show that $\mu\{(\theta : \bar{K}(\theta,\sigma) = 0\}) = 1$ or, equivalently, $\int \bar{K}(\theta,\sigma)\mu(d\theta) = 0$. To show this, observe that

$$\int \bar{K}(\theta,\sigma)\mu(d\theta) \le \left| \int \bar{K}(\theta,\sigma)\mu(d\theta) - \int \bar{K}(\theta,\sigma)\mu_{t(k)+1}(d\theta) \right| + \left| \int \bar{K}(\theta,\sigma)\mu_{t(k)+1}(d\theta) - \int \bar{K}(\theta,\sigma_{t(k)})\mu_{t(k)+1}(d\theta) \right| + \int \bar{K}(\theta,\sigma_{t(k)})\mu_{t(k)+1}(d\theta).$$

So it suffices to show that each term in the RHS vanishes as $k \to \infty$. The first term vanishes because $\mu_{t(k)}$ converges to μ and $\theta \mapsto \bar{K}(\theta,\sigma)$ is continuous (see Lemma 1(i)). The second term vanishes because continuity of $\bar{K}$ (see Lemma 1(i)) and compactness of $\Theta \times \Delta(\mathbb{X})$ imply that $(\theta,\sigma) \mapsto \bar{K}(\theta,\sigma)$ is uniformly continuous. Thus, for any $\varepsilon > 0$, there exists K such that $|\bar{K}(\theta,\sigma_{t(k)}) - \bar{K}(\theta,\sigma)| < \varepsilon$ for all $k \ge K$ and all $\theta \in \Theta$. By the uniformity over θ, it follows that $\left|\int \bar{K}(\theta,\sigma)\mu_{t(k)+1}(d\theta) - \int \bar{K}(\theta,\sigma_{t(k)})\mu_{t(k)+1}(d\theta)\right| < \varepsilon$ for all $k \ge K$. Finally, the last term vanishes by Theorem 1.

A.5 Proof of Proposition 2

Since $\mathbb{X}$ is nonempty and finite, $\Delta(\mathbb{X})$ is a nonempty, compact and convex subset of an Euclidean space. Moreover, it is immediate that $\Delta\left(\cup_{\mu\in\Delta(\Theta(\sigma))}F(\mu)\right)$ is convex for all σ. Next, we will establish the claim that $\sigma\mapsto\Delta\left(\cup_{\mu\in\Delta(\Theta(\sigma))}F(\mu)\right)$ is uhc: Let $(y_n,\sigma_n)_n$ converge to (y,σ) and suppose that $y_n\in\Delta\left(\cup_{\mu\in\Delta(\Theta(\sigma_n))}F(\mu)\right)$ for all n. Take any $x\in\mathbb{X}$ such that $y(x)>0$, then, for all n sufficiently large, $y_n(x)>0$ and, therefore, there exists $\mu_{x,n}\in\Delta(\Theta(\sigma_n))$ such that $x\in F(\mu_{x,n})$. By compactness of $\Delta(\Theta)$, we can take a further subsequence $n(k)$ such that $\mu_{x,n(k)}$ converges to μ_x. By Lemma 1, $\sigma\mapsto\Theta(\sigma)$ is uhc; hence $\sigma\mapsto\Delta(\Theta(\sigma))$ is also UHC (for a proof, see Aliprantis and Border (2006)). Therefore, $\mu_x\in\Delta(\Theta(\sigma))$ and, by uhc of F, $x\in F(\mu_x)$. Therefore, $y\in\Delta\left(\cup_{\mu\in\Delta(\Theta(\sigma))}F(\mu)\right)$ and the uhc claim is established. The existence of a solution to (9) then follows from Kakutani's fixed-point theorem.

A.6 Proof of Proposition 3

Fix any x such that $\sigma(x)>0$. Since $\sigma\in\Delta\left(\cup_{\mu\in\Delta(\Theta(\sigma))}F_\beta(\mu)\right)$, there exists $\mu_x\in\Delta(\Theta(\sigma))$ such that $x\in F_\beta(\mu_x)$. This means that, for any $x'\in\mathbb{X}$,

$$\begin{aligned}\int(\pi(x,y)+\beta V(B(x,y,\mu_x))\bar{Q}_{\mu_x}(dy\mid x) &= \int(\pi(x,y)\bar{Q}_{\mu_x}(dy\mid x)+\beta V(\mu_x)\\ &\geq \int\left(\pi(x',y)+\beta V(B(x',y,\mu_x))\right)\bar{Q}_{\mu_x}(dy\mid x')\\ &\geq \int(\pi(x',y)\bar{Q}_{\mu_x}(dy\mid x')+\beta V(\mu_x),\end{aligned}$$

where the first line follows from weak identification (which implies $B(x,y,\mu_x)=\mu_x$ for all y in the support of $\bar{Q}_{\mu_x}(\cdot\mid x)$), the second line follows from $x\in F_\beta(\mu_x)$, and the third line follows from the convexity of the value function and the martingale property of Bayesian updating (which imply, using Jensen's inequality, $\int V(B(x',y,\mu_x))\bar{Q}_{\mu_x}(dy\mid x')\geq V(\int B(x',y,\mu_x)\bar{Q}_{\mu_x}(dy\mid x'))=V(\mu_x)$). Therefore, x is myopically the best action, i.e., $x\in F_0(\mu_x)$.

Finally, we will show that any action in the support σ can be justified by the belief μ_x; thus, there is single belief that rationalizes all actions in the support of σ. Let x'' be an action in the support of σ. Since $\sigma\in\Delta\left(\cup_{\mu\in\Delta(\Theta(\sigma))}F_\beta(\mu)\right)$, we know, by repeating the argument of the previous step for action x'', that there exists $\mu_{x''}\in\Delta(\Theta(\sigma))$ such that $x''\in F_0(\mu_{x''})$. By weak identification and the fact that $\mu_{x''}$ and μ_x both belong to $\Delta(\Theta(\sigma))$, $\bar{Q}_{\mu_{x''}}(\cdot\mid\tilde{x})=\bar{Q}_{\mu_x}(\cdot\mid\tilde{x})$ for all $\tilde{x}$ in the support of σ. Therefore, for any $x' \in \mathbb{X}$,

$$\int \pi(y,x'')\bar{Q}_{\mu_x}(dy|x'') = \int \pi(y,x'')\bar{Q}_{\mu_{x''}}(dy|x'') \geq \int \pi(y,x')\bar{Q}_{\mu_{x''}}(dy|x')$$

and so $x'' \in F_0(\mu_x)$.

A.7 Proof of Theorem 3

The proof of Theorem 3 consists of three parts. Part 1 defines an enlargement of the set of actions that allows us to adopt the methods developed by BHS2005. Part 2 and 3 correspond to the arguments in the proofs of Proposition 1.3 and Theorem 4.2 in BHS2005, respectively, and we provide them here for completeness. Throughout the proof we fix a history from the set of histories with probability 1 defined by the statement of Theorem 1; we omit the history from the notation.

Part 1. Enlargement of the set $\Delta(F(\mu))$.

Let $\mathbb{S} = \{a - b \mid a,b \in \Delta(\mathbb{X})\}$ and let $\Xi : \mathbb{R}_+ \times \Delta(\mathbb{X}) \rightrightarrows \mathbb{S}$ be defined such that, for all $(\delta,\sigma) \in \mathbb{R}_+ \times \Delta(\mathbb{X})$,

$$\Xi(\delta,\sigma) = \left\{ y \in \mathbb{S} : \begin{array}{c} \exists \sigma' \in \Delta(\mathbb{X}), \mu' \in \Delta(\Theta) \;\; s.t. \;\; y \in \Delta(F(\mu')) - \sigma', \\ \mu' \in M(\delta,\sigma'), \|\sigma' - \sigma\| \leq \delta \end{array} \right\},$$

where $M : \mathbb{R}_+ \times \Delta(\mathbb{X}) \rightrightarrows \Delta(\Theta)$ is defined such that, for all $(\delta,\sigma') \in \mathbb{R}_+ \times \Delta(\mathbb{X})$,

$$M(\delta,\sigma') \equiv \{\mu' \in \Delta(\Theta) : \int_\Theta \bar{K}(\theta,\sigma')\mu'(d\theta) \leq \delta\},$$

where $\bar{K}(\theta,\sigma') \equiv K(\theta,\sigma') - K^*(\sigma')$. Note that $\Theta(0,\sigma) = \Theta(\sigma)$ and so $\Xi(0,\sigma) = \cup_{\mu \in \Delta(\Theta(\sigma))}\Delta(F(\mu)) - \sigma$.

Claim 1: $(\delta,\sigma) \mapsto \Xi(\delta,\sigma)$ is uhc.

Proof. Because $\mathbb{S}$ is compact, it suffices to show that Ξ has the closed graph property. For this purpose, we will first show that $(\delta,\sigma') \mapsto M(\delta,\sigma')$ is uhc. To establish this claim, note that $\Delta(\Theta)$ is compact because of the assumption that Θ is compact. Hence, we will show that M has the closed graph property. Take $(\mu'_n)_n$ converging to μ' (in the weak topology), $(\delta_n)_n$ converging to δ, and $(\sigma'_n)_n$ converging to σ'. Suppose that $\mu'_n \in M(\delta_n,\sigma'_n)$ for all n. We will show that $\mu' \in M(\delta,\sigma')$. Since $(\mu'_n)_n$ converges (weakly) to μ' and $\bar{K}(\theta,\cdot)$ is continuous (see

Lemma 1), it follows that

$$\lim_n \left(\int_\Theta \bar{K}(\theta, \sigma_n')\mu_n'(d\theta) - \int_\Theta \bar{K}(\theta, \sigma')\mu'(d\theta) \right) = \lim_n \left(\int_\Theta \bar{K}(\theta, \sigma_n')\mu_n'(d\theta) - \int_\Theta \bar{K}(\theta, \sigma')\mu_n'(d\theta) \right)$$
$$+ \lim_n \left(\int_\Theta \bar{K}(\theta, \sigma')\mu_n'(d\theta) - \int_\Theta \bar{K}(\theta, \sigma')\mu'(d\theta) \right)$$
$$= 0.$$

Also, since $\mu_n' \in M(\delta_n, \sigma_n')$, then $\int_\Theta \bar{K}(\theta, \sigma_n')\mu_n'(d\theta) \leq \delta_n$. Taking limits of this last expression on both sides, we obtain $\int_\Theta \bar{K}(\theta, \sigma')\mu'(d\theta) \leq \delta$, implying that $\mu' \in M(\delta, \sigma')$.

Next, to show that Ξ has the closed graph property, take $(y_n)_n$ converging to y, $(\delta_n)_n$ converging to δ, and $(\sigma_n)_n$ converging to σ. Suppose that $y_n \in \Xi(\delta_n, \sigma_n)$ for all n. We will show that $y \in \Xi(\delta, \sigma)$. Since $y_n \in \Xi(\delta_n, \sigma_n)$ for all n, there exists a sequence $(\mu_n', \sigma_n')_n$ such that $y_n \in \Delta(F(\mu_n')) - \sigma_n'$, $\|\sigma_n' - \sigma_n\| \leq \delta_n$, and $\mu_n' \in M(\delta_n, \sigma_n')$. Because the sequence $(\mu_n, \sigma_n')_n$ lives in a compact set, $\Delta(\Theta) \times \Delta(\mathbb{X})$, there exists a subsequence, $(\mu_{n(k)}', \sigma_{n(k)}')_k$ that converges to (μ', σ'). By uhc of M and of $\mu \mapsto \Delta(F(\mu))$ (due to the assumption that F is uhc), it follows that $y \in \Delta(F(\mu')) - \sigma'$, $\|\sigma' - \sigma\| \leq \delta$, and $\mu' \in M(\delta, \sigma')$. Thus, $y \in \Xi(\delta, \sigma)$.

Claim 2: There exists a sequence $(\delta_t)_t$ with $\lim_{t\to\infty} \delta_t = 0$ such that, for all t, $\sigma_{t+1} - \sigma_t \in \frac{1}{t+1}\Xi(\delta_t, \sigma_t)$.

Proof. By equation (14) in the text, $\sigma_{t+1} - \sigma_t \in \frac{1}{t+1}(\Delta(F(\mu_{t+1})) - \sigma_t)$ for all t. By Theorem 1, there exists a sequence $(\delta_t)_t$ with $\lim_{t\to\infty} \delta_t = 0$ such that, for all t, $\int_\Theta \bar{K}(\theta, \sigma_t)\mu_{t+1}(d\theta) \leq \delta_t$. Thus, $\Delta(F(\mu_{t+1})) - \sigma_t \subseteq \Xi(\delta_t, \sigma_t)$ for all t, and the claim follows.

Part 2. The interpolation of $(\sigma_t)_t$ is what BHS2005 call a perturbed solution of the differential inclusion.

Define $m(t) \equiv \sup\{k \geq 0 : t \geq \tau_k\}$, where $\tau_0 = 0$ and $\tau_k = \sum_{i=1}^k 1/i$. Let $\mathbf{w}$ be the continuous-time interpolation of $(\sigma_t)_t$, as defined in equation (15) in the text. By Claim 2, for any t, $\mathbf{w}(t) \in \sigma_{m(t)} + (t - \tau_{m(t)})\Xi(\delta_{m(t)}, \sigma_{m(t)})$; hence, $\dot{\mathbf{w}}(t) \in \Xi(\delta_{m(t)}, \sigma_{m(t)})$ for almost every t. Let $\boldsymbol{\gamma}(t) \equiv \delta_{m(t)} + \left\|\mathbf{w}(t) - \sigma_{m(t)}\right\|$. Then $\dot{\mathbf{w}}(t) \in \Xi(\boldsymbol{\gamma}(t), \mathbf{w}(t))$ for almost every t. In addition, note that $\lim_{t\to\infty} \boldsymbol{\gamma}(t) = 0$ because $(\delta_t)_t$ goes to zero, $m(t)$ goes to infinity, and $\boldsymbol{w}$ is the interpolation of $(\sigma_t)_t$.

Part 3. A perturbed solution is an asymptotic pseudotrajectory (i.e., it satisfies equation (17) in the text).

Let $\boldsymbol{v}(t) \equiv \dot{\mathbf{w}}(t) \in \Xi(\boldsymbol{\gamma}(t), \mathbf{w}(t))$ for almost every t. Then

$$\mathbf{w}(t+s) - \mathbf{w}(t) = \int_0^s \mathbf{v}(t+\tau) d\tau. \tag{21}$$

Since $\mathbb{S}$ is a bounded set, $\boldsymbol{v}$ is uniformly bounded; therefore, $\mathbf{w}$ is uniformly continuous. Hence, the family $\mathbf{S}^t(\mathbf{w})$, defined as $\mathbf{S}^t(\mathbf{w})(s) = \mathbf{w}(s+t)$, is equicontinuous and, therefore, relatively compact with respect to $L^\infty(\mathbb{R}, \Delta(\mathbb{X}), Leb)$, where Leb is the Lebesgue measure; all L^p spaces in the proof are with respect to Lebesgue, so we drop it from subsequent notation. Therefore, there exists a subsequence $(t(n))_n$ and a $\boldsymbol{w}^* \in L^\infty(\mathbb{R}, \Delta(\mathbb{X}))$ such that $\mathbf{w}^* = \lim_{t(n)\to\infty} \mathbf{S}^t(\mathbf{w})$. Set $t = t_n$ in (21) and define $\mathbf{v}_n(s) = \mathbf{v}(t_n + s)$. Then

$$\mathbf{w}^*(s) - \mathbf{w}^*(0) = \lim_{n\to\infty} \int_0^s \mathbf{v}_n(\tau) d\tau.$$

Since $\mathbf{v}_n \in L^\infty(\mathbb{R}, \mathbb{S})$ for all n, then $\mathbf{v}_n \in L^2([0,T], \mathbb{S})$. By the Banach-Alouglu Theorem, there exists a subsequence, which we still denote as $(t(n))_n$, and a $\mathbf{v}^* \in L^2([0,T], \mathbb{S})$ such that $(\mathbf{v}_n)_n$ converges in the weak topology to $\mathbf{v}^*$; therefore,

$$\lim_{n\to\infty} \int_0^s \mathbf{v}_n(\tau) d\tau = \int_0^s \mathbf{v}^*(\tau) d\tau \tag{22}$$

pointwise in $s \in [0,T]$. Indeed, convergence is uniform because the family $\int_0^s \mathbf{v}_n(\tau) d\tau$ is equicontinuous and $[0,T]$ is compact.

The proof concludes by showing the claim that $\mathbf{v}^*(\tau) \in \Gamma_F(\mathbf{w}(\tau)) - \mathbf{w}(\tau)$ Lebesgue-a.s. in $\tau \in [0,T]$. We will prove it by showing that $\mathbf{v}^*(\tau) \in Co(\Xi(0, \mathbf{w}(\tau)))$, where Co denotes the convex hull; the desired claim then follows because the facts that $\Gamma_F(\sigma) - \sigma$ is a convex set and contains $\Xi(0,\sigma)$ and, by definition, $Co(\Xi(0,\sigma))$ is the smallest convex set that contains $\Xi(0,\sigma)$, imply that $Co(\Xi(0,\sigma)) \subseteq \Gamma_F(\sigma) - \sigma$.

We will prove $\mathbf{v}^*(\tau) \in Co(\Xi(0, \mathbf{w}(\tau)))$ in several steps. First, we show that weak convergence of $(\mathbf{v}_n)_n$ to $\mathbf{v}^*$ implies almost sure convergence of a weighted average of $(\mathbf{v}_n)_n$ to $\mathbf{v}^*$. Formally, by Mazur's Lemma, for each $n \in \mathbb{N}$, there exists a $N(n) \in \mathbb{N}$ and a non-negative vector, $(\alpha_n, ..., \alpha_{N(n)})$, such that $\sum_{i=n}^{N(n)} \alpha_i = 1$, and $\lim_{n\to\infty} \|\bar{\boldsymbol{v}}_n - \boldsymbol{v}^*\|_{L^2([0,T],\mathbb{S})} = 0$ where $\bar{\boldsymbol{v}}_n \equiv \sum_{k=n}^{N(n)} \alpha_k \boldsymbol{v}_n$. Therefore, as $\lim_{n\to\infty} \|\bar{\boldsymbol{v}}_n - \boldsymbol{v}^*\|_{L^2([0,T],\mathbb{S})} = 0$, it follows that $\lim_{j\to\infty} \bar{\boldsymbol{v}}_n = \boldsymbol{v}^*$ a.s.-Lebesgue.

Fix $\tau \in [0,T]$ such that the previous claim holds; recall it holds Lebesgue-a.s. Define $\boldsymbol{\gamma}_n(\tau) \equiv \gamma(t_n + \tau)$ and $\mathbf{w}_n(\tau) \equiv \mathbf{w}(t_n + \tau)$. By uhc of Ξ at $(0,\sigma)$ for all σ (see Claim 1 in Part

1) and the facts that $\boldsymbol{\gamma}_n(\tau) \to 0$ and $\mathbf{w}_n(\tau) \to \boldsymbol{w}^*(\tau)$, it follows that, for any $\varepsilon > 0$, there exists N_ε such that, for all $n \geq N_\varepsilon$, $\Xi(\boldsymbol{\gamma}_n(\tau), \mathbf{w}_n(\tau)) \subseteq \Xi^\varepsilon(0, \mathbf{w}^*(\tau))$, where $\Xi^\varepsilon(0, \mathbf{w}^*(\tau)) \equiv \{y' \in \mathbb{S} : \|y' - y\| \leq \varepsilon, y \in \Xi(0, \mathbf{w}^*(\tau))\}$. Recall that $\mathbf{v}_n(\tau) \in \Xi(\gamma_n(\tau), \mathbf{w}_n(\tau))$ for all n; therefore, $\bar{\mathbf{v}}_n(\tau) \in Co(\Xi^\varepsilon(0, \mathbf{w}^*(\tau)))$ for all $n \geq N_\varepsilon$. Since $Co(\Xi^\varepsilon(0, \mathbf{w}^*(\tau)))$ is closed and $\lim_{j\to\infty} \bar{\boldsymbol{v}}_n(\tau) = \boldsymbol{v}^*(\tau)$, it follows that $\boldsymbol{v}^*(\tau) \in Co(\Xi^\varepsilon(0, \mathbf{w}^*(\tau)))$. Since this is true for all $\varepsilon > 0$, it follows that $\boldsymbol{v}^*(\tau) \in Co(\Xi(0, \mathbf{w}^*(\tau)))$.