Asymptotic Behavior of Bayesian Learners with Misspecified Models [*]

Ignacio Esponda   Demian Pouzo   Yuichi Yamamoto

(UC Santa Barbara)   (UC Berkeley)   (Hitotsubashi Univ.)

April 27, 2021

Abstract

We consider an agent who represents uncertainty about the 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. By first showing 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 differential inclusion that describes the evolution of the frequency of actions. We then present a series of implications that can be readily applied to economic applications, thus providing off-the-shelf tools that can be used to characterize behavior under misspecified learning.

Contents

1 Introduction

Over the last few decades, evidence of systematic mistakes and biases in beliefs has been collected in a large range of economic environments. Moreover, the evidence indicates that many of these mistakes persist with experience.[1] One approach to incorporating these findings in our theories is to simply postulate that economic agents have fixed, wrong beliefs about aspects of their environment, and never learn about these aspects. A different approach that has gained popularity over the last few years is to postulate that agents do learn about their environment, but they do so in the context of a misspecified model that misses some important aspects of reality. The researcher who follows this approach is forced to specify the agent's misspecification, and the direction of biases is often not ex-ante obvious without further analysis.

Examples of misspecified learning in economics date back to the 1970s: A firm estimates demand but wrongly excludes competitors' prices (Arrow and Green (1973), Kirman (1975)); a teacher assesses how praise and criticism affect student performance, but does not understand regression to the mean (Tversky and Kahneman (1973), Esponda and Pouzo (2016)); a person faces an increasing marginal income tax rate but behaves as if facing a constant marginal tax (Sobel (1984), Liebman and Zeckhauser (2004), Esponda and Pouzo (2016)); when learning the value of assets, policies, or investment projects, traders, voters, and investors fail to account for sample selection (Esponda (2008), Esponda and Pouzo (2017, 2019a), Jehiel (2018)); a seller estimates a constant-elasticity demand function, but elasticity is not constant (Nyarko (1991), Fudenberg, Romanyuk and Strack (2017)); a person inverts causal relationships and incorrectly believes that dieting affects health (Spiegler (2016)); overconfidence biases an agent's learning of a fundamental (Heidhues, Kőszegi and Strack (2018a)).

In these examples, the agent processes information through the lens of a simple model that misses some aspect of reality. The direction of the bias is often not obvious because the agent's behavior affects the feedback she observes, this feedback is in turn processed via the agent's misspecified model, and this processing leads to updated beliefs and subsequent changes in behavior, which in turn lead to changes in beliefs, and so on.

Despite these examples, we have not yet fully understood how model misspecification affects long-run learning outcomes. Indeed, most existing papers consider somewhat specialized setups, and we do not know whether the learning process converges beyond these particular cases. This paper develops a unified theory on Bayesian learning with model misspecification, which hopefully shapes our understanding of why different models in the literature lead

---

[1]For discussions of the evidence, see, for example, Camerer and Johnson (1997) and Section 3.D in Rabin (1998)

to different conclusions and allows behavior to be characterized in a wider range of settings.

We consider the following environment, which includes the examples above and many other situations of interest. Time is discrete and there is a single, infinitely-lived agent who discounts the future and must take an action in each period. The action potentially affects the distribution of an observable variable, which we call a consequence. The agent's per-period payoff depends on her action and the realized consequence. The true distribution over consequences as a function of an action $x \in X$ is given by $Q(\cdot \mid x) \in \Delta(Y)$, where Y is the set of consequences. The agent, however, does not know Q. She has a parametric model of it, given by $(Q_\theta(\cdot \mid x))_{x \in X}$, 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 the true distribution Q, and it is correctly specified otherwise.[2]

Our key point of departure from previous literature is that we begin by focusing 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 Esponda and Pouzo (2016), we show that eventually the posterior at time t roughly concentrates on the set of parameter values that minimize Kullback-Leibler divergence given the frequency of actions up to time t. This result allows us to write the evolution of the action frequency 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.[3] Finally, we present a series of implications that can be readily applied to economic applications. For the special case of one-dimensional models that are identified–a

---

[2]The correctly-specified version of this environment was originally studied by Easley and Kiefer (1988) and Aghion, Bolton, Harris and Jullien (1991). We focus on the case where Θ is finite dimensional because, in the infinite dimensional case, Bayesian updating need not converge to the truth for most priors and parameter values even if the model is correctly specified (Freedman (1963), Diaconis and Freedman (1986)).

[3]The generalization, called a differential inclusion, allows the derivative to take multiple values, and it has proven useful in previous work in economics (e.g., Gilboa and Matsui (1991)). In our environment, multiplicity arises whenever multiple actions are optimal for the agent, and the evolution of beliefs and subsequent actions depends on which action is followed. Multiple optimal actions arise naturally when the agent is indifferent given her belief, but they can also arise in a misspecified setting due to multiplicity of beliefs.

case that includes many of the applications in the literature–our results imply that convergence and stability can be characterized by examining a simple, two-dimensional figure.

Our results pertain to the agent's long-run behavior and benefit from previous work on misspecified learning.[4] Our environment is the single-agent version of the environment studied by Esponda and Pouzo, 2016; henceforth EP2016. They introduce the notion of Berk-Nash equilibrium and show that, under some conditions, if behavior converges then it must converge to a Berk-Nash equilibrium.[5] But they do not study convergence in general.[6]

We know from Nyarko (1991)'s example that the agent's action need not converge if she learns using a misspecified model, but until recently we knew little about convergence in general. Over the last few years, there has been substantial progress. The first papers to make general progress restricted attention to specific environments. Fudenberg, Romanyuk and Strack (2017) consider a model where the agent has a finite number of actions but still updates between two possible models (i.e., Θ has two 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 one-dimensional stochastic differential equation. Heidhues, Kőszegi and Strack (2018a) 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. Heidhues, Koszegi and Strack (2018b) consider a setting where action spaces are continuous, the state has a unidirectional effect on output, and the prior and noise are normal. These assumptions imply that the posterior admits a one-dimensional summary statistic, to which they apply tools from stochastic approximation theory to establish convergence. He (2018) establishes convergence results in an environment where agents suffer from the gambler's fallacy and mislearn from endogenously censored data.

Two recent papers have made progress in more general environments. Frick, Iijima and Ishii (2019b) provide conditions for convergence of the agent's beliefs when the set of models Θ is finite, but they do not explicitly model actions. They also identify environments where vanishing amounts of misspecification can lead to extreme failures of learning the truth. Fu-

---

[4]We focus on the systematic patterns that tend to arise as time goes by, as opposed to initial behavior which tends to be more dependent on random draws.

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

[6]EP2016 tackle the issue of convergence in Theorem 3, where they use an idea from Fudenberg and Kreps (1993) to show that, if agents are allowed to make possibly large but vanishing mistakes, then behavior can converge to any equilibrium. Here, as in the rest of the literature on misspecified learning, we consider the case where agents don't make these types of mistakes.

denberg, Lanzani and Strack (2020) focus on convergence of actions. They show that if the agent's action converges, then it converges to what they call a uniform Berk-Nash equilibrium, which is a refinement of (pure-action) Berk-Nash equilibrium. They also establish partial converses; for example, all uniformly *strict* Berk-Nash equilibria have an arbitrarily high probability of being the long-run outcome for some initial beliefs. They do not, however, establish general conditions under which the action converges.

These recent results provide a much more complete picture of action and belief convergence under model misspecification. By focusing on the frequency of actions, rather than the action itself or the belief, we are able to obtain a general asymptotic characterization of the agent's behavior, whether it converges or not. As we show, there are examples where neither the action nor the belief of the agent converges, but the frequency of actions does converge (to a mixed-action equilibrium). In particular, this result provides a new interpretation of a mixed-action equilibrium. We also present an example where not even the action frequency converges, but we can still characterize asymptotic behavior.

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 (1993)'s 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. We focus on the problem of a single agent for concreteness, though our tools can be applied to games by assuming that players believe that other players follow stationary strategies, as in Fudenberg and Kreps (1993) and EP2016.

Misspecified learning has also been studied in other environments. Rabin and Vayanos (2010) study a case where shocks are i.i.d. but agents believe them to be autoregressive. Esponda and Pouzo (2019b) 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 class of macroeconomic models where agents learn with misspecified models. Bohren and Hauser (2018) and Frick, Iijima and Ishii (2019a) characterize asymptotic behavior in social learning environments with model misspecification.[7] Finally, Frick, Iijima and Ishii (2019b) focus on convergence and robustness of the stability of equilibrium in both single-agent and social learning environments. Our results can probably be directly extended to some of these setting (such as Markov decision processes)

---

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

4

but perhaps not others (such as social learning, where, unlike our environment, the exact order of action sequences matters for belief updating).

Finally, we take the misspecification as a primitive and establish results for any fixed misspecified model. For work that could help understand which types of misspecified models agents are more likely to use, see, e.g., Aragones et al. (2005), Al-Najjar (2009), Al-Najjar and Pai (2013), Schwartzstein (2014), Olea et al. (2019), and Fudenberg and Lanzani (2020).

We present some motivating examples in Section 2 and introduce the model in Section 3. We characterize asymptotic beliefs in Section 4 and asymptotic behavior in Section 5, and present implications relevant to economic applications in Sections 6 and 7. Finally, we relate our findings to the notion of a Berk-Nash equilibrium in Section 8.

2 Examples

The main contribution of this paper is to develop new tools to study learning in misspecified settings. While we do not focus on any particular application, the tools we provide can be useful in a wide range of environments. To illustrate, we discuss two examples where our approach allows us to make progress in areas that were previously outside the scope of analysis.

2.1 Cyclic behavior

Cyclic behavior is prevalent in human behavior, such as in addiction, dieting, and interpersonal relations, where behavior fluctuates between different actions, such as substance abuse and abstention. In economics, cyclic behavior typically appears as a response to changing endogenous variables, such the real business cycle or periods of high inflation followed by low inflation. It might appear implausible, however, to explain cyclical behavior in single-agent decision environments that remain unchanged, at least using the standard expected utility framework, and without incorporating additional, fluctuating variables.

But, in fact, we show that a standard belief-based theory of behavior can explain cycles provided that the agent learns with a misspecified model. To illustrate, consider Spiegler's (2016) dieter's dilemma.[8] An agent decides in each period whether or not to follow a specific diet, in this case to drink a leafy green juice. The agent believes that the green juice (G) po-

---

[8]Spiegler (2016) interprets a mixed action equilibrium to be the proportion of subjects who take each action in the steady state of a dynamic environment where a sequence of short-lived agents face this problem and learn from a database that contains past realizations. In contrast, we consider the case of a single agent and directly tackle the cyclic behavior of this one agent who learns from her own past experience.

tentially affects her blood pressure (B), which in turn affects her propensity to have headaches (H). We can conveniently depict the agent's model using the graph $G \to B \to H$.[9] The agent learns the strength of these relationships over time from experience. The truth, however, is that there is no connection between drinking the green juice and headaches; instead, both the green juice and headaches affect blood pressure: $G \to B \leftarrow H$. More specifically, suppose that blood pressure is low if and only if the agent drinks green juice or has no headache.

Suppose that the agent's behavior were to converge to drinking the green juice every period. Then her blood pressure would remain low irrespective of her health, and the probability of headaches conditional on low blood pressure would equal the unconditional probability. Due to model misspecification, the agent will incorrectly associate this unconditional probability with the causal effect of blood pressure on headaches, and, if this probability is low relative to the cost of drinking the green juice every day, she will decide to stop drinking it.

Now suppose this is the case and the agent stops drinking the juice. If her behavior were to converge to never drinking the juice, then her blood pressure would be determined only by whether or not she has a headache. In particular, the probability that she has a headache conditional on high blood pressure would be equal to one, so she would believe that there is a perfect association between blood pressure and headaches, and, if the cost of drinking the green juice is not too high, she would decide it is worth drinking it to lower her blood pressure.

In this example, using existing tools, all we can say is that the agent's action cycles forever between drinking and not drinking the juice. In this paper, we will develop tools that allow us to characterize the frequency of times that the agent spends cycling between different actions, which also allows us to determine, for example, the agent's limiting welfare. Interestingly, cycling can occur both in cases where the agent has a coarse model (for example, she believes that the effect of blood pressure on headaches is either high or low) or a finer model where any strength of the relationship is a priori conceivable. With a coarse model, the agent's belief will also diverge, and there will be periods of time where she will be almost convinced of a strong causal relationship and decide to drink the juice, and also periods where she will be almost convinced of a weak relationship and not drink the juice. With a finer model, the agent will eventually be close to indifferent between drinking or not drinking the green juice, but asymptotic behavior will still be characterized by cycles.

In Example 1, we formally study the dieter's dilemma, which belongs to a family of misspecified learning examples where actions are negatively reinforcing: The more an agent chooses a specific action, the more the evidence (interpreted through her misspecified model)

---

[9]This corresponds to a directed acyclic graph (DAG), a concept used by Spiegler (2016) to study model misspecification in economics. In this paper, we do not rely on DAGs, but use it here for pedagogical convenience.

indicates that it was not a good action to take.

2.2 Misdirected learning

The tools we develop can also be applied to generalize conclusions in previous work and to highlight the economic conditions that drive a result as opposed to technical, non-essential restrictions. To illustrate, we consider the model of Heidhues, Kőszegi and Strack (2018a) (henceforth, HKS). A single agent chooses an action in each period to maximize output. Output depends on his action, his own ability, and an external fundamental. The agent does not know the fundamental and will over time use output observations to update his beliefs about it. HKS study this learning problem under the assumption that the agent is overconfident about his own ability, and they show how overconfidence systematically biases the agent away from the correct belief and towards lower output.

Formally, output is given by $y = Q(\theta, x, a) + \varepsilon$, where x is the agent's action (for example, effort), $\theta \in [\bar{\theta}, \bar{\theta}] \subset \mathbb{R}$ is an unknown fundamental, $a \in \mathbb{R}$ is the agent's ability, and ε is a random noise, which follows a log-concave distribution.[10] The agent is overconfident about his ability a, and thinks that output is given by $y = Q(\theta, x, A) + \varepsilon$, where $A > a$. As discussed by HKS, this model captures several applications, including delegation, control in organizations, assertiveness vs. deference in relationships, and public policy choices.

HKS focus on the case where Q has increasing differences in both $(-x, a)$ and (x, θ), so that the incremental gain from choosing a higher action is *lower* when ability a is higher and it is *higher* when the fundamental θ is higher. HKS show that this assumption leads to misdirected learning: Even an agent who starts on average with correct beliefs about the fundamentals becomes too pessimistic about it. As the agent changes his action in response to this pessimism, he lowers outcomes and therefore becomes even more pessimistic about the fundamental, a process that is perpetuated over time.

To reach this conclusion, HKS make a number of other assumptions, including that there is a unique equilibrium belief (this is guaranteed by restrictions on the Q function) and that the agent is myopic. They are able to extend this result to the non-myopic case by further assuming that Q is linear in the fundamental θ. Their intuition, however, suggests that the result is more general, and, as we discuss in Section 7, the tools developed in this paper can be used to show that these additional assumptions are not required. Moreover, for the case of an underconfident agent, i.e., $A < a$, they are able to show that misinference regarding the fundamental is self-

---

[10]To apply our results, we assume that the agent chooses an action from a finite set of actions, while HKS assume that the agent chooses an action from a bounded real-valued interval.

correcting in the steady state, but their methods for establishing convergence do not apply. As we discuss in Section 7, we can show that the agent's belief indeed converges to the steady state provided that an identification property holds. This convergence result is more generally true in a large set of environments where the agent's parameterized model is one-dimensional.

3 The environment

Objective environment. A single agent faces the following infinitely repeated problem. Each period $t = 1, 2, ...$, the agent chooses an action from a finite set X.[11] She then receives a consequence according to the consequence function $Q : X \to \Delta Y$, where Y is the set of consequences and ΔY is the set of all (Borel) probability measures over it. Finally, the payoff function $\pi : X \times Y \to \mathbb{R}$ determines the agent's current payoff. In particular, if $x_t \in X$ is the agent's action at time t, then $y_t \in Y$ is drawn according to the probability measure $Q(\cdot \mid x_t) \in \Delta Y$, and the agent's payoff at time t is $\pi(x_t, y_t)$.

Assumption 1. *(i) Y is a Borel subset of Euclidean space; (ii) There exists a Borel probability measure $v \in \Delta Y$ such that, for all $x \in X$, $Q(\cdot|x) \ll v$, i.e., $Q(\cdot|x)$ is absolutely continuous with respect to v (an implication is the existence of densities $q(\cdot \mid x) \in L^1(Y, \mathbb{R}, v)$ such that $\int_A q(y \mid x) v(dy) = Q(A|x)$ for any $A \subseteq Y$ Borel).*[12]

Assumption 1 collects standard technical conditions and allows for either a finite or nonfinite space of consequences Y: It includes both the case where the consequence is a continuous variable (v is the Lebesgue measure and $q(\cdot \mid x)$ is the density function) and the case where it is discrete (v is the counting measure and $q(\cdot \mid x)$ is the probability mass function).

If the agent knew the primitives and wished to maximize discounted expected utility, she would choose an action in each period from the set of actions that maximizes

$$\int_Y \pi(x, y) Q(dy \mid x) = \int_Y \pi(x, y) q(y|x) v(dy).$$

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

Subjective family of models. The agent is endowed with a parametric family of consequence functions, $\mathcal{Q}_\Theta = \{Q_\theta : \theta \in \Theta\}$, where each $Q_\theta : X \to \Delta Y$ is indexed by a *model*

---

[11]The assumption of a finite action space is important and it seems challenging to extend many of our results to a continuum of actions.

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

$\theta \in \Theta$. We refer to $\mathscr{Q}_\Theta$ as the family of models and say that it is *correctly specified* if $Q \in \mathscr{Q}_\Theta$ and *misspecified* otherwise.

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

Example 1. Consider the dieter's dilemma discussed in Section 2. Recall that the true model is given by $G \to B \leftarrow H$, where all variables are binary and $G = 1$ means 'drink the green juice', $B = 1$ means 'high blood pressure', and $H = 1$ means 'headache'. The payoff function is $\pi(H, G) = (1 - H) - CG$, where $C > 0$ is the cost of drinking the green juice. Let q be the probability that the agent has a headache, and recall that $B = 0$ if and only if either $G = 1$ or $H = 0$. A consequence is given by $y = (H, B)$ and the action is binary, $x = G \in \{0, 1\}$. The consequence function $Q(y \mid G)$ is given by the left panel of Figure 1. In particular, the probability of a headache is the same irrespective of G, and since drinking the juice is costly, $C > 0$, it is optimal to not drink the juice, $G = 0$.

In contrast, the agent believes that $G \to B \to H$. Let θ_j^H be the agent's perceived probability that she has a headache, $H = 1$, whenever her blood pressure is $B = j$. In addition, let θ^B be the agent's perceived probability that her blood pressure is high, $B = 1$, if she does not drink the juice. For simplicity, we assume that the agent knows that if she drinks the juice then her blood pressure is low, $B = 0$, for sure. Then a model is given by $\theta = (\theta_0^H, \theta_1^H, \theta^B)$, where $\Theta = \{\theta \in [0, 1]^3 : \theta_0^H \leq \theta_1^H\}$ is the set of models with the restriction that the perceived probability of a headache is nondecreasing in blood pressure.[13] The consequence function $Q_\theta(y \mid G)$ is given by the right panel of Figure 1. For example, the probability that $H = 1$ and $B = 1$ given $G = 0$ is given by the probability that $H = 1$ given $B = 1$, which is θ_1^H, times

---

[13] Technically, this set of models does not satisfy Assumption 2(iii), but we can easily replace it with the the set $\Theta^\varepsilon = \{\theta \in \Theta : \theta_0^H \geq \varepsilon, \theta^B \in [\varepsilon, 1 - \varepsilon]\}$, which does satisfy the assumption and yields the same solution for all sufficiently small $\varepsilon > 0$.

Figure 1: Dieter's dilemma consequence functions. The left panel described the true consequence function, Q, while the right panel describes the consequence function parameterized by θ.

the probability that $B = 1$ given $G = 0$, which is θ^B. Clearly, the agent's family of models is misspecified. $\square$

Bayesian learning. The agent is Bayesian and starts with a prior μ_0 over the space of models 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.[14] 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 $f : \Delta\Theta \to X$ specifying the action $f(\mu) \in X$ that the agent takes at any moment in time in which her belief is μ.[15] A *history* is a sequence $h = (x_0, y_0, ..., x_t, y_t, ...) \in \mathbb{H} \equiv (X \times Y)^\infty$. Together with the primitives of the problem, a policy f induces a probability distribution over the set of histories, which we will denote by P^f.

Policy correspondence. A *policy correspondence* is a mapping $F : \Delta\Theta \rightrightarrows X$, where $F(\mu) \subseteq X$ denotes the set of actions that the agent might choose any time her belief is $\mu \in \Delta\Theta$. We sometimes abuse notation and, for a set of probability measures $A \subseteq \Delta\Theta$, we let $F(A)$ represent the set of actions x such that $x \in F(\mu)$ for some $\mu \in A$. 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.

---

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

[15]We do not allow the agent to mix to simplify the exposition and to highlight the fact that a mixed distribution over actions may describe limiting behavior despite the fact that the agent never actually mixes. In the more general case where f maps into ΔX, our main result (Theorem 2) holds exactly as stated but some of the statements in Section 6 need to be modified accordingly.

10

Assumption 4. *The policy correspondence F is upper hemi-continuous (uhc).*[16]

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

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

where $\mu' = B(x,y,\mu)$ is the Bayesian posterior, and $\bar{Q}_\mu \equiv \int_\Theta Q_\theta \mu(d\theta)$. In this case, the correspondence mapping beliefs to optimal actions is uhc provided that $\mu \mapsto \int_Y \pi(x,y) \bar{Q}_\mu(dy|x)$ is continuous and bounded.[17]

In this paper, we characterize behavior for a given set of policies and not for a single policy. The reason is that our method works for behavior that is continuous in beliefs, and this condition is formally achieved by assuming the policy correspondence to be uhc. Continuity is important because we approximate the agent's belief, and we need to know what actions can be chosen at nearby beliefs, and continuity allows us to do so. A single policy, on the other hand, may not be continuous (e.g. the optimal policy is discontinuous at beliefs where the agent is indifferent between multiple actions). Note that expanding beyond a single policy to a policy correspondence is not much of a limitation in most applications since the objective is often to characterize behavior for *all* policies that share a particular characteristic (e.g., the set of all optimal policies).[18]

Action frequency. Our main objective is to study regularities in asymptotic behavior. Previous work has focused on characterizing the limit of the sequence of actions, whenever it exists. 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 action *frequency*. 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, limiting average payoffs (welfare). Second, as we will show, asymptotic beliefs depend crucially on the action frequency. Because actions in turn depend on beliefs, future actions depend crucially on the frequency of past actions.

---

[16]As usual, $\Delta(\Theta)$ is endowed with the weak topology.

[17]Given our assumption that $\theta \mapsto q_\theta$ is continuous, a sufficient condition is that, for all $x \in X$, there exists $h_x \in L^1(Y, \mathbb{R}, \nu)$ such that, for all $\theta \in \Theta$, $|\pi(x, \cdot) q_\theta(\cdot \,|\, x)| \leq h(\cdot)$ a.s.-$Q(\cdot \,|\, x)$. For example, this is satisfied if Y is compact and $\pi(x, \cdot)$ is continuous.

[18]An alternative approach would be to focus on a single, *continuous* policy function (for example, a smooth approximation of the optimal correspondence, where mixed actions are feasible). Theorem 2 holds exactly as stated under this alternative approach since all that it requires about behavior is that it be continuous in beliefs.

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

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

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

4 Asymptotic characterization of beliefs

In this section, we take as given the sequence of action frequencies, $(\sigma_t)_t$, 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.[19]

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

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

The *set of closest models 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. *(i) $(\theta, \sigma) \mapsto K(\theta, \sigma) - K^*(\sigma)$ is continuous; (ii) $\Theta(\cdot)$ is uhc, nonempty-, and compact-valued.*

Proof. The proof of this lemma and all other results in the paper appear in the appendix. $\square$

If the actions were drawn from an i.i.d. distribution $\sigma \in \Delta X$, we could apply Berk's (1966) result to conclude that the posterior eventually concentrates on the set of closest models given σ (i.e., for all open sets $U \supseteq \Theta(\sigma)$, $\lim_{t \to \infty} \mu_t(U) = 1$ P^f-a.s.).[20] EP2016 showed that this conclusion extends to non-i.i.d. actions, provided that the distribution over actions at time t

---

[19]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 $X \times Y$.

[20]See also Bunke and Milhaud (1998). Relatedly, White (1982) shows that the Kullback-Leibler divergence characterizes the limiting behavior of the maximum quasi-likelihood estimator.

converges to a distribution σ. This result is useful to characterize behavior under the assumption that it stabilizes, but it is insufficient to determine whether or not behavior stabilizes.

We next provide a characterization of beliefs without assuming 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

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

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 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,* $\lim_{t\to\infty}\sup_{\theta\in\Theta}|L_t(\theta) - K(\theta, \sigma_t)| = 0$ P^f-a.s.

The next step is to replace $L_t(\cdot)$ in (1) with $K(\cdot, \sigma_t)$. By Lemma 2, for sufficiently large t, we obtain

$$\mu_{t+1}(A) \approx \frac{\int_A e^{-tK(\theta,\sigma_t)}\mu_0(d\theta)}{\int_\Theta e^{-tK(\theta,\sigma_t)}\mu_0(d\theta)}. \tag{2}$$

As $t \to \infty$, the posterior concentrates on models 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 elements, θ_1 and θ_2. In this case, (2) becomes

$$\mu_{t+1}(\theta_1) \approx 1/(1 + \frac{\mu_0(\theta_2)e^{-tK(\theta_2,\sigma_t)}}{\mu_0(\theta_1)e^{-tK(\theta_1,\sigma_t)}}). \tag{3}$$

Suppose, for example, that $(\sigma_t)_t$ converges to σ and that KLD is uniquely 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 (3) that $\mu_{t+1}(\theta_1)$ converges to 1, so the posterior concentrates on the model 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 $t \to \infty$. 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 models 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.

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

$$\lim_{t \to \infty} \int_\Theta (K(\theta, \sigma_t) - K^*(\sigma_t))\mu_{t+1}(d\theta) = 0 \quad P^f\text{-a.s.} \tag{4}$$

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 a vanishing approximation error, and not on μ_t.

5 Asymptotic characterization of action frequencies

In this section, we propose a method to study the asymptotic behavior of the action frequency. 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 action frequency. We then use tools from stochastic approximation developed by Benaïm, Hofbauer and Sorin (2005) (henceforth, BHS2005) to characterize the solutions of this difference equation in terms of the solution to a generalization of a differential equation.

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

$$\sigma_{t+1} = \sigma_t + \frac{1}{t+1} \left(\mathbf{1}(x_{t+1}) - \sigma_t \right), \tag{5}$$

where $\mathbf{1}(x_{t+1}) = (1_x(x_{t+1}))_{x \in X}$ and $1_x(x_{t+1})$ 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

$$\sigma_{t+1} = \sigma_t + \frac{1}{t+1} \left(E\left[\mathbf{1}(x_{t+1}) \mid \mu_{t+1}\right] - \sigma_t \right) + \frac{1}{t+1} \left(\underbrace{\mathbf{1}(x_{t+1}) - E\left[\mathbf{1}(x_{t+1}) \mid \mu_{t+1}\right]}_{=0} \right). \tag{6}$$

14

The last term in equation (6) is exactly equal to zero because the agent chooses pure actions.[21] The reason it is hard to characterize $(\sigma_t)_t$ using (6) 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 4 is useful. This theorem will allow us to approximate μ_{t+1} with a set of probability measures that depends on σ_t.

The objective is not to approximate μ_{t+1} but rather the conditional expectation $E\left[\mathbf{1}(x_{t+1}) \mid \mu_{t+1}\right]$ in equation (6). The conditional expectation, however, is typically discontinuous in the belief (e.g., if 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. We tackle this discontinuity issue by replacing the function $\mu \mapsto E\left[\mathbf{1}(x_{t+1}) \mid \mu\right]$ 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 (6) as a particular case of the following stochastic difference *inclusion*:

$$\sigma_{t+1} = \sigma_t + \frac{1}{t+1}\left(r_{t+1} - \sigma_t\right), \tag{7}$$

where $r_{t+1} \in \Delta F(\mu_{t+1})$. It is called a difference inclusion because r_{t+1} can take multiple values. 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, equation (7) 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 convenient to work with its continuous-time interpolation. Because of the multiplicity inherent in equation (7), we apply the specific methods developed by BHS2005, who extend Benaim (1996)'s ordinary-differential equation method to the case of differential inclusions.[22]

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]With mixed actions, this last term would be a Martingale difference sequence, essentially adding a noise term that can be controlled asymptotically in a standard manner. Theorem 2 would hold as stated, where now $\Delta F(\mu)$ would be a set of *compound* lotteries, i.e., the set of all distributions over actions $\hat{\sigma}$ that are induced by some compound lottery z chosen from $\Delta F(\mu)$, that is, $\hat{\sigma}(x) = \int_{\sigma \in F(\mu)} z(\sigma)\sigma(x)d\sigma$ for each x.

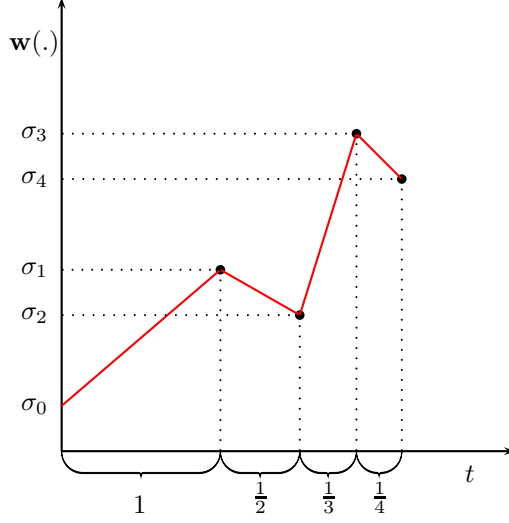[22]See Borkar (2009) for a textbook treatment of the ODE method in stochastic approximation.

Figure 2: Example of a continuous-time interpolation.

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

$$\mathbf{w}(\tau_t + s) = \sigma_t + s\frac{\sigma_{t+1} - \sigma_t}{\tau_{t+1} - \tau_t}, \quad s \in [0, \frac{1}{t+1}). \tag{8}$$

Figure 2 illustrates this simple interpolation for a specific value of $x \in X$. A convenient property of the interpolation is that it preserves the accumulation points of the discrete process.

Equations (7) and (8) 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 $\Delta F(\Delta\Theta(\sigma_t))$. Thus, the derivate approximately takes values in $\Delta F(\Delta\Theta(\sigma_t)) - \sigma_t$. 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 $\dot{\mathbf{w}}(t) \in \Delta F(\Delta\Theta(\mathbf{w}(t))) - \mathbf{w}(t)$. Thus, we can show that the continuous-time interpolation of $(\sigma_t)_t$ is well approximated by solutions to the following differential inclusion:

$$\dot{\sigma}(t) \in \Delta F(\Delta\Theta(\sigma(t))) - \sigma(t). \tag{9}$$

In the special case where the agent is correctly specified and the KLD minimizer is unique for all action frequencies (i.e., $\Theta(\sigma) = \theta^*$ for all σ), the right-hand side of (9) reduces to $\delta_{x^*} - \sigma(t)$, where δ_{x^*} is the degenerate distribution at x^* and x^* is the unique optimal action

16

given the correct model θ^*.[23] In this case, all solutions to the differential inclusion converge to this optimal action δ_{x^*} regardless of the initial prior. However, if the model is misspecified, the KLD minimizer $\Theta(\sigma(t))$ depends on the action frequency $\sigma(t)$, and so does the first-term of the right-hand side of (9). Hence solutions to the differential inclusions show more complicated dynamics in general.

To state the main result formally, we first define what we mean by a solution to the dif-ferential inclusion. A *solution to the differential inclusion (9) with initial point* $\sigma \in \Delta X$ is a mapping $\sigma : \mathbb{R} \to \Delta X$ that is absolutely continuous over compact intervals with the properties that $\sigma(0) = \sigma$ and that (9) is satisfied for almost every t. Let S_σ^T denote the set of solutions to (9) 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 (9); see, e.g., Aubin and Cellina (2012).

We now state the main characterization result.

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.: For all $T > 0$,*

$$\lim_{t \to \infty} \inf_{\sigma \in S_{\mathbf{w}(t)}^T} \sup_{0 \leq s \leq T} \|\mathbf{w}(t+s) - \sigma(s)\| = 0. \tag{10}$$

Theorem 2 says that, for any $T > 0$, the curve $\mathbf{w}(t + \cdot) : [0,T] \to \Delta X$ defined by the continuous-time interpolation of $(\sigma_t)_t$ approximates some solution to the differential inclu-sion (9) with initial condition $\mathbf{w}(t)$ over the interval $[0,T]$ with arbitrary accuracy for suffi-ciently large t. As we will show, this result is convenient because it allows us to characterize asymptotic properties of $(\sigma_t)_t$ by solving the differential inclusion in (9).

BHS2005 refer to a function $\mathbf{w}$ satisfying (10) as an asymptotic pseudotrajectory of the differential inclusion. They show that the limit set of a (bounded) asymptotic pseudotrajectory is internally chain transitive.[24] Thus, one corollary of Theorem 2 is that the frequency of actions converges almost surely to an internally chain transitive set of the differential inclusion. Because the notion of internally chain transitive is fairly complex, in the next two sections we provide a series of results that help characterize behavior in economic applications.

The differential inclusion (9) becomes a differential equation in the special case in which $\Delta F(\Delta \Theta(\sigma))$ is a singleton for all σ. This is not true if F is the correspondence of optimal actions because there are typically beliefs at which multiple actions are optimal. But even if we smoothed the best response function (as in Fudenberg and Kreps, 1993) and turned F into

---

[23]For this particular statement, we assume a unique optimal action given the true parameter value θ^*.

[24]For a definition of an internally chain transitive set, see BHS2005, Section 3.3, Definition VI.

a function, $\Delta\Theta(\sigma)$ could still be multivalued. This is typically the case, for example, in bandit problems, but it is also true when the space of models is coarse.

6 Convergence results

We now present a series of implications of Theorem 2 that can be readily applied to economic applications. In the next two sections, we assume that the agent chooses a policy f that is a selection from F and that Assumptions 1-4 are satisfied. All probabilistic statements are with respect to the corresponding probability measure P^f.

6.1 Equilibrium

We define an equilibrium to be a stationary point of the differential inclusion.

Definition 2. $\sigma \in \Delta X$ is an *equilibrium* if $\sigma \in \Delta F(\Delta\Theta(\sigma))$.

If σ is an equilibrium, then there is a solution of the differential inclusion that starts at σ and forever remains at σ. The next result shows that, if the action frequency converges with positive probability, then it must converge to an equilibrium.

Proposition 1. *Let H^* be the set of all sample paths which satisfy the property stated in Theorem 1; note that $P^f(H^*) = 1$. If there is a sample path $h \in H^*$ such that $\sigma_t(h)$ converges to σ^*, then σ^* must be an equilibrium.*

6.2 Attracting sets and unstable equilibrium

Proposition 1 applies only to the case in which the action frequency converges. It does not tell us what happens when the action frequency does not converge, and also it is not clear when the action frequency converges. Also, even when the action frequency converges, if there are multiple equilibria, the proposition does not tell us which one will arise as a long-run outcome. In this section, we introduce two concepts, *attracting sets* and *unstable equilibria*, which are useful to make better predictions about the asymptotic behavior of the action frequency.

Let $d(\sigma, A)$ denote the distance from a point σ to a set A, that is, let $d(\sigma, A) = \inf_{\tilde{\sigma} \in A} \|\sigma - \tilde{\sigma}\|$. The following definition is standard in the stochastic approximation literature (e.g. BHS2005).

Definition 3. A set $A \subseteq \triangle X$ is *attracting* if there is a set $\mathscr{U}$ such that $A \subset \text{int}\mathscr{U}$ and such that for any $\varepsilon > 0$, there is T such that $d(\boldsymbol{\sigma}(t), A) < \varepsilon$ for any initial value $\boldsymbol{\sigma}(0) \in \mathscr{U}$, for any solution $\boldsymbol{\sigma} \in S^\infty_{\boldsymbol{\sigma}(0)}$ to the differential inclusion, and for any $t > T$.

In this definition, we require uniform convergence, in that as long as the initial value is chosen from $\mathcal{U}$, $\boldsymbol{\sigma}(t)$ is in the ε-neighborhood of A for *all* periods $t > T$. Intuitively, this implies that once $\boldsymbol{\sigma}(t)$ enters the ε-neighborhood of A, it will never leave it. The largest set $\mathcal{U}$ which satisfies the property in this definition is the *basin of attraction* of A, and we will denote it by $\mathcal{U}_A$. A set A is *globally attracting* if it is attracting and its basin of attraction is the whole space $\triangle X$. An equilibrium σ^* is *attracting* if the set $A = \{\sigma^*\}$ is attracting.

The following proposition shows that an attracting set appears as a long-run outcome in some sense. Let E denote the set of all equilibria.

Proposition 2. *The following results hold:*

(i) If A is globally attracting, then the action frequency σ_t approaches this set A almost surely: $\lim_{t\to\infty} d(\sigma_t, A) = 0$. In particular, if A is a globally attracting equilibrium, σ_t converges to that equilibrium almost surely.

(ii) Suppose that there are finitely many attracting sets $(A_1, \cdots, A_N)$ such that $\triangle X$ is the union of the basins $(\mathcal{U}_{A_1}, \cdots, \mathcal{U}_{A_N})$ of these attractors and of the equilibrium set E. Then almost surely, σ_t approaches the equilibrium set E or one of these attractors: $\lim_{t\to\infty} d(\sigma_t, E) = 0$ or $\lim_{t\to\infty} d(\sigma_t, A_n) = 0$ for some n.

The property of uniform convergence required in the definition of attracting sets is crucial to obtain Proposition 2. To see this, let $\boldsymbol{w}(t)$ denote the current action frequency. Theorem 2 implies that the motion of the action frequency in the future is approximated by a solution $\boldsymbol{\sigma} \in S^{\infty}_{\boldsymbol{w}(t)}$ to the differential inclusion for some (long but) *finite time T*; but it does not guarantee that the action frequency $\boldsymbol{w}$ is approximated by $\boldsymbol{\sigma}$ forever. So even if all solutions $\boldsymbol{\sigma} \in S^{\infty}_{\boldsymbol{w}(t)}$ starting from the current value $\boldsymbol{w}(t)$ converge to some equilibrium σ^*, the action frequency $\boldsymbol{w}$ may not converge there.[25] Formally, Theorem 2 implies that, for any T and $\varepsilon > 0$, if t is large enough, then $\|\boldsymbol{w}(t+T) - \boldsymbol{\sigma}(T)\| < \varepsilon$ for some $\boldsymbol{\sigma} \in S^{\infty}_{\boldsymbol{w}(t)}$, so the action frequency $\boldsymbol{w}(t+T)$ in time $t+T$ is close to the equilibrium σ^*. However, after time $t+T$, the action frequency $\boldsymbol{w}$ can be quite different from $\boldsymbol{\sigma}$, and it may move away from the equilibrium σ^*.

This suggests that in order to guarantee convergence to σ^*, we need a stronger assumption, and uniform convergence is precisely the property we want. To see how it works, note that Theorem 2 can be applied iteratively, so that the action frequency $\boldsymbol{w}$ from time $t+T$ to $t+2T$ is approximated by a solution $\boldsymbol{\sigma}' \in S^{\infty}_{\boldsymbol{w}(t+T)}$ starting from $\boldsymbol{w}(t+T)$. As mentioned earlier, this value $\boldsymbol{w}(t+T)$ is close to the equilibrium σ^*. So if the equilibrium σ^* is attracting, then

---

[25]This is the "shadowing" problem in the stochastic approximation literature (e.g., Benaim (1999), section 8).

$w(t+T)$ is in the basin of σ^*, and the solution σ' starting from this point stays around the equilibrium σ^*. This in turn implies that the action frequency $w(t+2T)$ in time $t+2T$ is also close to σ^*. A similar argument shows that $w(t+nT)$ in time $t+nT$ is close to σ^* for every $n = 1, 2, \cdots$. The proof of Proposition 2 generalizes this idea and shows convergence to attracting sets.[26]

Next, we apply the result to an example where the action diverges but the frequency converges.

Example 1 (Dieter's dilemma, continued) Using the consequence functions Q and Q_θ depicted in Figure 1, the KLD function in this example is

$$K(\sigma, \theta) = \sigma(1)\left(q\ln\frac{q}{\theta_0^H} + (1-q)\ln\frac{1-q}{1-\theta_0^H}\right) + \sigma(0)\left(q\ln\frac{q}{\theta_1^H\theta^B} + (1-q)\ln\frac{1-q}{(1-\theta_0^H)(1-\theta^B)}\right),$$

where $\sigma(1)$ is the probability of drinking the juice, i.e., $G = 1$. Assuming $\sigma(1) < 1$, there is a unique minimizer given by $\theta_1^H = 1$, $\theta^B = q$, and $\theta_0^H = q\sigma(1)/(q\sigma(1)+1-q) < q$. Intuitively, if blood pressure is high, $B = 1$, it must be that the agent did not drink the juice, $G = 0$, and that she had a headache, $H = 1$. Therefore, there is a perfect correlation between B and H, thus explaining the belief that the probability of a headache conditional on high blood pressure is one, $\theta_1^H = 1$. In order to fit the true joint probability of $H = 1$ and $B = 1$ when $G = 0$, which is q, the agent believes that the probability of having high blood pressure conditional on $G = 0$ is $\theta^B = q$, since, as mentioned above, she believes that conditional on high blood pressure, she will have a headache for sure. Finally, the agent believes that the probability of a headache conditional on low blood pressure is lower than the unconditional probability of a headache, since low blood pressure is indicative of good health. This is particularly true when $G = 0$, which is why θ_0^H is lowest, and equal to zero, when $\sigma(1) = 0$, and it is increasing in the probability that the agent takes the green juice, $\sigma(1)$.

If, however, $\sigma(1) = 1$, so that the agent takes the green juice with probability one, then there are multiple minimizers of KLD, all satisfying: $\theta_1^H \geq \theta_0^H$, $\theta^B \in [0, 1]$, and $\theta_0^H = q$. Intuitively, there is no information to identify θ_1^H and θ^B (these parameters only enter the second term of the KLD function, but this term is now multiplied by zero). Moreover, because the agent always drinks the juice, then $B = 0$ occurs irrespective of the realization of H, and so the probability of H conditional on $B = 0$ is given by the unconditional probability, q.

Next, we turn to the optimal decision and suppose, for simplicity, that the agent is myopic.

---

[26]Formally, the whole path of w is approximated by a *chain* of trajectories $(\sigma_1, \sigma_2, \cdots)$ where $\|\sigma_n(T) - \sigma_{n+1}(0)\| < \varepsilon$, and uniform convergence ensures that this chain of trajectories converges to σ^*.
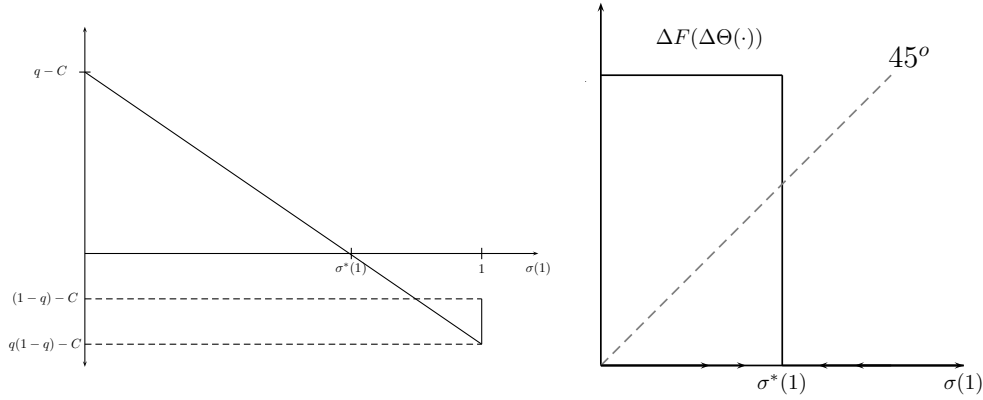
Figure 3: The dieter's dilemma. The left panel plots the difference in expected payoff from dieting vs. not dieting as a function of the probability of dieting. The right hand side plots the mapping $\sigma \mapsto \Delta F(\Delta\Theta(\sigma))$ that characterizes the the asymptotic dynamics of the action frequency.

The expected payoff is $(1 - \theta_0^H) - C$ if $G = 1$ and $(1 - \theta_0^H)(1 - \theta^B) + (1 - \theta_1^H)\theta^B$ if $G = 0$. Using the minimized values of the parameter values discussed above, the difference in expected payoff is a function of $\sigma(1)$ and given by $Diff(\sigma(1)) = q(1-q)/(q\sigma(1)+1-q) - C$ for $\sigma(1) < 1$. This expression is decreasing in $\sigma(1)$. Intuitively, as the probability of drinking the green juice increases, the perceived probability that low blood pressure causes headaches increases, which decreases the relative benefit of drinking green juice in order to lower blood pressure.

In the case $\sigma(1) = 1$, where there are multiple minimizers of KLD, the fact that $\theta_1^H \geq \theta_0^H$ implies that $(1 - \theta_0^H)$ is an upper bound for the expected payoff under $G = 0$. Thus, $Diff(\sigma(1)) \in [-C, 1-q-C]$.

The left hand side of Figure 3 plots the difference in expected payoff between $G = 1$ and $G = 0$ as a function of the probability of $G = 1$, i.e., $Diff(\sigma(1))$, under the assumption that $1-q < C < q$. This assumption says that the cost is large enough that the agent would not find it profitable to take the juice in order to lower the probability of a headache from one to q, but that the cost is small enough that the agent *would* find it profitable to take the juice in order to lower the probability of a headache from q to zero. As shown in the figure, the assumption implies that $Diff(0) > 0$ and $Diff(1) < 0$. Since $Diff(\cdot)$ is decreasing, there is a unique $\sigma^*(1)$, depicted in Figure 3, such that $Diff(\sigma^*(1)) = 0$.

From the expression $Diff$, we can characterize the correspondence $F(\Delta\Theta(\sigma))$. In particular, $F(\Delta\Theta(\sigma)) = \{1\}$ if $\sigma(1) < \sigma^*(1)$, $F(\Delta\Theta(\sigma)) = \{0\}$ if $\sigma > \sigma^*(1)$, and $F(\Delta\Theta(\sigma)) =$

21

$\{0,1\}$ if $\sigma = \sigma^*(1)$. Actions are negatively reinforcing in the sense that doing more of one action makes the agent want to do less of it. This feature can be seen in Figure 3, where we have plotted $\sigma \mapsto \Delta F(\Delta \Theta(\sigma))$. For example, if the agent chooses pure action $G = 0$, i.e., $\sigma(1) = 0$, then the KLD minimizer is such that it is optimal to choose the opposite action, $G = 1$. Similarly, if the agent takes pure action $G = 1$, she then prefers to take action $G = 0$.

This feature of negatively reinforcing actions is present in several examples in the literature, and previous work has shown that the action does not converge in those examples (e.g., Nyarko (1991), Esponda and Pouzo (2016), Fudenberg, Romanyuk and Strack (2017)).[27] We can use our differential inclusion to go beyond this result and show that the action *frequency* does converge. In the example, $\sigma^*(1)$ is the unique equilibrium point: Given $\sigma^*(1)$, the closest model is $\theta^* = (q\sigma^*(1)/(q\sigma^*(1)+1-q), 1, q)$, and, given the belief δ_{θ^*}, the agent is indifferent between each of the actions in the support of σ^*. Moreover, as Figure 3 shows, for any initial condition, the solutions to the differential inclusion converge to $\sigma^*(1)$, and so $\{\sigma^*(1)\}$ is a globally attracting set. Proposition 2(i) implies that the frequency of action $G = 1$ almost surely converges to $\sigma^*(1)$.

In this example, the action diverges, but the action frequency converges to σ^* and the belief converges to a belief degenerate at θ^*. Consider next a modified example with only two models in the support of the prior, $\Theta = \{(\bar{\theta}_0^H, 1, q), (\hat{\theta}_0^H, 1, q)\}$, where we pick $\bar{\theta}_0^H$ and $\hat{\theta}_0^H$ such that $K(\sigma^*, \bar{\theta}_0^H) = K(\sigma^*, \hat{\theta}_0^H)$; in particular, we can think of a low and high perceived value of the probability of a headache conditional on low blood pressure (note that the unrestricted KLD minimizer that we found above is between these two values). Then the frequency of actions converges to the mixed action σ^*, though now the agent's belief does not converge.[28] So the action frequency may converge even if both the action and the belief diverge. $\square$

In the next example, we show that the attracting set need not be an equilibrium.

Example 2. The consequence space is $Y = \mathbb{R}^3$. There are three actions, x_1, x_2, and x_3. Given an action x_k, the consequence y follows the normal distribution $N(e_k, I)$ where $e_k \in \mathbb{R}^3$ is the unit vector whose kth component is one, and I is the identity matrix. However, the agent does not recognize that the action influences the consequence. Formally, the model space is the probability simplex $\Theta = \{\theta = (\theta_1, \theta_2, \theta_3) | \sum_{k=1}^3 \theta_k = 1, \ \theta_k \geq 0 \ \forall k\}$, and for each model $\theta = (\theta_1, \theta_2, \theta_3)$, the agent believes that y follows the normal distribution $N(\theta, I)$. Assume that for each degenerate belief δ_θ, the policy $F(\delta_\theta)$ is given as in Figure 4, where the triangle

---

[27]Negative reinforcement is also present in other examples in Spiegler (2016) as well as in voting (Esponda and Pouzo (2017, 2019a); Esponda and Vespa (2018)) and investment (Jehiel (2018)) environments.

[28]The argument is identical to Berk's (1966) example of a fair coin that is believed to be biased.

represents the model space Θ. For example, if the current belief puts probability one on the model $\theta = e_1$, then the policy F selects the action x_2.
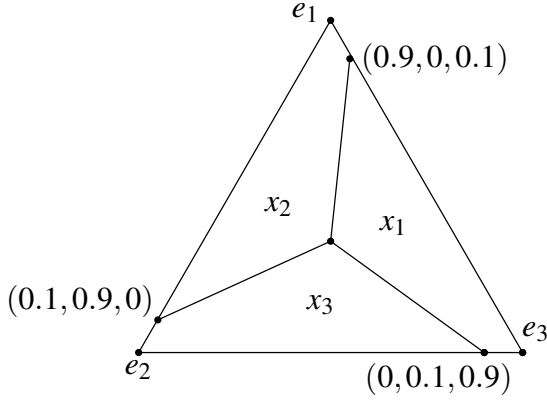

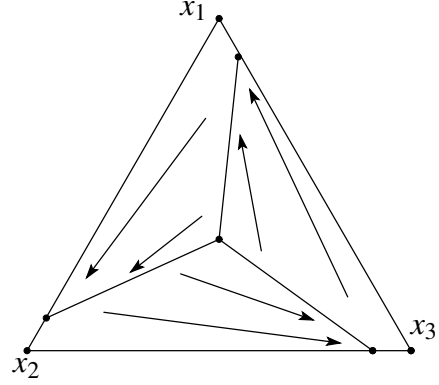Figure 4: Policy $F(\delta_\theta)$ for each model θ Figure 5: Differential Inclusion

Simple algebra shows that, for each mixed action σ, the KLD minimizer is $\theta = \sigma$. Intuitively, the agent's subjective model cannot explain the correlation between the action and the consequence anyway, so the best model is the one which explains the marginal distribution of the consequence y. Accordingly, a solution to the differential inclusion is described as in Figure 5, where the triangle represents the whole action space $\triangle X$ and each arrow points to the corresponding vertex in the large triangle.

This example has a unique equilibrium, $\sigma^* = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. This equilibrium is not attracting. Indeed, starting from any nearby point $\sigma \neq \sigma^*$, a solution to the differential inclusion moves away from the equilibrium, as described in Figure 6.

On the other hand, the cycle described by the arrows in Figure 7 is attracting. The basin of attraction is the whole space $\triangle X$ except the equilibrium point $\sigma^* = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. That is, given any initial value $\sigma \neq \sigma^*$, any solution to the differential inclusion will eventually follow this cycle. (The proof is straightforward and hence omitted.) $\square$

Proposition 2(ii) implies that in Example 2, the action frequency σ_t must converge to the (non-attracting) equilibrium $\sigma = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ or follow the limit cycle described in Figure 7. But which one is more likely to occur? It turns out that the equilibrium σ^* in the example above is unstable, in that the action frequency never converges there. So the action frequency follows the limit cycle almost surely.

To see why the equilibrium σ^* is unstable, suppose that the current action frequency is exactly this equilibrium, i.e., $\sigma_t = \sigma^*$. Suppose also that the agent chooses some action today, say x_1. This changes the action frequency in the next period, and we have $\sigma_{t+1} = \frac{1}{t+1} 1_{x_1} +$
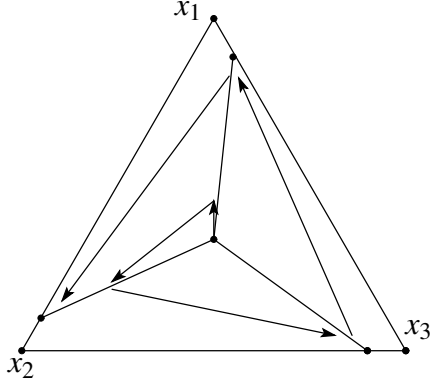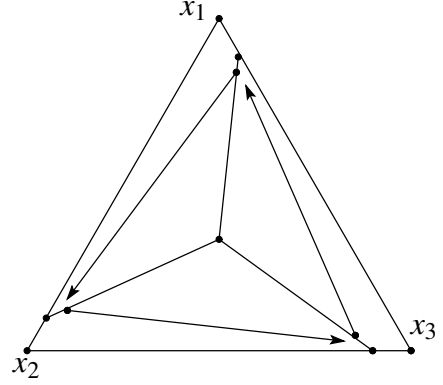
23

Figure 6: Instability of the equilibrium

Figure 7: Limit cycle.

$\frac{t}{t+1}\sigma^*$. Note that this new action frequency is slightly different from the equilibrium σ^*. Then starting from this action frequency, a solution to the differential inclusion moves away from the equilibrium (See Figure 6), which implies instability of σ^*. More formally, this equilibrium σ^* is *unstable* in the following sense:

Definition 4. An equilibrium σ^* is *unstable* if there is a number $T > 0$ and an open neighborhood $\mathscr{U}$ of σ^* such that for any $\sigma \in \mathscr{U}$, for any $x \in F(\triangle\Theta(\sigma^*))$, and for any $\overline{\beta} \in (0,1)$, there is $\beta \in (\overline{\beta},1)$ such that for any $\boldsymbol{\sigma} \in S^{\infty}_{\beta\sigma+(1-\beta)\delta_x}$, we have $\boldsymbol{\sigma}(t) \notin \mathscr{U}$ for some $t \in [0,T]$.

In Example 2, starting from *any* nearby point $\sigma \neq \sigma^*$ of the equilibrium σ^*, the solution to the differential inclusion moves away from the equilibrium σ^*. In such a case, the condition stated in the definition is satisfied, so this equilibrium σ^* is unstable.

But our definition of unstable equilibrium is a bit more general. Roughly, an equilibrium σ^* is unstable if starting from *almost all* nearby points $\sigma \neq \sigma^*$ of the equilibrium σ^*, all the solutions to the differential inclusion eventually leave its neighborhood. This is illustrated in the following example:

Example 3. We add one more action x_3' to Example 2. This new action x_3' is redundant, and is identical to the action x_3. Formally, the signal distribution given the action x_3' is $N(e_3,I)$, and the policy $F(\mu)$ contains x_3' for all μ such that $F(\mu)$ contains x_3 in Example 2. The agent still believes that the action does not influence the signal distribution.

This example has a continuum of equilibria; any mixed action σ with $\sigma(x_1) = \sigma(x_2) = \sigma(x_3) + \sigma(x_3') = \frac{1}{3}$ is an equilibrium. Pick one equilibrium σ^*, and pick an open neighborhood $\mathscr{U}$. This neighborhood $\mathscr{U}$ contains equilibrium points and non-equilibrium points. The set of equilibrium points is continuous, but has measure zero; so almost all the points in $\mathscr{U}$ are

24

non-equilibrium points. Starting from these non-equilibrium points, all the solutions to the differential inclusion leave the neighborhood $\mathscr{U}$, as described in Figure 6. However, starting from the equilibrium points, a solution to the differential inclusion can stay there forever. So $\mathscr{U}$ contains some points from which the solution to the differential inclusion does not leave $\mathscr{U}$. Still, this equilibrium σ^* is unstable. Indeed, given any point $\sigma \in \mathscr{U}$ and given any action x, if we choose β sufficiently close to one, the perturbed point $\beta\sigma + (1-\beta)\delta_x$ is not an equilibrium; so starting from this perturbed point, the solution to the differential inclusion eventually leaves $\mathscr{U}$. $\square$

The following proposition asserts that unstable equilibria do not arise as long-run outcomes.

Proposition 3. *If σ^* is an unstable equilibrium, then the action frequency σ_t converges to σ^* with probability zero.*

6.3 Convergence to attracting sets for some prior

Proposition 2 provides useful conditions under which the action frequency converges to an attracting set, such as the set of equilibria. Moreover, Proposition 3 shows that the frequency cannot converge to unstable equilibria. These propositions, however, do not imply that the action frequency converges to any one specific attracting set or equilibrium (unless it is globally attracting). We will show that if an attracting set A satisfies some additional property, then the action frequency converges to it with positive probability for some initial prior.[29] Throughout this section, let $B_\varepsilon(A)$ denote the ε-neighborhood of A, i.e., the set of all σ such that $d(\sigma, A) < \varepsilon$.

We first introduce the idea of a "perturbed differential inclusion." Given an initial value $\boldsymbol{\sigma}(0)$, let $\boldsymbol{S}_{\boldsymbol{\sigma}(0)}^{\infty,\varepsilon}$ denote the set of all solutions to the following differential inclusion:

$$\dot{\boldsymbol{\sigma}}(t) \in \bigcup_{\tilde{\sigma} \in B_\varepsilon(\boldsymbol{\sigma}(t))} \triangle F(\triangle \Theta(\tilde{\sigma})) - \boldsymbol{\sigma}(t). \tag{11}$$

---

[29]Theorem 7.3 of Benaim (1999) gives a sufficient condition for convergence to an attracting set for some prior. This result, however, relies on a technical assumption ((24) in his paper), which roughly requires that if we take sufficiently large t, then regardless of the past history h^t, the motion of $\mathbf{w}$ in the continuation problem is approximated by a solution to the differential inclusion. This assumption is not satisfied in our model, because there are histories in which the posterior belief μ_{t+1} is not concentrated on $\Theta(\sigma^t)$ (note that this happens when realized signals are very different from ex-ante expectation), and after such histories, the motion of $\mathbf{w}$ could be very different from any solution to the differential inclusion.

Recall that in the original differential inclusion, the agent chooses an action from $F(\triangle\Theta(\boldsymbol{\sigma}(t)))$. In (11), this choice set is expanded, so that the agent chooses an action from $F(\triangle\Theta(\tilde{\sigma}))$, where $\tilde{\sigma}$ is a perturbation of the current action frequency $\boldsymbol{\sigma}(t)$.

Definition 5. A set A is *robustly attracting* if it is attracting and there is $\zeta > 0$ and $\varepsilon > 0$ such that for any initial value $\boldsymbol{\sigma}(0) \in B_\zeta(A)$, any solution $\boldsymbol{\sigma} \in S_{\boldsymbol{\sigma}(0)}^{\infty,\varepsilon}$ to the perturbed differential inclusion never leaves the basin $\mathscr{U}_A$; i.e., $\boldsymbol{\sigma}(t) \in \mathscr{U}_A$ for all $t \geq 0$.

In some special cases, attracting sets and robustly attracting sets are equivalent. For example, as will be explained in Proposition 7, attracting sets are robustly attracting when Θ is the one-dimensional interval $[0,1]$. The same result holds when there are only two actions. (The proof is straightforward and hence omitted.) However, in general, attracting sets need not be robustly attracting.[30]

A sufficient condition for a set A to be robustly attracting is that the (non-perturbed) differential inclusion has a contraction property in a neighborhood of A. Formally, let $V(\boldsymbol{\sigma}) = d(\boldsymbol{\sigma}, A)$, and suppose that there is an open neighborhood U of A such that $(\tilde{\sigma} - \sigma) \cdot \nabla V(\sigma) < 0$ for all $\sigma \in U \setminus A$ and $\tilde{\sigma} \in \triangle F(\triangle\Theta(\sigma))$.[31] Then this A is robustly attracting.[32] Note that this contraction property is satisfied by any strict equilibrium; a pure action δ_x is a *strict equilibrium* if there is an open neighborhood U of δ_x such that $F(\triangle\Theta(\tilde{\sigma})) = \{x\}$ for all $\tilde{\sigma} \in U$. So any strict equilibrium is robustly attracting.

We will show that the action frequency converges to a robustly attracting set with positive probability, at least for some initial prior.

Proposition 4. *For each robustly attracting set A, there is an initial prior μ_0^* with full support such that $\lim_{t\to\infty} d(\sigma_t, A) = 0$ with positive probability.*

The following chain of set inclusions summarizes the relationships between several of the

---

[30]We provide such an example in the Online Appendix.

[31]More generally, A is robustly attracting if there is an open neighborhood U of A and a function $V : U \to \boldsymbol{R}_+$ such that (i) $V(\sigma) = 0$ if and only if $\sigma \in A$, (ii) $(\tilde{\sigma} - \sigma) \cdot \nabla V(\sigma) < 0$ for all $\sigma \in U \setminus A$ and $\tilde{\sigma} \in \triangle F(\triangle\Theta(\sigma))$, and (iii) ∇V is Lipchitz-continuous. Note that condition (ii) here is a bit more demanding than Lyapunov stability, which requires $V(\boldsymbol{\sigma}(t)) < V(\boldsymbol{\sigma}(0))$ for all $\boldsymbol{\sigma}(0)$ and $\boldsymbol{\sigma} \in S_{\boldsymbol{\sigma}(0)}^\infty$.

[32]The proof is provided in the Online Appendix.

concepts considered above:

strict equilibrium $\subset$ contraction property in a neighborhood

$\subset$ robustly attracting singleton set

$\subset$ attracting singleton set

$\subset$ equilibrium.

7   One-dimensional models

In this section, we focus on the special case where the model space is one-dimensional: $\Theta = [0,1]$. This case includes many of the current applications in the literature and it allows us to provide a more powerful characterization of the action frequency and the belief.

We will first discuss the case where the model is identified, in the sense that, for any mixed action σ, there is a unique minimizer of Kullback-Leibler divergence, $\theta(\sigma)$.[33] In this case, our differential inclusion reduces to a one-dimensional problem; this reduction considerably simplifies our analysis, because in general the action frequency σ is multi-dimensional and solving the differential inclusion can be a difficult task. We will show that the belief converges to an equilibrium belief almost surely, and we will provide a simple characterization of attracting/unstable equilibria.

In some applications, identification does not hold for all mixed actions, but it naturally holds for all pure actions. We will show in Section 7.2 how to obtain convergence results in this case provided we make other assumptions about the structure of the environment.

7.1   Identification for all mixed actions

Throughout this subsection, we will impose the following identifiability assumption:

Assumption 5. *The following two conditions hold:*

(i) *For each σ, there is a unique minimizer of $K(\sigma,\theta)$ which we denote by $\theta(\sigma) \in [0,1]$, that is, $\Theta(\sigma) = \{\theta(\sigma)\}$.*

(ii) *For each σ with $\theta(\sigma) \in (0,1)$, we have $\left. \frac{\partial^2 K(\sigma,\theta)}{\partial \theta^2} \right|_{\theta=\theta(\sigma)} > 0.$*

---

[33]This assumption rules out bandit problems as well as cases where the model is coarse, such as the version of Example 1 where Θ contains only two elements.

Part (i) is what we call the identifiability condition, which asserts that for each mixed action σ, there is a unique model which best fits the true world. EP2016 provide a more detailed discussion about this identifiability condition. Note that the best model $\theta(\sigma)$ is continuous in σ, because $\Theta(\sigma)$ is upper hemi-continuous in σ,

Part (ii) requires that whenever $\theta(\sigma)$ is an interior solution (so that the first-order condition is satisfied at the minimum), it satisfies the second-order condition. This assumption is crucial for the strict monotonicity result (Proposition 5(iii)), which is needed to prove instability of unstable models (Proposition 8). But all other results remain true without it.

We will show that the agent's belief converges almost surely under Assumption 5. This result strengthens that of Heidhues et al. (2018b), who show a similar convergence result under the assumptions that (i) the model space Θ is one-dimensional, (ii) both the initial prior and a noise term in the consequence y are normally distributed, (iii) the agent's utility is concave in action x, so that there is a unique optimal action for each belief, and (iv) the mean value of the (both objective and subjective) output is monotone in model θ. Our Assumption 5 is weaker than their assumption (ii), and we drop their assumptions (iii) and (iv) and consider general payoff functions. In particular, in our setup, the agent can be indifferent given some beliefs, and there may be a mixed-action equilibrium. Also, we characterize (robustly) attracting and unstable equilibria in this one-dimensional environment. This allows us to identify which equilibrium is more likely to arise asymptotically, when there are multiple equilibria.

The following proposition shows that the closest model $\theta(\sigma)$ is monotone with respect to the action σ. In the proof, we first show that the KLD function $K(\sigma, \theta)$ has the increasing differences property. Then the result follows from the monotone selection theorem of Topkis (1998) and Edlin and Shannon (1998).

Proposition 5. *Suppose that Assumption 5 holds. Pick any σ and $\tilde{\sigma}$, and for each $\beta \in [0,1]$, let $\sigma_\beta = \beta\sigma + (1-\beta)\tilde{\sigma}$. Then the following results are true:*

(i) *If $\theta(\sigma) = \theta(\tilde{\sigma})$, then $\theta(\sigma_\beta) = \theta(\sigma)$ for all $\beta \in [0,1]$.*

(ii) *If $\theta(\tilde{\sigma}) < \theta(\sigma)$, then $\theta(\sigma_\beta)$ is weakly increasing with respect to β.*

(iii) *If $\theta(\tilde{\sigma}) < \theta(\sigma)$, then $\theta(\sigma_{\beta_1}) < \theta(\sigma_{\beta_2})$ for any β_1 and β_2 such that $\beta_1 < \beta_2$ and $\theta(\sigma_{\beta_1}) \in (0,1)$.*

The monotonicity result above ensures that the motion of the closest model $\theta(\sigma)$ is characterized by a simple, one-dimensional problem. Note that when $\Theta = [0,1]$, the best model $\theta(\sigma)$

can move in only three directions; it can go up, down, or stay the same. In particular, since the motion of the action frequency is approximated by

$$\dot{\boldsymbol{\sigma}} = \sigma - \boldsymbol{\sigma}(t)$$

for some $\sigma \in \triangle F(\delta_{\theta(\boldsymbol{\sigma}(t))})$, the result in Proposition 5 implies that, at each time t,

- $\theta(\boldsymbol{\sigma}(t))$ moves up if $\theta(\sigma) > \theta(\boldsymbol{\sigma}(t))$ for all $\sigma \in \triangle F(\delta_{\theta(\boldsymbol{\sigma}(t))})$.

- $\theta(\boldsymbol{\sigma}(t))$ moves down if $\theta(\sigma) < \theta(\boldsymbol{\sigma}(t))$ for all $\sigma \in \triangle F(\delta_{\theta(\boldsymbol{\sigma}(t))})$.

To better understand the motion of $\theta(\boldsymbol{\sigma}(t))$, consider the following example:

Example 4. The consequence space is $Y = \mathbb{R}$, and the agent has two actions, x_0 and x_1. Given an action x_k, the consequence y follows the normal distribution $N(k, 1)$. The agent does not recognize that the action influences the consequence, and she believes that given a model $\theta \in [0, 1]$, y follows the normal distribution $N(\theta, 1)$ regardless of the chosen action. Consider an upper hemi-continuous policy F which satisfies

$$F(\delta_\theta) = \begin{cases} \{x_0\} & \text{if } \theta \in [0, \frac{1}{3}) \cup (\frac{2}{3}, 1] \\ \{x_1\} & \text{if } \theta \in (\frac{1}{3}, \frac{2}{3}) \\ \{x_0, x_1\} & \text{if } \theta \in \{\frac{1}{3}, \frac{2}{3}\} \end{cases}.$$

Given a mixed action σ, the consequence follows the normal distribution $N(\sigma(x_1), 1)$, so the closest model is $\theta(\sigma) = \sigma(x_1)$. Hence the motion of $\theta(\boldsymbol{\sigma}(t))$ can be described by the arrows in Figure 8: $\theta(\boldsymbol{\sigma}(t))$ will move up in the middle region (i.e., $\theta(\boldsymbol{\sigma}(t)) \in (\frac{1}{3}, \frac{2}{3})$), because the agent chooses the action x_1 and the corresponding model is $\theta(\delta_{x_1}) = 1$. For the other region, $\theta(\boldsymbol{\sigma}(t))$ will move down because the agent chooses the action x_0 and the corresponding model is $\theta(\delta_{x_0}) = 0$. □
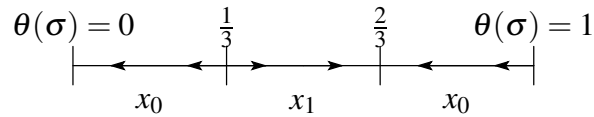

Figure 8: Motion of $\theta(\boldsymbol{\sigma}(t))$

Using the fact that the closest model $\theta(\sigma)$ follows the simple rule above, we will now show that it converges almost surely. This convergence result is not a corollary of Proposition 2 because the equilibrium set is not globally attracting in general. The key feature that we use

29

to establish convergence in this case is that, if the model space is $\Theta = [0,1]$ and $\theta(\sigma)$ follows a recursive rule as in Figure 8, then it cannot be cyclic. This implies that $\theta(\sigma)$ cannot oscillate forever and it must converge.

A model θ^* is an *equilibrium model* if there is an equilibrium σ^* such that $\theta(\sigma^*) = \theta^*$. In Example 4, there are three equilibrium models, 0, 1/3, and 2/3. Let $\Theta^* \subseteq \Theta$ denote the set of all equilibrium models.

Proposition 6. *Suppose that Assumption 5 holds, and that Θ^* is finite. Then almost surely, $\lim_{t\to\infty} \theta(\sigma_t)$ exists and $\lim_{t\to\infty} \theta(\sigma_t) \in \Theta^*$.*

This proposition, together with Theorem 1, implies that the posterior belief μ_t converges almost surely, and the limit belief is a degenerate belief on some equilibrium model.

Example 5. (HKS, 2018) Each period, the agent chooses effort x and observes stochastic output $y = Q(\theta, x, a) + \varepsilon$, where θ is an unknown fundamental, a is the agent's ability, and ε is a random noise, which follows a log-concave distribution. The agent thinks that the signal is given by $y = Q(\theta, x, A) + \varepsilon$, where $A > a$ means the agent is overconfident and $A < a$ means she is underconfident. An implication of Proposition 6 is that beliefs converge provided that the identification condition holds. Identification depends on the shape of the function Q. For example, if Q is linear in θ, then $K(\theta, \delta_x)$ is convex for each pure action x, which in turn implies that $K(\theta, \sigma)$ is convex for all mixed actions σ.[34] As we show in the next subsection, we can relax the identification assumption to study the overconfident case, $A > a$. $\square$

When there are multiple equilibrium models, Proposition 6 does not tell us which one will arise as a long-run outcome. To address this concern, we define *attracting models* as follows.

Definition 6. A model $\theta^* \in [0,1]$ is *attracting* if there is $\varepsilon > 0$ such that

- $\theta(\delta_x) \geq \theta^*$ for any $\theta \in (\theta^* - \varepsilon, \theta^*)$ and for any $x \in F(\delta_\theta)$.

- $\theta(\delta_x) \leq \theta^*$ for any $\theta \in (\theta^*, \theta^* + \varepsilon)$ and for any $x \in F(\delta_\theta)$.

Intuitively, a model θ^* is attracting if it is locally absorbing, in that $\theta(\sigma(t))$ moves toward θ^* in its neighborhood. Indeed, the first bullet point in the definition asserts that if $\theta(\sigma(t))$ is slightly lower than θ^* in the current period t, then it will go up, and hence be closer to θ^* at the next instant. Similarly, the second bullet point in the definition ensures that if $\theta(\sigma(t))$

---

[34]Identification may hold even if $K(\theta, \delta_x)$ is not convex for each pure action x. For example, if there is a function f increasing in θ such that $Q(\theta, x, a) = f(\theta, a)(a + x) - c(x)$, then identification holds, though if, say, $f(\theta, a) = \tan^{-1}(\theta)$, then the KLD function is not convex for pure actions.

is slightly higher than θ^* in the current period t, then it will go down. In Example 4, the equilibrium models 0 and 2/3 are attracting, while 1/3 is not.

Given an attracting model θ^*, let $A = \{\sigma \in \triangle F(\delta_{\theta^*}) | \theta(\sigma) = \theta^*\}$ be the set of equilibria σ in which the agent has a degenerate belief on θ^*.[35] The following proposition shows that this set A is robustly attracting, which means that these equilibria should arise as a long-run outcome for some initial prior. Also the proposition shows that the converse is true, i.e., if a set $A = \{\sigma \in \triangle F(\delta_{\theta^*}) | \theta(\sigma) = \theta^*\}$ is robustly attracting, then θ^* is an attracting model.

Proposition 7. *Under Assumption 5, for each θ^*, the following properties are equivalent:*

(a) *θ^* is attracting.*

(b) *The set $A = \{\sigma \in \triangle F(\delta_{\theta^*}) | \theta(\sigma) = \theta^*\}$ is attracting.*

(c) *The set A is robustly attracting.*

In the same spirit, we define *unstable models* as follows:

Definition 7. A model $\theta^* \in (0,1)$ is *unstable* if $\theta^* \neq \theta(\delta_x)$ for each pure action $x \in F(\delta_{\theta^*})$ and there is $\varepsilon > 0$ such that

- $\theta(\delta_x) \leq \theta^* - \varepsilon$ for any $\theta \in (\theta^* - \varepsilon, \theta^*)$ and for any $x \in F(\delta_\theta)$.

- $\theta(\delta_x) \geq \theta^* + \varepsilon$ for any $\theta \in (\theta^*, \theta^* + \varepsilon)$ and for any $x \in F(\delta_\theta)$.

In words, a model θ^* is unstable if $\theta(\boldsymbol{\sigma}(t))$ moves away from θ^* in its neighborhood. Indeed, the first bullet point implies that if $\theta(\boldsymbol{\sigma}(t))$ is slightly below θ^*, it will move down further at the next instant. The second bullet point implies that if $\theta(\boldsymbol{\sigma}(t))$ is slightly above θ^*, it will go up at the next instant. In Example 4, the equilibrium model $\theta = \frac{1}{3}$ is unstable. In the definition above, we consider only interior models $\theta \in (0,1)$. This is so because whenever an extreme point $\theta = 0, 1$ is supported by some equilibrium (i.e., there is an equilibrium σ such that $\theta(\sigma) = 0$), there is a pure-strategy equilibrium δ_x supporting it.

The following proposition shows that if θ^* is unstable, then any equilibrium in which the agent has a degenerate belief on this model θ^* is unstable; hence these equilibria do not arise as long-run outcomes.

---

[35]Upper hemi-continuity of F ensures that this set A is non-empty, which in turn implies that any attracting model is an equilibrium model. For the special case in which $F(\delta_{\theta^*})$ contains only one component, this set A is a singleton. Similarly, even when $F(\delta_{\theta^*})$ contains only two components, the set A is a singleton for generic parameters. On the other hand, when $F(\delta_{\theta^*})$ contains three or more actions, the set A is typically continuous.

Proposition 8. *Under Assumption 5, $\theta^* \in (0,1)$ is* unstable *if and only if it is not supported by a pure equilibrium, there is at least one mixed equilibrium σ^* with $\theta(\sigma^*) = \theta^*$, and all mixed equilibria σ^* with $\theta(\sigma^*) = \theta^*$ are* unstable.

7.2 Identification for all pure actions

There are economic environments where identification holds for all pure, but not mixed, actions. One prominent example is the environment of Heidhues, Kőszegi and Strack (2018a).

Example 5 (HKS, continued) In Example 5, without specific restrictions on Q, there may exist more than one minimizer for some mixed action, in which case identification fails. But it is nevertheless the case that, given a pure action x, the Kullbuck-Leibler divergence is single-peaked, and hence has a unique minimizer. So identification holds for *pure* actions. If, in addition, we assume that Q has increasing differences in both $(-x, a)$ and (x, θ) and that the agent is overconfident $(A > a)$, then the monotone properties discussed in Section 2 hold, and, as we show next, the action converges to a pure-action equilibrium. $\square$

We will show that, even if identifiability fails, as in the environment of HKS, the belief still converges, provided that the payoffs and the information structure are "monotone" in some sense. Compared to HKS's environment, our monotonicity assumption is more general, we allow multiple (and possibly mixed action) equilibria, and we do not place further restrictions on Q or the discount factor. Specifically, we will consider the following environment.[36]

- $\Theta = [0,1]$. $X = \{x_1, x_2, \cdots, x_N\}$.

- For each pure action x, the KLD function $K(\theta, \delta_x)$ is single-peaked, in the sense that for each x, there is $\theta(x)$ such that $\frac{\partial K(\theta, \delta_x)}{\partial \theta} < 0$ for all $\theta < \theta(x)$ and $\frac{\partial K(\theta, \delta_x)}{\partial \theta} > 0$ for all $\theta > \theta(x)$.

- Higher action induces higher KLD minimizer: $\theta(x_1) < \theta(x_2) < \cdots < \theta(x_N)$.

- Higher beliefs induce higher actions: There are $0 = \theta_0 < \theta_1 < \cdots < \theta_N = 1$ such that $x_n \notin F(\mu)$ for each n and μ such that $[\theta_{n-1}, \theta_n] \cap \text{co}(\text{supp}\,\mu) = \emptyset$, where co$B$ denotes the convex hull of a set B.

In this environment, x_n is a *pure-action equilibrium* if $\theta(x_n) \in [\theta_{n-1}, \theta_n]$.

---

[36] One difference with HKS is that they consider a continuum of actions, while we assume a finite number of actions. The conditions below are also satisfied in other environments (e.g., adverse selection, Esponda (2008)).

Proposition 9. *Let H^* be the set of all sample paths which satisfy the property stated in Theorem 1; note that $P^f(H^*) = 1$. For any sample path $h \in H^*$, the action frequency $\sigma_t(h)$ converges to a degenerate distribution on a pure-action equilibrium.*

This proposition shows that the action frequency converges. If an equilibrium is strict, we can strengthen the result and show that the action converges.[37]

8 Relationship to Berk-Nash equilibrium

In this section, we relate the notion of equilibrium from the differential inclusion approach to EP2016's definition of Berk-Nash equilibrium. To facilitate comparisons, we assume that the agent maximizes discounted expected utility, where F_β is the correspondence of optimal actions and $\beta \in [0,1)$ is the discount factor ($\beta = 0$ is the case of a myopic agent).

The definition of equilibrium that emerges from the differential inclusion approach is that of a probability distribution over actions satisfying $\sigma \in \Delta F_\beta(\Delta\Theta(\sigma)) = \Delta \cup_{\mu \in \Delta\Theta(\sigma)} F_\beta(\mu)$ (see Definition 2). Equivalently, σ is an equilibrium if and only if for every action x in the support of σ there exists a belief $\mu_x \in \Delta\Theta(\sigma)$ such that $x \in F_\beta(\mu_x)$. In contrast, EP2016 define a Berk-Nash equilibrium to be a probability distribution over actions satisfying $\sigma \in \cup_{\mu \in \Delta\Theta(\sigma)} \Delta F_0(\mu)$.[38] Note that σ is a Berk-Nash equilibrium if and only if there exists a belief $\mu \in \Delta\Theta(\sigma)$ such that, for every x in the support of σ, $x \in F_0(\mu)$.

There are two differences between the definition of equilibrium in this paper and a Berk-Nash equilibrium: The latter concept (1) restricts actions to be supported by the same belief; and (2) requires actions to be myopically optimal. These two properties are common in most other standard equilibrium concepts, such as Nash equilibrium. Following Fudenberg and Levine (1993), the first property is known as the unitary-belief property, and puts restrictions on the set of mixed actions that can constitute an equilibrium.[39] The second property is convenient because myopic optimality is easier to characterize than general optimality.

These differences can be relevant. In particular, it is possible for the action frequency to

---

[37]Formally, suppose that an action x is a strict pure-action equilibrium, in that it is uniquely optimal given the equilibrium belief (i.e. $F(\delta_{\theta(x)}) = x$). Then for any sample path h in H^* (the set of histories satisfying the condition of Theorem 1) with $\lim \sigma_t(h) = \delta_x$, the action converges and we have $\lim x_t(h) = x$. This result follows from the fact that if the action frequency converges to δ_x for some sample path in H^*, then after some time, the belief stays in a neighborhood of $\delta_{\theta(x)}$ forever, and the action x is uniquely optimal given such beliefs.

[38]This is the definition for the single agent case; EP2016 also consider the case of multiple agents.

[39]Fudenberg and Levine (1993) showed that non-unitary equilibria make sense in a game where there are multiple players and, for each player, there is an underlying population of agents in the role of that player, and different agents may have different experiences (hence, beliefs) about other players.

converge to an equilibrium with a non-unitary belief, which is not a Berk-Nash equilibrium.[40] We now provide, however, a sufficient condition under which the two properties of equilibrium highlighted above hold, and so limiting action frequencies are always Berk-Nash equilibria.

Definition 8. The family of models is **weakly identified given** $\sigma \in \Delta X$ 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 family of models 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 10. *Suppose that the family of models is weakly identified given σ. Then $\Delta \cup_{\mu \in \Delta\Theta(\sigma)} F_\beta(\mu) \subseteq \Delta \cup_{\mu \in \Delta\Theta(\sigma)} F_0(\mu) = \cup_{\mu \in \Delta\Theta(\sigma)} \Delta F_0(\mu)$. Moreover, if there is a unique KLD minimizer $\theta(\sigma)$, then the previous inclusion is an equality.*

Proposition 10 says that when the agent is myopic, the set of equilibria coincides with the set of Berk-Nash equilibria; here, weak identification guarantees that any mixed action supported by a non-unitary belief is supported by a unitary belief. It also shows that, when the agent is non-myopic, the set of equilibria is contained in the set of Berk-Nash equilibria under weak identification, and the two sets coincide under the stronger property of identification (i.e., unique KLD minimizer for all σ). Intuitively, under identification, the agent's belief is degenerate at $\theta(\sigma)$, so the optimal action coincides with the myopically optimal action.

Similarly, Proposition 10 implies that the set of solutions to the differential inclusion for a non-myopic agent is a subset of that for a myopic agent under weak identification. So under weak identification, if a set A is (globally or robustly) attracting for a myopic agent, the same is true for a non-myopic agent. If σ is an unstable equilibrium for a myopic agent, the action frequency does not converge there regardless of the discount factor.[41] These results are useful, as the optimal policy for a myopic agent has a simpler structure than that of non-myopic agent. Moreover, Proposition 10 implies that, under identifiability (i.e., unique minimizer), the set of solutions to the differential inclusion does not depend on the discount factor. For example, if the action frequency converges to some mixed Berk-Nash equilibrium for a myopic agent, the same is true for a patient, forward-looking agent. An implication is that the long-run welfare for a patient agent is simply the expected payoff in the mixed equilibrium.

---

[40]We provide such an example in the Online Appendix.

[41]Indeed, when the agent is non-myopic, σ is an unstable equilibrium or not an equilibrium.

We now relate Proposition 1 in Section 6 to EP2016's result that if the sequence of distributions over actions converges, then it converges to a Berk-Nash equilibrium. They study distributions of actions because payoff perturbations give agents a motive to mix. In our environment there is no motive for mixing, so convergence of the sequence of distributions over actions implies that the actions converge. Propositions 1 and 10 strengthen EP2016's conclusion by showing that, under weak identification, even though actions may not converge, if the action *frequency* converges, then it converges to a Berk-Nash equilibrium.

We conclude by discussing the relationship to Fudenberg, Lanzani and Strack (2020); henceforth FLS. As mentioned in the introduction, FLS provide conditions under which the action converges with arbitrarily high probability (specifically, if the action is a uniformly strict Berk-Nash equilibrium). In contrast, our differential inclusion describes the evolution of the action frequency, not the action itself. In particular, it is possible that the action frequency converges to a degenerate action but that the action itself diverges. So to see if the action itself converges, one could use FLS's approach and check whether the limiting point is a uniformly strict Berk-Nash equilibrium. On the other hand, if one is interested in convergence of action frequency only, then requiring the uniformity condition of FLS is not necessary.[42]

---

[42]For an example where the action frequency converges but the action diverges, consider the two-model version of Example 1 (the dieter's dilemma). Suppose that $q(1-q) = C$, so that $\sigma^*(1) = 1$, and pick $\bar{\theta}_0^H$ and $\hat{\theta}_0^H$ such that $K(\sigma^*, \bar{\theta}_0^H) = K(\sigma^*, \hat{\theta}_0^H)$. Because the action $G = 1$ is a non-uniform Berk-Nash equilibrium, then Theorem 1 in FLS implies the action does not converge to $G = 1$. The analysis using the differential inclusion, however, looks exactly like Figure 3, except with $\sigma^*(1) = 1$. Thus, the action frequency converges to $\sigma^*(1) = 1$, even though the action diverges because $G = 0$ is chosen infinitely often.

References

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

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.

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.

__ , "Dynamics of stochastic approximation algorithms," in "Seminaire de Probabilites XXXIII," Vol. 1709 of *Lecture Notes in Mathematics*, Springer Berlin Heidelberg, 1999, pp. 1–68.

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.

Camerer, Colin F and Eric J Johnson, "The process-performance paradox in expert judgment: How can experts know so much and predict so badly," *Research on judgment and decision making: Currents, connections, and controversies*, 1997, *342*.

Diaconis, P. and D. Freedman, "On the consistency of Bayes estimates," *The Annals of Statistics*, 1986, pp. 1–26.

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

Edlin, Aaron S and Chris Shannon, "Strict monotonicity in comparative statics," *Journal of Economic Theory*, 1998, *81* (1), 201–219.

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

_ and _ , "Eqilibrium in Misspecified Markov Decision Processes," *working paper*, 2019b.

_ 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.

Freedman, D.A., "On the asymptotic behavior of Bayes' estimates in the discrete case," *The Annals of Mathematical Statistics*, 1963, *34* (4), 1386–1403.

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

— , — , and — , "Stability and Robustness in Misspecified Learning Models," 2019b.

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.

Fudenberg, Drew and Giacomo Lanzani, "The Stability of Misperceptions under Mutations," *Working Paper*, 2020.

— , — , and Philipp Strack, "Limits Points of Endogenous Misspecified Learning," *Available at SSRN*, 2020.

— , 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.

Gilboa, Itzhak and Akihiko Matsui, "Social stability and equilibrium," *Econometrica: Journal of the Econometric Society*, 1991, pp. 859–867.

He, Kevin, "Mislearning from Censored Data: The Gambler's Fallacy in Optimal-Stopping Problems," *arXiv preprint arXiv:1803.08170*, 2018.

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

— , Botond Koszegi, and Philipp Strack, "Convergence in Misspecified Learning Models with Endogenous Actions," *Available at SSRN 3312968*, December 2018b.

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.

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.

Olea, José Luis Montiel, Pietro Ortoleva, Mallesh M Pai, and Andrea Prat, "Competing Models," *arXiv preprint arXiv:1907.03809*, 2019.

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.

Rabin, Matthew, "Psychology and economics," *Journal of economic literature*, 1998, *36* (1), 11–46.

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.

Topkis, Donald M, *Supermodularity and complementarity*, Princeton university press, 1998.

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

White, Halbert, "Maximum likelihood estimation of misspecified models," *Econometrica: Journal of the Econometric Society*, 1982, pp. 1–25.

A   Appendix: Proofs

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 $x \in X$. This fact follows from Assumptions 1-2.

Proof of Lemma 1. Continuity of K: For any $(\theta, \sigma) \in \Theta \times \Delta X$ take a sequence $(\theta_n, \sigma_n)_n$ in $\Theta \times \Delta X$ 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 $\sigma \in \Delta X$, $\theta \mapsto \log \frac{q(Y|X)}{q_\theta(Y|X)}$ 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 $\sigma \in \Delta X$. 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)

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

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

Finally, continuity of K, compactness of Θ (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. $\square$

Proof of Lemma 2. Let $(\theta, z) \mapsto g(\theta, z) \equiv \log \frac{q(y|x)}{q_\theta(y|x)}$, where $z = (y, x) \in Y \times X$. For any $\theta \in \Theta$ and any $\varepsilon > 0$, let $O(\theta, \varepsilon) \equiv \{\theta' : ||\theta' - \theta|| < \varepsilon\}$.

STEP 1. Pointwise convergence. Fix any $\varepsilon > 0$ and any $\theta \in \Theta$. For any $\tau \geq 0$ 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 P^f and the filtration generated by $\{h^t \equiv (x_0(h), y_0(h), x_1(h), y_1(h), ..., x_t(h)) : t \geq 0\}$, because $E_{P^f(\cdot | h^t)}[\zeta_t(h)] = 0$ for all t. Define $h \mapsto \zeta^t(h) \equiv \sum_{\tau=0}^t (1 + \tau)^{-1} \zeta_\tau(h)$ for any $t \geq 0$. 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 $\zeta \in L^2(\mathbb{H}, \mathbb{R}, P^f)$ such that $P^f(\mathscr{H}) = 1$ and, for any $h \in \mathscr{H}$, $\zeta^t(h) \to \zeta(h)$, provided $\sup_t E_{P^f} \left[(\zeta^t)^2 \right] < \infty$. This condition is satisfied because

$$E_{P^f} \left[(\zeta^t)^2 \right] = E_{P^f} \left[\sum_{\tau=0}^t (1 + \tau)^{-2} (\zeta_\tau)^2 \right] + 2E_{P^f} \left[\sum_{\tau > \tau'} (1 + \tau)^{-1} (1 + \tau')^{-1} \zeta_\tau \zeta_{\tau'} \right]$$

$$= \sum_{\tau=0}^t (1 + \tau)^{-2} E_{P^f} \left[(\zeta_\tau)^2 \right]$$

$$\leq \sum_{\tau=0}^t (1 + \tau)^{-2} E_{P^f} \left[\int \left(\sup_{\theta' \in O(\theta, \varepsilon)} g(\theta', y, X_\tau) \right)^2 Q(dy | X_\tau) \right]$$

$$\leq C \max_{x \in X} \int \sup_{\theta' \in O(\theta, \varepsilon)} \left(g(\theta', y, x) \right)^2 Q(dy | x),$$

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

41

$0\ P^f$-a.s. Therefore, we have established that, for all $\theta \in \Theta$,

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

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

$$E_{Q(\cdot|x)}\left[\sup_{\theta'\in O(\theta,\delta(\theta,\varepsilon))}g(\theta',Y,x)-g(\theta,Y,x)\right]<0.25\varepsilon \tag{12}$$

for all $x\in X$. To see this claim, note that, since $\theta\mapsto g(\theta,Y,x)$ is continuous $Q(\cdot|x)-a.s.$ for all $x\in X$, $\lim_{\delta\to 0}\sup_{\theta'\in O(\theta,\delta)}|g(\theta',Y,x)-g(\theta,Y,x)|=0$ a.s.$-Q(\cdot\mid x)$ for all $x\in X$. 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 $x\in X$.

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 $(O(\theta_j,\delta(\theta_j,\varepsilon)))_{j=1,\ldots J(\varepsilon)}$. Thus, for all $\varepsilon>0$,

$$\sup_{\theta\in\Theta}\left|(1+t)^{-1}\sum_{\tau=0}^{t}\left(g(\theta,z_\tau)-E_{Q(\cdot|x_\tau)}\left[g(\theta,Y,x_\tau)\right]\right)\right|$$

$$\leq\max_{j}\sup_{\theta\in O(\theta_j,\delta(\theta_j,\varepsilon))}\left|(1+t)^{-1}\sum_{\tau=0}^{t}\left(g(\theta,z_\tau)-E_{Q(\cdot|x_\tau)}\left[g(\theta,Y,x_\tau)\right]\right)\right|$$

$$\leq\max_{j}(1+t)^{-1}\sum_{\tau=0}^{t}\left(\sup_{\theta\in O(\theta_j,\delta(\theta_j,\varepsilon))}\left|g(\theta,z_\tau)-E_{Q(\cdot|x_\tau)}\left[g(\theta,Y,x_\tau)\right]\right|\right)$$

$$\leq\max_{j}(1+t)^{-1}\sum_{\tau=0}^{t}\left(\left|\sup_{\theta\in O(\theta_j,\delta(\theta_j,\varepsilon))}g(\theta,z_\tau)-E_{Q(\cdot|x_\tau)}\left[\inf_{\theta\in O(\theta_j,\delta(\theta_j,\varepsilon))}g(\theta,Y,x_\tau)\right]\right|\right)$$

$$\leq\max_{j}(1+t)^{-1}\sum_{\tau=0}^{t}\left(\left|\sup_{\theta\in O(\theta_j,\delta(\theta_j,\varepsilon))}g(\theta,z_\tau)-E_{Q(\cdot|x_\tau)}\left[\sup_{\theta\in O(\theta_j,\delta(\theta_j,\varepsilon))}g(\theta,Y,x_\tau)\right]\right|\right)$$

$$+\max_{j}(1+t)^{-1}\sum_{\tau=0}^{t}\left(E_{Q(\cdot|x_\tau)}\left[\sup_{\theta\in O(\theta_j,\delta(\theta_j,\varepsilon))}g(\theta,Y,x_\tau)-\inf_{\theta\in O(\theta_j,\delta(\theta_j,\varepsilon))}g(\theta,Y,x_\tau)\right]\right)$$

$$=I+II.$$

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 P^f-a.s. For the second term, note that (12) implies that

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

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

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

P^f-a.s. The statement in the lemma then follows by noting that

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

Proof of Theorem 1. 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 $t \geq t_\eta$,

$$|L_t(\theta) - K(\theta, \sigma_t)| < \eta \quad (13)$$

for all $\theta \in \Theta$.

Let $\bar{K}(\theta, \sigma) \equiv K(\theta, \sigma) - K^*(\sigma)$. Fix any $\varepsilon > 0$. Using (1) and the facts that $0 \leq K^*(\sigma)$ (the proof is standard) and $K^*(\sigma) < \infty$ for all σ (follows from Assumption 2(iii)), we obtain

$$\int \bar{K}(\theta, \sigma_t) \mu_{t+1}(d\theta) = \frac{\int_\Theta \bar{K}(\theta, \sigma_t) e^{-tL_t(\theta)} \mu_0(d\theta)}{\int_\Theta e^{-tL_t(\theta)} \mu_0(d\theta)}$$

$$= \frac{\int_\Theta \bar{K}(\theta, \sigma_t) e^{-t(L_t(\theta) - K^*(\sigma_t))} \mu_0(d\theta)}{\int_\Theta e^{-t(L_t(\theta) - K^*(\sigma_t))} \mu_0(d\theta)}$$

$$\leq \varepsilon + \frac{\int_{\{\theta: \bar{K}(\theta, \sigma_t) \geq \varepsilon\}} \bar{K}(\theta, \sigma_t) e^{-t(L_t(\theta) - K^*(\sigma_t))} \mu_0(d\theta)}{\int_{\{\theta: \bar{K}(\theta, \sigma_t) \leq \varepsilon/2\}} e^{-t(L_t(\theta) - K^*(\sigma_t))} \mu_0(d\theta)}$$

$$=: \varepsilon + \frac{A_t^\varepsilon}{B_t^\varepsilon}.$$

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

By (13), there exists t_η such that, for all $t \geq t_\eta$,

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

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

$$
\frac{A_t^\varepsilon}{B_t^\varepsilon} \leq e^{t2\eta} \frac{e^{-t\varepsilon/2}}{\mu_0(\{\theta : \bar{K}(\theta,\sigma_t) \leq \varepsilon/2\})}. \tag{14}
$$

At the end of this proof, we establish that continuity of $\bar{K}$ and compactness of ΔX imply that

$$
\kappa_\varepsilon \equiv \inf_{\sigma \in \Delta X} \mu_0(\{\theta : \bar{K}(\theta,\sigma) \leq \varepsilon/2\}) > 0 \tag{15}
$$

for all $\varepsilon > 0$. Thus, setting $\eta = \varepsilon/8 > 0$, (14) implies that, for all $t \geq \max\{t_\eta, 1/\varepsilon\}$,

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

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

Proof of equation (15): 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 $\Theta \times \Delta X$ 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, $\{\theta' : \|\theta_\sigma - \theta'\| < \delta_k\} \subseteq \{\theta : \bar{K}(\theta,\sigma') \leq k\}$ for all $\sigma' \in B(\sigma,\delta_k) \equiv \{\sigma' : \|\sigma - \sigma'\| < \delta_k\}$. Thus, for all σ, $\mu_0(\{\theta : \bar{K}(\theta,\sigma') \leq k\}) \geq \mu_0(\{\theta' : \|\theta_\sigma - \theta'\| < \delta_k\})$ for all $\sigma' \in B(\sigma,\delta_k)$. The balls $\{B(\sigma,\delta_k)\}_\sigma$ form an open cover for ΔX. Since ΔX is compact, there exists a finite subcover $\{B(\sigma^i,\delta_k)\}_{i=1}^n$. Let $r \equiv \min_{i \in \{1,\ldots,n\}} \mu_0(\{\theta' : \|\theta_{\sigma^i} - \theta'\| < \delta_k\})$ 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 $\mu_0(\{\theta : \bar{K}(\theta,\sigma') \leq k\}) \geq \mu_0(\{\theta' : \|\theta_{\sigma^i} - \theta'\| < \delta_k\}) \geq r > 0$. $\square$

Proof of Theorem 2. The proof of Theorem 2 consists of three parts. Part 1 defines an enlargement of the set of actions that allows us to adopt the methods developed by BHS2005.

44

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

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

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

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

where $\bar{K}(\theta, \sigma') \equiv K(\theta, \sigma') - K^*(\sigma')$. Note that $M(0, \sigma) = \Theta(\sigma)$ and so $\Xi(0, \sigma) = \cup_{\mu \in \Delta \Theta(\sigma)} \Delta F(\mu) - \sigma$. We will establish two properties in this part:

(1a) $(\delta, \sigma) \mapsto \Xi(\delta, \sigma)$ is uhc: 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 $(\mu_n, \sigma'_n)_n$

45

lives in a compact set, $\Delta\Theta \times \Delta X$, 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, $y \in \Xi(\delta, \sigma)$.

(1b) There exists a sequence $(\delta_t)_t$ with $\lim_{t\to\infty} \delta_t = 0$ such that, for all t, $\sigma_{t+1} - \sigma_t \in \frac{1}{t+1}\Xi(\delta_t, \sigma_t)$: By equation (7) 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 (8) in the text. By property (1b), 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 $\gamma(t) \equiv \delta_{m(t)} + \|\mathbf{w}(t) - \sigma_{m(t)}\|$. Then $\dot{\mathbf{w}}(t) \in \Xi(\gamma(t), \mathbf{w}(t))$ for almost every t. In addition, note that $\lim_{t\to\infty} \gamma(t) = 0$ because $(\delta_t)_t$ goes to zero, $m(t)$ goes to infinity, and $\mathbf{w}$ is the interpolation of $(\sigma_t)_t$.

Part 3. A perturbed solution is an asymptotic pseudotrajectory (i.e., it satisfies equation (10) in the text). Let $v(t) \equiv \dot{\mathbf{w}}(t) \in \Xi(\gamma(t), \mathbf{w}(t))$ for almost every t. Then

$$\mathbf{w}(t+s) - \mathbf{w}(t) = \int_0^s \mathbf{v}(t+\tau)d\tau. \tag{16}$$

Since $\mathbb{S}$ is a bounded set, v is uniformly bounded; therefore, $\mathbf{w}$ is uniformly continuous. Hence, the family of functions $\{s \mapsto \mathbf{S}^t(\mathbf{w})(s) : t \in \mathbb{R}\}$ — where for each (t, s) $\mathbf{S}^t(\mathbf{w})(s) = \mathbf{w}(s+t)$ — is equicontinuous and, therefore, relatively compact with respect to $L^\infty(\mathbb{R}, \Delta X, Leb)$, 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 $\mathbf{w}^* \in L^\infty(\mathbb{R}, \Delta X)$ such that $\mathbf{w}^* = \lim_{t_n \to \infty} \mathbf{S}^{t_n}(\mathbf{w})$.

Set $t = t_n$ in (16) 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-Alaoglu 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,

$$\lim_{n\to\infty} \int_0^s \mathbf{v}_n(\tau)d\tau = \int_0^s \mathbf{v}^*(\tau)d\tau \tag{17}$$

46

pointwise in $s \in [0, T]$. Indeed, convergence is uniform because the family $\{s \mapsto \int_0^s \mathbf{v}_n(\tau) d\tau : n \in \mathbb{N}\}$ is equicontinuous and $[0, T]$ is compact. In addition, $\mathbf{v}^* \in L^2([0, T], \mathbb{S})$, then $\mathbf{w}^*$ is absolutely continuous in $[0, T]$.

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

We will prove $\mathbf{v}^*(\tau) \in Co(\Xi(0, \mathbf{w}^*(\tau)))$ Lebesgue-a.s., 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 $\lim_{n \to \infty} \|\bar{\mathbf{v}}_n - \mathbf{v}^*\|_{L^2([0,T],\mathbb{S})} = 0$ where $\bar{\mathbf{v}}_n \equiv \sum_{k=n}^{N(n)} \alpha_k \mathbf{v}_n$. Therefore, as $\lim_{n \to \infty} \|\bar{\mathbf{v}}_n - \mathbf{v}^*\|_{L^2([0,T],\mathbb{S})} = 0$, it follows that $\lim_{n \to \infty} \bar{\mathbf{v}}_n = \mathbf{v}^*$ a.s.-Lebesgue.

Fix $\tau \in [0, T]$ such that the previous claim holds. Define $\gamma_n(\tau) \equiv \gamma(t_n + \tau)$ and $\mathbf{w}_n(\tau) \equiv \mathbf{w}(t_n + \tau)$. By uhc of Ξ at $(0, \sigma)$ for all σ (see property (1a)) and the facts that $\gamma_n(\tau) \to 0$ and $\mathbf{w}_n(\tau) \to w^*(\tau)$, it follows that, for any $\varepsilon > 0$, there exists N_ε such that, for all $n \geq N_\varepsilon$, $\Xi(\gamma_n(\tau), \mathbf{w}_n(\tau)) \subseteq \Xi^\varepsilon(0, \mathbf{w}^*(\tau))$, 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 $\lim_{j \to \infty} \bar{\mathbf{v}}_n(\tau) = \mathbf{v}^*(\tau)$, it follows that $\mathbf{v}^*(\tau) \in Co(\Xi^\varepsilon(0, \mathbf{w}^*(\tau)))$. Since this is true for all $\varepsilon > 0$, it follows that $\mathbf{v}^*(\tau) \in Co(\Xi(0, \mathbf{w}^*(\tau)))$. $\square$

Proof of Proposition 1. Let σ^* be an arbitrary non-equilibrium point. Then there is a pure action x such that $\sigma^*(x) > 0$ and $x \notin F(\triangle\Theta(\sigma^*))$. Choose such x. By upper hemi-continuity of F (Assumption 4) and $\Theta(\cdot)$ (Lemma 1) it follows that there exists a $\varepsilon > 0$ such that $x \notin F(\triangle\Theta(\sigma))$ for all $\sigma \in B_\varepsilon(\sigma^*)$ and such that $\inf_{\sigma \in B_\varepsilon(\sigma^*)} \sigma(x) > 0$. Pick such $\varepsilon > 0$. Then there is some $T > 0$ such that for any initial value in this ε-neighborhood, $\boldsymbol{\sigma}(0) \in B_\varepsilon(\sigma^*)$ and any solution $\boldsymbol{\sigma} \in S_{\boldsymbol{\sigma}(0)}^\infty$ to the differential inclusion leaves this neighborhood within time T, i.e., we have

$$\|\boldsymbol{\sigma}(\tau) - \sigma^*\| \geq \varepsilon \tag{18}$$

for some $\tau < T$. Such T exists, because the share of the action x decreases whenever $\boldsymbol{\sigma}(\tau)$ is in the set $B_\varepsilon(\sigma^*)$.

Now, pick a sample path h such that the property stated in Theorem 2 holds. We will show that σ_t cannot stay in the $\frac{\varepsilon}{2}$-neighborhood of σ^* forever. This completes the proof, because it

47

implies that almost surely, σ_t cannot converge to any non-equilibrium point σ^*.

Pick $\tilde{T}$ such that for any time $t > \tilde{T}$,

$$\inf_{\boldsymbol{\sigma} \in S^T_{\boldsymbol{w}(t)}} \|\boldsymbol{w}(t+s) - \boldsymbol{\sigma}(s)\| < \frac{\varepsilon}{2} \quad \forall s \in [0, T] \tag{19}$$

Suppose there exists a $t > \tilde{T}$ such that $\boldsymbol{w}(t) \in B_{\frac{\varepsilon}{2}}(\sigma^*)$ (if no such t exists, then the proof is finished because it follows that σ_t is outside a $\varepsilon/2$ neighborhood of σ^* for all $t > \tilde{T}$). Then from (18) and (19), there is $s \in [0, T]$ such that $\|\boldsymbol{w}(t+s) - \sigma^*\| \geq \varepsilon/2$. So σ_t cannot stay in the $(\varepsilon/2)$-neighborhood forever. $\square$

Proof of Proposition 2. Part (i) directly follows from part (ii). Proof of part (ii): Pick a history from the set of histories with probability one defined by the statement of Theorem 2, and let $\boldsymbol{w}$ denote the interpolation of the action frequency σ_t given this path. If there is t^* such that $\boldsymbol{w}(t) \in E$ for all $t > t^*$, the result follows. So we will focus on the case in which for any t^*, there is $t > t^*$ such that $\boldsymbol{w}(t) \notin E$.

Pick attracting sets $(A_1, \cdots, A_N)$ as stated. Pick an arbitrarily small $\varepsilon > 0$. Without loss of generality, we assume that for each attracting set A_n, the ε-neighborhood of A_n is in the basin of attraction $\mathscr{U}_{A_n}$.

Pick T large enough that for any attracting set A_n, for any initial value $\boldsymbol{\sigma}(0) \in \mathscr{U}_{A_n}$, for any $\boldsymbol{\sigma}^{2T}_{\boldsymbol{\sigma}(0)}$, and for any $s \in [T, 2T]$,

$$d(\boldsymbol{\sigma}(s), A_n) < \frac{\varepsilon}{2}. \tag{20}$$

Also, pick $\tilde{T}$ large enough that for any $t > \tilde{T}$ and for any $s \in [0, 2T]$

$$\inf_{\boldsymbol{\sigma} \in S^{2T}_{\boldsymbol{w}(t)}} \|\boldsymbol{w}(t+s) - \boldsymbol{\sigma}(s)\| < \frac{\varepsilon}{2}. \tag{21}$$

Recall that for any t^*, there is $t > t^*$ such that $\boldsymbol{w}(t) \notin E$. This implies that there is $t > \tilde{T}$ and an attracting set A_n such that $\boldsymbol{w}(t) \in \mathscr{U}_{A_n}$. Pick such t and A_n. From (20) and (21), we have $d(\boldsymbol{w}(t+s), A_n) < \varepsilon$ for all $s \in [T, 2T]$. This implies that $\boldsymbol{w}(t+s) \in \mathscr{U}_{A_n}$ for all $s \in [T, 2T]$, so applying the same argument iteratively, we have $d(\boldsymbol{w}(t+s), A_n) < \varepsilon$ for all $s \geq T$, which means that $\boldsymbol{w}$ will stay in the ε-neighborhood of the attracting set A_n forever. Since ε can be arbitrarily small, $d(\boldsymbol{w}(t) - A_n)$ converges to zero as $t \to \infty$. (Note that choosing smaller ε does not influence A_n.) $\square$

Proof of Proposition 3. Let σ^* be an unstable equilibrium, and pick $\mathscr{U}$ and T as in the

definition of unstable equilibrium. Pick a history from the set of histories with probability one defined by the statements of Theorems 1 and 2. Let $\mathbf{w}$ denote the interpolation of the action frequency σ_t given this path. It suffices to show that $\mathbf{w}(t)$ does not converge to σ^* given this history.

Pick a sufficiently small $\varepsilon > 0$, so that 2ε-neighborhood of σ^* is a subset of $\mathscr{U}$. Without loss of generality, we can assume that there is $\eta > 0$ such that $F(\mu) \subseteq F(\triangle \Theta(\sigma^*))$ for any μ such that $\int (K(\theta, \sigma) - K^*(\sigma))\mu(d\theta) < \eta$ for some $\sigma \in B_\varepsilon(\sigma^*)$. (If necessary, take ε small.) Pick such $\eta > 0$.

From Theorems 1 and 2, there is T^* and τ^* such that $T^* = \sum_{i=1}^{\tau^*} \frac{1}{i}$,

$$\int (K(\theta, \sigma_\tau) - K^*(\sigma_\tau))\mu_{\tau+1}(d\theta) < \eta \tag{22}$$

for all $\tau \geq \tau^*$, and

$$\inf_{\boldsymbol{\sigma} \in S_{\mathbf{w}(t)}^T} \sup_{0 \leq s \leq T} \|\mathbf{w}(t+s) - \boldsymbol{\sigma}(s)\| < \varepsilon \tag{23}$$

for all $t \geq T^*$.

Suppose that $\mathbf{w}(t)$ is in the ε-neighborhood of σ^* for some $t > T^*$ such that $t = \sum_{i=1}^{\tau} \frac{1}{i}$ for some τ. We will show that there is $t' > 0$ such that $\mathbf{w}(t+t')$ is not in the ε-neighborhood of σ^*. This completes the proof, because it implies that $\mathbf{w}$ cannot stay around σ^* forever. Let $\sigma = \mathbf{w}(t)$ satisfy the condition in the definition of unstable equilibrium. From (22) and the definition of η, the agent chooses some action $x \in F(\triangle \Theta(\sigma^*))$ in the current period. This means that $\mathbf{w}(\tilde{t})$ moves toward δ_x during the time $\tilde{t} \in [t, t + \frac{1}{\tau+1}]$. Then from the condition in the definition of unstable equilibrium, there is $\tilde{t} \in [t, t + \frac{1}{\tau+1}]$ such that for any $\boldsymbol{\sigma} \in S_{\mathbf{w}(\tilde{t})}^\infty$, we have $\boldsymbol{\sigma}(t) \notin \mathscr{U}$ for some $t \in [0, T]$. Then as in the previous case, we can show that there is $t' \leq T$ such that $\|\mathbf{w}(\tilde{t} + t') - \sigma\| > \varepsilon$, as desired. $\square$

Proof of Proposition 4. We will first present a few preliminary results. We have seen in Lemma 2 that given any initial prior μ_0 and given any policy f, there is T such that with positive probability, the consequence frequency is close to the mean (more formally, the sample average of the likelihood L_t is close to the mean) for *all* periods after T. The following lemma shows that this T can be chosen independently of μ_0 and f. The proof can be found in the online appendix.

Lemma 1. *For any $\eta > 0$, there is T and $\underline{q} > 0$ such that for any initial prior μ_0 with full*

49

support and for any f,

$$P^f(\forall t \geq T \forall \theta \ |L_t(\theta) - K(\theta, \sigma_t)| < \eta) > \underline{q}. \tag{24}$$

The next claim just summarizes what we have seen in the proof of Theorem 1: It shows that if the past consequence frequency is close to the mean as stated in the above claim, and if the initial prior μ_0 satisfies some technical condition, then the posterior belief μ_{t+1} concentrates on the states which approximately minimize $K(\theta, \sigma_t)$ for large t. The proof directly follows from the proof of Theorem 1.

Claim 1. *For any $\eta > 0$ and for any $\kappa > 0$, there is T such that for any initial prior μ_0 and for any $t > T$ such that $|L_t(\theta) - K(\theta, \sigma_t)| < \frac{\eta}{16}$ and $\mu_0(\{\theta : K(\theta, \sigma_t) - K^*(\sigma_t) \leq \frac{\eta}{4}\}) \geq \kappa$,*

$$\int (K(\theta, \sigma_t) - K^*(\sigma_t))\mu_{t+1}(d\theta) < \eta.$$

The next claim shows that if the posterior belief μ_{t+1} is concentrated as stated in the above claim then the motion of the action frequency σ_t is described by the perturbed differential inclusion. A difference from Theorem 2 is that here the motion of the action frequency $\mathbf{w}$ exactly matches a solution to the perturbed differential inclusion. In contrast, in Theorem 2, we take the limit as $t \to \infty$, so a solution to the differential inclusion is an approximation of the action frequency $\mathbf{w}$.

Claim 2. *Let F be an uhc policy correspondence. Then for any $\varepsilon > 0$, there is $\eta > 0$ such that given a sample path h, for any $t > \frac{1}{\varepsilon}$ such that $\int (K(\theta, \sigma_t) - K^*(\sigma_t))\mu_t(d\theta) < \eta$, there is $\boldsymbol{\sigma} \in S_{\mathbf{w}(T)}^{\infty, \varepsilon}$ such that $\mathbf{w}(T+s) = \boldsymbol{\sigma}(s)$ for all $s \in [0, \frac{1}{t+1}]$, where $T = \sum_{\tau=1}^{t} \frac{1}{\tau}$.*

Proof. Pick $\varepsilon > 0$ arbitrarily. It is sufficient to show that there is $\eta > 0$ such that for any σ and for any μ such that $\int (K(\theta, \sigma) - K^*(\sigma))\mu(d\theta) < \eta$, there is $\tilde{\sigma} \in B_{\frac{\varepsilon}{2}}(\sigma)$ such that $F(\mu) \subseteq F(\triangle\Theta(\tilde{\sigma}))$.

Note that for each σ, there is $\varepsilon_\sigma < \varepsilon$ and η_σ such that $F(\mu) \subseteq F(\triangle\Theta(\sigma))$ for all μ such that $\int (K(\theta, \tilde{\sigma}) - K^*(\tilde{\sigma}))\mu(d\theta) < \eta_\sigma$ for some $\tilde{\sigma} \in B_{\varepsilon_\sigma}(\sigma)$. Since $\triangle X$ is compact, there is a finite subcover $\{B_{\varepsilon_{\sigma_1}}(\sigma_1), \cdots, B_{\varepsilon_{\sigma_M}}(\sigma_M)\}$. Let $\eta = \min_m \eta_{\sigma_m} > 0$. This η satisfies the property we want. Indeed, for any σ and μ such that $\int (K(\theta, \sigma) - K^*(\sigma))\mu(d\theta) < \eta$, if we set $\tilde{\sigma} = \sigma_m$ such that $\sigma \in B_{\varepsilon_{\sigma_m}}(\sigma_m)$, we have $F(\mu) \subseteq F(\triangle\Theta(\tilde{\sigma}))$. $\square$

The next claim shows that to prove convergence to an attracting set A, it suffices to show that σ_t visits the basin of A infinitely often with positive probability.

50

Claim 3. *Suppose that given an initial prior μ_0 and a policy f, σ_t visits the basin of A infinitely often with positive probability, i.e., $P^f(\forall T \exists t > T \quad \sigma_t \in \mathcal{U}_A) > 0$. Then $P^f(\lim_{t \to \infty} d(\sigma_t, A) = 0) > 0$.*

Proof. Let $\mathcal{H}$ be the set of all h which satisfies the property stated in Theorem 1. Note that $P^f(\mathcal{H}) = 1$. Also, let $\tilde{\mathcal{H}}$ be the set of all h such that σ_t visits the basin of A infinitely often, i.e., it is the set of all h such that for any T, there is $t > T$ such that $\sigma_t \in \mathcal{U}_A$. Let $\mathcal{H}^* = \mathcal{H} \cap \tilde{\mathcal{H}}$. By the assumption, we have $P^f(\mathcal{H}^*) = P^f(\tilde{\mathcal{H}}) > 0$.

Pick a sample path $h \in \mathcal{H}^*$. To prove the claim, it suffices to show that $\lim_{t \to \infty} d(\sigma_t, A) = 0$ given this path. Pick an arbitrary $\varepsilon > 0$. Without loss of generality, we assume that $B_\varepsilon(A)$ is in the basin of attraction $\mathcal{U}_A$.

Pick T large enough that (20) holds for any initial value $\boldsymbol{\sigma}(0) \in \mathcal{U}_A$, for any $\boldsymbol{\sigma} \in S^{2T}_{\boldsymbol{\sigma}(0)}$, and for any $s \in [T, 2T]$. Also, pick $\tilde{T}$ large enough that (21) holds for any $t > \tilde{T}$ and for any $s \in [0, 2T]$.

Since σ_t visits $\mathcal{U}_A$ infinitely often, there is $t > \tilde{T}$ such that $\boldsymbol{w}(t) \in \mathcal{U}_A$. Pick such t. Then as in the proof of Proposition 2(ii), we can show that $\boldsymbol{w}$ will stay in the ε-neighborhood of the set A forever. Since ε can be arbitrarily small, $\lim_{t \to \infty} d(\boldsymbol{w}(t), A) = 0$. $\square$

Now we will prove the proposition. Let A be a robustly attracting set, and let $\overline{\zeta} > 0$ be such that $B_{\overline{\zeta}}(A) \subset \mathcal{U}_A$. Let ζ and ε be as in the definition of robustly attracting set. Then pick η as in Claim 2, pick an arbitrary $\kappa > 0$, and pick T as stated in Claim 1.

Pick t^* large enough that $\frac{1}{t^*} < \varepsilon$ and

$$\frac{t^*}{t^* + T} \sigma + \frac{T}{t^* + T} \tilde{\sigma} \in B_\zeta(A) \tag{25}$$

for all $\sigma \in B_{\frac{\zeta}{2}}(A)$ and $\tilde{\sigma} \in \triangle X$. Now, consider the following hypothetical situation:

(a) The initial prior is μ_0 such that $\mu_0(\{\theta : K(\theta, \sigma) - K^*(\sigma) \le \frac{\eta}{4}\}) > \kappa$ for all σ. The current period is $t^* + 1$.

(b) The action frequency in the past is close to A, in that $\sigma_{t^*} \in B_{\frac{\zeta}{2}}(A)$.

(c) The past observation is close to the mean, in that $|L_{t^*}(\theta) - K(\theta, \sigma_{t^*})| < \frac{\eta}{16}$ for all θ.[43]

Let h^{t^*} be a history which satisfies all the properties above. (Given a policy f, the probability of such a history h^{t^*} may be zero, but this does not affect the following argument.) Let $\mathcal{H}$ be

---

[43]Claim 1 ensures that this condition can be satisfied by some consequence sequence.

the set of histories such that the history during the first t^* periods is exactly h^{t^*} and

$$|L_{t^*+1,t}(\theta) - K(\theta, \sigma_{t^*+1,t})| < \frac{\eta}{16} \qquad (26)$$

for all $t \geq t^* + T$, where $L_{t^*+1,t}$ is the sample average of the likelihood from period t^*+1 to period t, and $\sigma_{t^*+1,t}$ is the action frequency from period t^*+1 to period t. From Lemma 1, we know $P^f(\mathscr{H}|h^{t^*}) > q$.

Pick a path $h \in \mathscr{H}$. We claim that given this path, σ_t never leaves the basin of A after period t^*.

Claim 4. *For each path $h \in \mathscr{H}$, $\sigma_t \in \mathscr{U}_A$ for all $t > t^*$.*

Proof. Pick h as stated. From (b) and (25), we have $\sigma_t \in B_\zeta(A) \subseteq \mathscr{U}_A$ for all $t \in \{t^* + 1, \cdots, t^* + T\}$, regardless of the agent's play during these periods. So what remains is to show that $\sigma_t \in \mathscr{U}_A$ for all $t > t^* + T$. From (26), (a), (c), and Claim 1, we have $\int (K(\theta, \sigma_t) - K^*(\sigma_t))\mu_{t+1}(d\theta) < \eta$ for all $t \geq t^* + T$. Then Claim 2 implies that the motion of the action frequency after period $t^* + T$ is described by some solution to the ε-perturbed differential inclusion. Since $\sigma_{t+T^*} \in B_\zeta(A)$ and σ^* is robustly attracting, we have $\sigma_t \in B_{\overline{\zeta}}(A) \subseteq \mathscr{U}_A$ for all $t \geq t^* + T$. $\square$

Let μ^* be the posterior belief induced by the initial prior μ_0 and the history h^{t^*} above.

Now, consider a new "game" in which the agent's initial prior is μ^*. Since the agent's action is determined by the belief, her play in this new game is exactly the same as her play in the continuation game induced by the initial prior μ_0 and the history h^{t^*}. So Claim 4 implies that in this new game, with positive probability, the action frequency σ_t will stay in the basin $\mathscr{U}_A$ in all periods $t > \tilde{T}$, where $\tilde{T}$ is a sufficiently large number. (This is so because the action frequency σ_{t^*} during the first t^* periods has almost no impact on the action frequency σ_t for large t.) Then Claim 3 implies that in this new game, the action frequency σ_t converges to A with positive probability. $\square$

Proof of Proposition 5. We will start with a useful lemma, which shows that Assumption 5 essentially requires single-peakedness of the Kullback-Leibler divergence $K(\theta, \delta_x)$. Let $\underline{\theta} = \min_{\sigma \in \triangle X} \theta(\sigma)$, and let $\overline{\theta} = \max_{\sigma \in \triangle X} \theta(\sigma)$. The proof of the following lemma can be found in the online appendix.

Lemma 2. *If Assumption 5 holds, then for each action frequency σ, the Kullback-Leibler divergence $K(\theta, \sigma)$ is single-peaked with respect to θ in $[\underline{\theta}, \overline{\theta}]$, that is, we have $K(\theta, \sigma) >$*

$K(\tilde{\theta}, \sigma)$ *for each* $\theta \in [\underline{\theta}, \theta(\sigma))$ *and* $\tilde{\theta} \in (\theta, \theta(\sigma)]$, *and* $K(\theta, \sigma) < K(\tilde{\theta}, \sigma)$ *for each* $\theta \in [\theta(\sigma), \overline{\theta})$ *and* $\tilde{\theta} \in (\theta, \overline{\theta}]$.

We now prove each part of the proposition.

Part (i): A standard algebra shows that

$$K(\theta, \sigma_\beta) = \beta K(\theta, \sigma) + (1-\beta) K(\theta, \tilde{\sigma})$$

for each θ. Then the result follows immediately.

Part (ii): We first show that $\theta(\sigma_\beta) \geq \theta(\tilde{\sigma})$ for all β. Suppose not so that there is $\beta_1 \in (0,1)$ such that $\theta(\sigma_{\beta_1}) < \theta(\tilde{\sigma})$. Then since $\theta(\sigma_\beta)$ is continuous in β and $\theta(\sigma_{\beta_1}) < \theta(\tilde{\sigma}) < \theta(\sigma_1)$, there must be some β_2 such that $\beta_1 < \beta_2 < 1$ and $\theta(\sigma_{\beta_2}) = \theta(\tilde{\sigma})$. But then from part (i), we have $\theta(\sigma_\beta) = \theta(\tilde{\sigma})$ for all $\beta \in [0, \beta_2]$, and in particular $\theta(\sigma_{\beta_1}) = \theta(\tilde{\sigma})$. This is a contradiction.

Similarly, we can show that $\theta(\sigma_\beta) \leq \theta(\sigma)$ for all β. Taken together, we have $\theta(\sigma_\beta) \in [\theta(\tilde{\sigma}), \theta(\sigma)]$ for all β. Now, from Claim 8 in the proof of Lemma 2, $K(\theta, \sigma)$ has increasing differences, in that

$$\frac{\partial^2 K(\theta, \sigma)}{\partial \theta \partial \beta} = \frac{\partial K(\theta, \sigma)}{\partial \theta} - \frac{\partial K(\theta, \tilde{\sigma})}{\partial \theta} \geq 0.$$

for all β and $\theta \in [\theta(\tilde{\sigma}), \theta(\sigma)]$. So the monotone selection theorem of Topkis implies the result we want.

Part (iii): Pick β_1 and β_2 as stated. Let $\theta^* = \theta(\sigma_{\beta_1})$. This θ^* is an interior solution, so it must solve the first-order condition $\frac{\partial K(\theta^*, \sigma_{\beta_1})}{\partial \theta} = 0$, which is equivalent to

$$\beta_1 \frac{\partial K(\theta^*, \sigma)}{\partial \theta} + (1-\beta_1) \frac{\partial K(\theta^*, \tilde{\sigma})}{\partial \theta} = 0. \tag{27}$$

We claim that each term in the left-hand side is non-zero:

Claim 5. $\frac{\partial K(\theta^*, \sigma)}{\partial \theta} \neq 0$.

Proof. Suppose not so that $\frac{\partial K(\theta^*, \sigma)}{\partial \theta} = 0$. Then from (27), we have $\frac{\partial K(\theta^*, \tilde{\sigma})}{\partial \theta} = 0$, that is, θ^* satisfies the first-order condition for σ and $\tilde{\sigma}$. Then we must have $\frac{\partial^2 K(\theta^*, \sigma)}{\partial \theta^2} \geq 0$. Indeed, if not and $\frac{\partial^2 K(\theta^*, \sigma)}{\partial \theta^2} < 0$, θ^* becomes the local maxima for $K(\theta, \sigma)$, which contradicts with the single-peakedness of $K(\theta, \sigma)$. Similarly we have $\frac{\partial^2 K(\theta^*, \tilde{\sigma})}{\partial \theta^2} \geq 0$

Also, from Assumption 5, we know that the second-order condition, $\frac{\partial^2 K(\theta^*, \sigma_{\beta_1})}{\partial \theta^2} > 0$, is

53

satisfied for σ_{β_1}, which is equivalent to

$$\beta_1 \frac{\partial^2 K(\theta^*, \sigma)}{\partial \theta^2} + (1 - \beta_1) \frac{\partial^2 K(\theta^*, \tilde{\sigma})}{\partial \theta^2} > 0.$$

This inequality implies $\frac{\partial^2 K(\theta^*, \sigma)}{\partial \theta^2} > 0$ or $\frac{\partial^2 K(\theta^*, \tilde{\sigma})}{\partial \theta^2} > 0$. Suppose for now that $\frac{\partial^2 K(\theta^*, \sigma)}{\partial \theta^2} > 0$. (The argument for the case with $\frac{\partial^2 K(\theta^*, \tilde{\sigma})}{\partial \theta^2} > 0$ is symmetric, so we will omit it.) Then since $\frac{\partial^2 K(\theta^*, \tilde{\sigma})}{\partial \theta^2} \geq 0$, we have $\frac{\partial^2 K(\theta^*, \sigma_\beta)}{\partial \theta^2} > 0$ for all $\beta \neq 0$. Also, since $\frac{\partial K(\theta^*, \sigma)}{\partial \theta} = \frac{\partial K(\theta^*, \tilde{\sigma})}{\partial \theta} = 0$, we have $\frac{\partial K(\theta^*, \sigma_\beta)}{\partial \theta} = 0$ for all β. So θ^* satisfies both the first-order and the second-order conditions, which implies that $\theta(\sigma_\beta) = \theta^*$ for all $\beta \neq 0$. Then since $\theta(\sigma_\beta)$ is continuous in β, we have $\theta(\sigma_\beta) = \theta^*$ for all $\beta \in [0, 1]$. But this is a contradiction, because we have $\theta(\tilde{\sigma}) < \theta(\sigma)$. □

The above claim and (27) imply that

$$\frac{\partial K(\theta^*, \sigma_{\beta_2})}{\partial \theta} = \beta_2 \frac{\partial K(\theta^*, \sigma)}{\partial \theta} + (1 - \beta_2) \frac{\partial K(\theta^*, \tilde{\sigma})}{\partial \theta} \neq 0,$$

which means that θ^* cannot be the optimal solution for β_2. (Note that θ^* is an interior value, so the first-order condition is necessary for it to be optimal.) Then from part (ii), the result follows. □

Proof of Proposition 6. Let Θ^{**} be the union of the equilibrium models and the boundary points, that is, $\Theta^{**} = \Theta^* \cup \{0, 1\}$. Since Θ^* is finite, it can be written as $\Theta^{**} = \{\theta_0, \theta_1, \cdots, \theta_N\}$ where $0 = \theta_0 < \cdots < \theta_N = 1$.

We first show that each interval (θ_n, θ_{n+1}) has a useful property.

Lemma 3. *Each interval (θ_n, θ_{n+1}) must satisfy one of the following properties:*

   (i) For each $\theta \in (\theta_n, \theta_{n+1})$ and for each $x \in F(\delta_\theta)$, we have $\theta(\delta_x) > \theta$.

   (ii) For each $\theta \in (\theta_n, \theta_{n+1})$ and for each $x \in F(\delta_\theta)$, we have $\theta(\delta_x) < \theta$.

Proof. If there is $\theta \in (\theta_n, \theta_{n+1})$ such that $\theta(\delta_x) = \theta$ for some $x \in F(\delta_\theta)$, then this θ is an equilibrium model, which is a contradiction. So such θ does not exist. Similarly, if there is $\theta \in (\theta_n, \theta_{n+1})$ such that $\theta(\delta_x) < \theta < \theta(\delta_{\tilde{x}})$ for some $x, \tilde{x} \in F(\delta_\theta)$, then there is a mixture σ of x and $\tilde{x}$ such that $\theta(\sigma) = \theta$, which implies that θ is a mixed-strategy equilibrium model. So again such θ does not exist. Accordingly, (θ_n, θ_{n+1}) must be the union of the two sets, Θ_1 and Θ_2: Θ_1 is the set of all $\theta \in (\theta_n, \theta_{n+1})$ such that $\theta(\delta_x) > \theta$ for all $x \in F(\delta_\theta)$. Θ_2 is the

set of all $\theta \in (\theta_n, \theta_{n+1})$ such that $\theta(\delta_x) < \theta$ for all $x \in F(\delta_\theta)$. However, since $F(\delta_\theta)$ is upper hemi-continuous in θ, one of these sets must be empty. This implies the result. □

Next, we characterize how the KLD minimizer $\theta(\sigma_t)$ changes over time, when the motion of σ_t is determined by the differential inclusion. Consider an interval (θ_n, θ_{n+1}) which satisfies property (i) in the lemma above. Pick a solution $\boldsymbol{\sigma}$ to the differential inclusion, and suppose that $\theta(\boldsymbol{\sigma}(t)) \in (\theta_n, \theta_{n+1})$ in the current period t. Then from property (i), the agent will choose an action x such that $\theta(\delta_x) > \theta(\boldsymbol{\sigma}(t))$, which means that $\theta(\boldsymbol{\sigma}(t))$ should move up and eventually reaches (a neighborhood of) θ_{n+1}. Also, once $\theta(\boldsymbol{\sigma}(t))$ goes above θ_{n+1}, it cannot be lower than θ_{n+1} in any later period. Formally, we have the following result:

Lemma 4. *Suppose that the interval (θ_n, θ_{n+1}) satisfies property (i) stated in Lemma 3. Then for any $\varepsilon > 0$, there is $T > 0$ such that given any initial value $\boldsymbol{\sigma}(0)$ with $\theta(\boldsymbol{\sigma}(0)) > \theta_n$ and given any solution $\boldsymbol{\sigma} \in S_{\boldsymbol{\sigma}(0)}^\infty$ to the differential inclusion, we have $\theta(\boldsymbol{\sigma}(t)) > \theta_{n+1} - \varepsilon$ for all $t \geq T$.*

Proof. Let $X^* = \cup_{\theta \in (\theta_n, \theta_{n+1})} F(\delta_\theta)$. We first consider the special case in which $\theta(\delta_x) \geq \theta_{n+1}$ for all $x \in X^*$. Then we will explain how to extend the proof for a general case.

Case 1: $\theta(\delta_x) \geq \theta_{n+1}$ for all $x \in X^*$. Let $\mathscr{X}$ be the set of all mixed strategies σ such that $\theta(\sigma) \geq \theta_{n+1}$. From Proposition 5(ii), this set is convex. Similarly, the set $\triangle X \setminus \mathscr{X}$ is convex. So there is a hyperplane H which separates these two sets; i.e., there is a vector $\lambda \in \mathbb{R}^{|X|}$ and $k \in \mathbb{R}$ such that $\lambda \cdot \sigma \geq k$ for all σ such that $\theta(\sigma) \geq \theta_{n+1}$, and $\lambda \cdot \sigma < k$ for all σ such that $\theta(\sigma) < \theta_{n+1}$. From Proposition 5(ii), for any $\sigma \in \triangle X^*$, we have $\theta(\sigma) \geq \theta_{n+1}$ and hence $\lambda \cdot \sigma \geq k$.

Pick an arbitrary solution $\boldsymbol{\sigma}$ to the differential inclusion. Pick any time t such that $\theta(\boldsymbol{\sigma}(t)) \in (\theta_n, \theta_{n+1})$. Then we have

$$\dot{\boldsymbol{\sigma}}(t) = \sigma - \boldsymbol{\sigma}(t) \tag{28}$$

for some $\sigma \in \triangle X^*$, and also we have

$$\lambda \cdot \dot{\boldsymbol{\sigma}}(t) = \lambda \cdot (\sigma - \boldsymbol{\sigma}(t)) \geq k - \lambda \cdot \boldsymbol{\sigma}(t) > 0. \tag{29}$$

The first equation (28), together with Proposition 5(ii), implies that $\theta(\boldsymbol{\sigma}(t))$ weakly increases as time goes for all these t. That is, if $\theta(\boldsymbol{\sigma}(t)) \in (\theta_n, \theta_{n+1})$ in the current time t, then we have $\theta(\boldsymbol{\sigma}(t+\eta)) \geq \theta(\boldsymbol{\sigma}(t))$ at the next instant $t + \eta$. The second equation (29) implies that $\lambda \cdot \dot{\boldsymbol{\sigma}}(t)$

strictly increases as time goes. So $\boldsymbol{\sigma}(t)$ moves toward the hyperplane H if $\theta(\boldsymbol{\sigma}(t)) \in (\theta_n, \theta_{n+1})$ in the current time t.

These observations immediately imply the result we want. Pick an arbitrary $\varepsilon > 0$, and let $\tilde{\varepsilon} > 0$ be such that $\theta(\sigma) > \theta_{n+1} - \varepsilon$ for all σ such that $\lambda \cdot \sigma > k - \tilde{\varepsilon}$. Pick T large enough that

$$\tilde{\varepsilon} T > k - \lambda \cdot \sigma \tag{30}$$

for all σ. From (29), if $\lambda \cdot \boldsymbol{\sigma}(t) > k - \tilde{\varepsilon}$ in the current period t, we have $\lambda \cdot \dot{\boldsymbol{\sigma}}(t) \geq \tilde{\varepsilon}$ that is, $\lambda \cdot \boldsymbol{\sigma}(t)$ increases at a rate at least $\tilde{\varepsilon}$. Then from (30), given any initial value $\boldsymbol{\sigma}(0)$ with $\theta(\boldsymbol{\sigma}(0)) > \theta_n$, there is $t < T$ such that $\lambda \cdot \boldsymbol{\sigma}(t) > k - \tilde{\varepsilon}$, which implies $\theta(\boldsymbol{\sigma}(t)) > \theta_{n+1} - \varepsilon$. Also (28) implies that after this time t, $\theta(\boldsymbol{\sigma}(\tilde{t}))$ cannot fall below $\theta_{n+1} - \varepsilon$, that is, $\theta(\boldsymbol{\sigma}(\tilde{t})) > \theta_{n+1} - \varepsilon$ for all $\tilde{t} > t$. This implies the result, because $t < T$.

Case 2: $\theta(\delta_x) < \theta_{n+1}$ for some $x \in X^*$. Let $X^{**} = \{x_1, x_2, \cdots, x_M\}$ denote the set of all $x \in X^*$ such that $\theta(\delta_x) < \theta_{n+1}$. For each action x_m, let ξ_m denote the maximal value of $\theta \in (\theta_n, \theta_{n+1})$ such that $x_m \in F(\delta_\theta)$. Note that the maximum exists, because F is upper hemi-continuous. Also, by the assumption, we have $\xi_m < \theta(\delta_{x_m})$. Without loss of generality, assume that $\theta_n < \xi_1 \leq \xi_2 \cdots \leq \xi_M < \theta_{n+1}$.

Then we can show that there is T_1 such that given any initial value $\boldsymbol{\sigma}(0)$ with $\theta(\boldsymbol{\sigma}(0)) \in (\theta_1, \xi_1]$ and given any solution $\boldsymbol{\sigma}$ to the differential inclusion, we have $\theta(\boldsymbol{\sigma}(t)) > \xi_1$ for some time $t < T_1$. The proof is very similar to the argument in the previous case: Let λ_1 and k_1 be such that $\lambda_1 \cdot \sigma \geq k_1$ for all $\theta(\sigma) \geq \theta(\delta_{x_1})$ and $\lambda_1 \cdot \sigma < k_1$ for all $\theta(\sigma) < \theta(\delta_{x_1})$. Then for any t such that $\theta(\boldsymbol{\sigma}(t)) \in (\theta_n, \xi_1]$, we have $\dot{\boldsymbol{\sigma}}(t) = \sigma - \boldsymbol{\sigma}(t)$ for some $\sigma \in \triangle X^*$, and also

$$\lambda_1 \cdot \dot{\boldsymbol{\sigma}}(t) = \lambda_1 \cdot (\sigma - \boldsymbol{\sigma}(t)) > k_1 - \lambda_1 \cdot \boldsymbol{\sigma}(t) > 0.$$

Note that $k_1 - \lambda_1 \cdot \boldsymbol{\sigma}(t)$ is bounded away from zero uniformly in $\boldsymbol{\sigma}(t)$ with $\theta(\boldsymbol{\sigma}(t)) \in (\theta_n, \xi_1]$, because property (i) in Lemma 3 ensures $\theta(\delta_{x_m}) > \xi_m$ for each m. This immediately implies the existence of T_1.

Similarly, there is T_2 such that given any initial value $\boldsymbol{\sigma}(0)$ with $\theta(\boldsymbol{\sigma}(0)) \in (\xi_1, \xi_2]$ and given any solution $\boldsymbol{\sigma}$ to the differential inclusion, we have $\theta(\boldsymbol{\sigma}(t)) > \xi_2$ for some time $t < T_2$. Again the proof is very similar to the argument in Case 1; the only difference is that here we use the fact that the action x_1 is never chosen when $\theta(\boldsymbol{\sigma}(t)) \in (\xi_1, \xi_2]$.

We iterate this process and define T_1, T_2, $\cdots$, T_M. Also, pick an arbitrarily small $\varepsilon > 0$, and let T_{M+1} be such that given any initial value $\boldsymbol{\sigma}(0)$ with $\theta(\boldsymbol{\sigma}(0)) \in (\xi_M, \theta_{n+1})$ and given any solution $\boldsymbol{\sigma}$ to the differential inclusion, we have $\theta(\boldsymbol{\sigma}(t)) > \theta_{n+1} - \varepsilon$ for some time $t <$

T_{M+1}. Then let $T = T_1 + \cdots + T_{M+1}$. This (ε, T) obviously satisfies the property stated in the lemma. $\qquad\square$

The next lemma relates the result in the previous lemma to the motion of $\theta(\boldsymbol{w}(t))$, where $\boldsymbol{w}(t)$ is the actual frequency. It shows that if $\theta(\boldsymbol{w}(t))$ visits the interval (θ_n, θ_{n+1}) infinitely often, then after a long time, $\theta(\boldsymbol{w}(t))$ cannot be less than θ_{n+1}. That is, $\theta(\boldsymbol{w}(t))$ cannot move against the solution to the differential inclusion in the long run.

Lemma 5. *Consider an interval (θ_n, θ_{n+1}) which satisfies property (i) in Lemma 3. Pick a sample path h such that the property stated in Theorem 3 is satisfied and such that $\theta(\boldsymbol{w}(t))$ exceeds θ_n infinitely often, i.e., for any $T > 0$, there is $t > T$ such that $\theta(\boldsymbol{w}(t)) > \theta_n$. Then $\liminf_{t\to\infty} \theta(\boldsymbol{w}(t)) \geq \theta_{n+1}$.*

Proof. The proof is very similar to that of Proposition 2(ii) and is provided in the online appendix. $\qquad\square$

Now we will show that $\theta(\sigma_t)$ converges almost surely. Suppose not, so that we have $\liminf_{t\to\infty} \theta(\sigma_t) < \limsup_{t\to\infty} \theta(\sigma_t)$ with positive probability. Then there is a path h such that the property stated in Theorem 3 is satisfied and such that $\liminf_{t\to\infty} \theta(\sigma_t) < \limsup_{t\to\infty} \theta(\sigma_t)$. Pick such h.

Let (θ_n, θ_{n+1}) be an interval such that the intersection of the interval and $[\liminf_{t\to\infty} \theta(\sigma_t), \limsup_{t\to\infty} \theta(\sigma_t)]$ is non-empty. Assume for now that this interval satisfies property (i) stated in lemma 3. By the definition of h, $\theta(\boldsymbol{w}(t))$ must exceed θ_n infinitely often, so Lemma 5 implies $\liminf_{t\to\infty} \theta(\boldsymbol{w}(t)) \geq \theta_{n+1}$. But this is a contradiction, because it implies that the intersection of (θ_n, θ_{n+1}) and $[\liminf_{t\to\infty} \theta(\sigma_t), \limsup_{t\to\infty} \theta(\sigma_t)]$ is empty.

Likewise, if the interval (θ_n, θ_{n+1}) satisfies property (ii) in Lemma 3, there is a contradiction. Hence we must have $\liminf_{t\to\infty} \theta(\sigma_t) = \limsup_{t\to\infty} \theta(\sigma_t)$ almost surely.

Also, Lemma 5 implies that for each interval (θ_n, θ_{n+1}) which satisfies property (i) in Lemma 3, we have $\lim_{t\to\infty} \theta(\sigma_t) \in (\theta_n, \theta_{n+1})$ with zero probability. Obviously the same is true for each interval (θ_n, θ_{n+1}) which satisfies property (ii). Hence $\lim_{t\to\infty} \theta(\sigma_t) \in \Theta^{**}$ almost surely.

So for the case in which the boundary points $\{0, 1\}$ are equilibrium models, we have $\lim_{t\to\infty} \theta(\sigma_t) \in \Theta^*$. If $\theta = 0$ is not an equilibrium model, then as in Lemma 3, we can show that $\theta(\delta_x) > \theta$ for any model $\theta \in [\theta_0, \theta_1)$ and for any $x \in F(\delta_\theta)$. Then as in Lemma 5, we can show that if a sample path h satisfies the property stated in Theorem 3 and $\theta(\boldsymbol{w}(t)) \in [\theta_0, \theta_1)$ infinitely often, then $\liminf_{t\to\infty} \theta(\boldsymbol{w}(t)) \geq \theta_1$. This immediately implies that σ_t converges to

57

$\theta_0 = 0$ with zero probability. Similarly, if $\theta = 1$ is not an equilibrium model, then σ_t converges to this model with zero probability. Hence the result follows. $\square$

Proof of Proposition 7. It is obvious that (c) implies (b). So in this proof, we will show that (a) implies (c), and (b) implies (a).

Proof that (a) implies (c). Pick an attracting model θ^*, and let $A = \{\sigma \in \triangle F(\delta_{\theta^*}) | \theta(\sigma) = \theta^*\}$. This set is non-empty, because F is upper hemi-continuous in σ and $\theta(\sigma)$ is continuous. We will show that this set A is robustly attracting.

The following notation is useful. Let $\mathscr{X}_1$ be the set of all mixed strategies σ such that $\theta(\sigma) < \theta^*$. From Proposition 5, this set is convex. Similarly, the set $\triangle X \setminus \mathscr{X}_1$ is convex. So there is a hyperplane H_1 which separates these two sets; i.e., there is a vector $\lambda_1 \in \mathbb{R}^{|X|}$ and k_1 such that $\lambda_1 \cdot \sigma < k_1$ for all σ such that $\theta(\sigma) < \theta^*$, and $\lambda_1 \cdot \sigma \geq k_1$ for all σ such that $\theta(\sigma) \geq \theta^*$. Similarly, letting $\mathscr{X}_2$ be the set of all σ such that $\theta(\sigma) > \theta^*$, there is a hyperplane H_2 which separates $\mathscr{X}_2$ and $\triangle X \setminus \mathscr{X}_2$, i.e., there is a vector $\lambda_2 \in \mathbb{R}^{|X|}$ and k_2 such that $\lambda_2 \cdot \sigma < k_2$ for all σ such that $\theta(\sigma) > \theta^*$, and $\lambda_2 \cdot \sigma \geq k_2$ for all σ such that $\theta(\sigma) \leq \theta^*$. (These hyperplanes H_1 and H_2 may or may not coincide.) Let $\mathscr{X}^*$ be the set of all σ such that $\theta(\sigma) = \theta^*$.

We first consider the special case in which $F(\delta_{\theta^*}) = X$, i.e., the agent is indifferent over all actions in the model θ^*. In this case, $A = \mathscr{X}^*$, i.e., the set A is the set of all mixed actions σ with $\theta(\sigma) = \theta^*$. Later on, we will explain how to extend the proof technique to the case with $F(\delta_{\theta^*}) \subset X$.

Case 1: $F(\delta_{\theta^*}) = X$.

Pick $\varepsilon > 0$ as in the definition of attracting models. Then let $\mathscr{X}_\varepsilon$ be the set of all σ such that $|\theta(\sigma) - \theta^*| < \varepsilon$. We show that this set $\mathscr{X}_\varepsilon$ is (a subset of) the basin of attraction. That is, given any initial value $\boldsymbol{\sigma}(0) \in \mathscr{X}_\varepsilon$, any solution $\boldsymbol{\sigma} \in S^\infty_{\boldsymbol{\sigma}(0)}$ to the differential inclusion will enter a neighborhood of the set $A = \mathscr{X}^*$ in finite time and stay there forever.

So pick any initial value $\boldsymbol{\sigma}(0) \in \mathscr{X}_\varepsilon$ and any solution $\boldsymbol{\sigma} \in S^\infty_{\boldsymbol{\sigma}(0)}$. We first show that this solution $\boldsymbol{\sigma}$ never leaves the set $\mathscr{X}_\varepsilon$.

Lemma 6. $\boldsymbol{\sigma}(t) \in \mathscr{X}_\varepsilon$ *for all t, that is, $|\boldsymbol{\sigma}(t) - \theta^*| < \varepsilon$ for all t.*

Proof. Suppose that $\theta(\boldsymbol{\sigma}(t)) \in (\theta^* - \varepsilon, \theta^*)$ for some t. Then we have $\dot{\boldsymbol{\sigma}}(t) = \sigma - \boldsymbol{\sigma}(t)$ for some $\sigma \in \triangle F(\delta_{\theta(\boldsymbol{\sigma}(t))})$. By the definition of ε, we must have $\theta(\sigma) \geq \theta^*$; then from Proposition 5, at the next instant $t + \eta$, we have $\theta(\boldsymbol{\sigma}(t + \eta)) \geq \theta(\boldsymbol{\sigma}(t))$, i.e., $\theta(\boldsymbol{\sigma}(t))$ is weakly increasing in t if $\boldsymbol{\sigma}(t) \in [\theta^* - \varepsilon, \theta^*)$. Similarly, if $\theta(\boldsymbol{\sigma}(t)) \in (\theta^*, \theta^* + \varepsilon)$, then $\theta(\boldsymbol{\sigma}(t))$ is weakly decreasing in t. This implies the result we want. $\square$

58

Next, we show that A is attracting. It suffices to prove the following lemma:

Lemma 7. *For any $\varepsilon > 0$, there is $T > 0$ such that for any initial value $\boldsymbol{\sigma}(0) \in \mathscr{X}_\varepsilon$ and any solution $\boldsymbol{\sigma} \in S_{\boldsymbol{\sigma}(0)}^\infty$, we have $d(\boldsymbol{\sigma}(t), \mathscr{X}^*) < \varepsilon$ for all $t > T$.*

Proof. Suppose that $\theta(\boldsymbol{\sigma}(t)) \in (\theta^* - \varepsilon, \theta^*)$ for some t. Then as shown in the proof of the previous lemma, we have $\dot{\boldsymbol{\sigma}}(t) = \sigma - \boldsymbol{\sigma}(t)$ for some σ such that $\theta(\sigma) \geq \theta^*$. This in turn implies that

$$\lambda_1 \cdot \dot{\boldsymbol{\sigma}}(t) = \lambda_1 \cdot (\sigma - \boldsymbol{\sigma}(t)) \geq k_1 - \lambda_1 \cdot \boldsymbol{\sigma}(t) > 0.$$

Here the weak inequality follows from $\theta(\sigma) \geq \theta^*$, and the strict inequality follows from $\theta(\sigma) < \theta^*$. Note that $k_1 - \lambda_1 \cdot \boldsymbol{\sigma}(t)$ measures the current distance from $\boldsymbol{\sigma}(t)$ to the hyperplane H_1, and $\lambda_1 \cdot \dot{\boldsymbol{\sigma}}(t)$ measures how much $\boldsymbol{\sigma}(t)$ gets closer to the hyperplane H_1 at the next instant, So the equation above implies that $\boldsymbol{\sigma}(t)$ gets closer to H_1 as time goes, and the speed of convergence is bounded away from zero until $\boldsymbol{\sigma}(t)$ enters a neighborhood of H_1.

Similarly, if $\theta(\boldsymbol{\sigma}(t)) \in (\theta^*, \theta^* + \varepsilon)$ for some t, then $\boldsymbol{\sigma}(t)$ gets closer to the hyperplane H_2 as time goes, and the speed of convergence is bounded away from zero until $\boldsymbol{\sigma}(t)$ enters a neighborhood of H_2. This implies the result we want, because the set $A = \mathscr{X}^*$ is the space sandwiched by H_1 and H_2 (formally, $A = \triangle X \setminus (\mathscr{X}_1) \cup \mathscr{X}_2)$). $\qquad\square$

As a last step, we show that the set A is robustly attracting:

Lemma 8. *The set A is robustly attracting.*

Proof. Let H_1' be the set of all σ with $\theta(\sigma) = \theta^* - \frac{\varepsilon}{2}$, and H_2' be the set of all σ with $\theta(\sigma) = \theta^* + \frac{\varepsilon}{2}$. Take a small $\varepsilon^* > 0$ such that $\theta(\sigma) \in (\theta^* - \varepsilon, \theta^*)$ for all σ with $d(\sigma, H_1') < \varepsilon^*$, and such that $\theta(\sigma) \in (\theta^*, \theta^* + \varepsilon)$ for all σ with $d(\sigma, H_2') < \varepsilon^*$. Note that such ε^* exists because $H_1', H_2', \mathscr{X}^*, \{\sigma | \theta(\sigma) = \theta^* - \varepsilon\}$, and $\{\sigma | \theta(\sigma) = \theta^* + \varepsilon\}$ are all compact and disjoint.

Consider any solution to the ε^*-perturbed differential inclusion, and suppose that $\boldsymbol{\sigma}(t) \in H_1'$ for some t, i.e., suppose that $\theta(\boldsymbol{\sigma}(t)) = \theta^* - \frac{\varepsilon}{2}$. Then by the definition of ε^*, $\dot{\boldsymbol{\sigma}}(t) = \sigma - \boldsymbol{\sigma}(t)$ for some σ such that $\theta(\sigma) \geq \theta^*$, which implies that $\theta(\boldsymbol{\sigma}(t))$ moves up at the next instant. Likewise, if $\theta(\boldsymbol{\sigma}(t)) = \theta^* + \frac{\varepsilon}{2}$ for some t, then $\theta(\boldsymbol{\sigma}(t))$ moves down at the next instant. Accordingly, if the initial value is in the set $\{\sigma | \theta^* - \frac{\varepsilon}{2} \leq \theta(\sigma) \leq \frac{\varepsilon}{2}\}$, any solution to the ε^*-perturbed differential inclusion cannot leave this set. This implies the result. $\qquad\square$

Case 2: $F(\delta_{\theta^*}) \subset X$.

59

Pick small $\varepsilon > 0$ as stated in the the definition of attracting models. Without loss of generality, we assume that $F(\delta_{\tilde{\theta}}) \subseteq F(\delta_{\theta^*})$ for all θ such that $|\theta - \theta^*| < \varepsilon$. (Take ε small, if necessary.)

Let $\mathscr{X}_{\varepsilon}$ be as in the previous case. We show that this set $\mathscr{X}_{\varepsilon}$ is (a subset of) the basin of attraction. That is, given any initial value $\boldsymbol{\sigma}(0) \in \mathscr{X}_{\varepsilon}$, any solution $\boldsymbol{\sigma} \in S^{\infty}_{\boldsymbol{\sigma}(0)}$ to the differential inclusion will enter a neighborhood of the set A in finite time and stay there forever. Note that now the set A is a strict subset of $\mathscr{X}^*$.

Pick any initial value $\boldsymbol{\sigma}(0) \in \mathscr{X}_{\varepsilon}$ and any solution $\boldsymbol{\sigma} \in S^{\infty}_{\boldsymbol{\sigma}(0)}$. Then Lemma 6 still holds, that is, $\boldsymbol{\sigma}(t)$ never leaves the set $\mathscr{X}_{\varepsilon}$. Also Lemma 7 still holds, that is, $\boldsymbol{\sigma}(t)$ moves toward to the set $\mathscr{X}^*$ as time goes.

Also, by the definition of ε, we have $F(\triangle\Theta(\sigma)) \subseteq F(\delta_{\theta^*})$ for any $\sigma \in \mathscr{X}_{\varepsilon}$. This implies that at every time t, we have $\dot{\boldsymbol{\sigma}}(t)[x] = -\boldsymbol{\sigma}(t)[x]$ for each $x \notin F(\delta_{\theta^*})$. This implies that $\boldsymbol{\sigma}(t)$ assigns probability zero on any action $x \notin F(\delta_{\theta^*})$ in the limit as $t \to \infty$, and in particular, for any $\varepsilon > 0$, there is T such that $\dot{\boldsymbol{\sigma}}(t)[x] = -\boldsymbol{\sigma}(t)[x]$ for all $x \notin F(\delta_{\theta^*})$ and $t > T$. This and Lemma 7 imply that the set A is an attractor.

Also, we can show that the set A is robustly attracting; the proof is very similar to that of Lemma 8, and hence omitted. $\square$

Proof that (b) implies (a). Pick an arbitrary θ^*, and let $A = \{\sigma \in \triangle F(\delta_{\theta^*}) | \theta(\sigma) = \theta^*\}$. Suppose that A is an attractor. We will show that the model θ^* is attracting.

Let $\mathscr{U}_A$ be the basin of the set A, and let $\boldsymbol{\sigma}(0)$ be such that $\theta(\boldsymbol{\sigma}(0)) < \theta^*$. Then we have the following lemma:

Lemma 9. *For any $\theta \in (\theta(\boldsymbol{\sigma}(0)), \theta^*)$ and for any $\sigma \in F(\delta_{\theta})$, we have $\theta(\sigma) > \theta$.*

Proof. Suppose not, so that there is $\overline{\theta} \in (\theta(\boldsymbol{\sigma}(0), \theta^*)$ and $\overline{\sigma} \in F(\delta_{\overline{\theta}})$ such that $\theta(\overline{\sigma}) \leq \overline{\theta}$. Consider a solution to the differential inclusion $\boldsymbol{\sigma} \in S^{\infty}_{\boldsymbol{\sigma}(0)}$ such that for any time t such that $\theta(\boldsymbol{\sigma}(t)) = \overline{\theta}$, we have $\dot{\boldsymbol{\sigma}}(t) = \overline{\sigma} - \boldsymbol{\sigma}(t)$. Then $\boldsymbol{\sigma}(t) \leq \overline{\theta}$ for all t; by the definition of $\overline{\sigma}$, $\theta(\boldsymbol{\sigma}(t))$ must go down whenever it hits $\theta(\boldsymbol{\sigma}(t)) = \overline{\theta}$. This contradicts with the fact that $\boldsymbol{\sigma}(0)$ is in the basin of attraction. $\square$

Since $F(\delta_{\theta})$ is upper hemi-continuous in θ, there is $\varepsilon > 0$ such that $F(\delta_{\theta}) = F(\delta_{\tilde{\theta}})$ for all $\theta, \tilde{\theta} \in (\theta^* - \varepsilon, \theta^*)$. Then the lemma above implies that for any $\theta \in (\theta^* - \varepsilon, \theta^*)$ and for any $\sigma \in F(\delta_{\theta})$, we have $\theta(\sigma) \geq \theta^*$.

Similarly, we can show that there is $\tilde{\varepsilon} > 0$ such that for any $\theta \in (\theta^*, \theta^* + \tilde{\varepsilon})$ and for any $\sigma \in F(\delta_{\theta})$, we have $\theta(\sigma) \leq \theta^*$. Hence θ^* is attracting. $\square$

Proof of Proposition 8. Only if: Suppose that a model $\theta^* \in (0,1)$ is unstable. Then upper hemi-continuity of F implies that there are pure actions x and $\tilde{x}$ such that $\theta(\delta_x) < \theta^* < \theta(\delta_{\tilde{x}})$ and and $x, \tilde{x} \in F(\delta_{\theta^*})$. Then from Proposition 5, there is a mixture σ^* of these actions x and $\tilde{x}$ such that $\theta(\sigma^*) = \theta^*$. Obviously this σ^* is a mixed equilibrium with $\theta(\sigma^*) = \theta^*$. So it suffices to show that all mixed equilibria with $\theta(\sigma^*) = \theta^*$ are unstable.

Choose $\varepsilon > 0$ as stated in the definition of unstable models. Then as in the proof of Proposition 7 there is a hyperplane which separates mixed actions σ with $\theta(\sigma) \geq \theta^* + \varepsilon$ from others. That is, there is $\lambda_1 \in \mathbb{R}^{|X|}$ and $k_1 \in \mathbb{R}$ such that $\lambda_1 \cdot \sigma \geq k_1$ if and only if $\theta(\sigma) \geq \theta^* + \varepsilon$. Likewise, there is λ_2 and k_2 such that $\lambda_2 \cdot \sigma \geq k_2$ if and only if $\theta(\sigma) \leq \theta^* - \varepsilon$.

Let $\mathscr{U}$ be the set of all σ such that $\theta(\sigma) \in (\theta^* - \frac{\varepsilon}{2}, \theta^* + \frac{\varepsilon}{2})$. Also, choose T sufficiently large so that

$$(k_1 - \lambda_1 \cdot \sigma)T > k_1 - \lambda_1 \cdot \tilde{\sigma} \tag{31}$$

for all σ with $\theta(\sigma) \in (\theta^*, \theta^* + \frac{\varepsilon}{2})$ and for all $\tilde{\sigma}$ with $\theta(\tilde{\sigma}) \in (\theta^*, \theta^* + \frac{\varepsilon}{2})$, and that

$$(k_2 - \lambda_2 \cdot \sigma)T > k_2 - \lambda_2 \cdot \tilde{\sigma} \tag{32}$$

for all σ with $\theta(\sigma) \in (\theta^* - \frac{\varepsilon}{2}, \theta^*)$ and for all $\tilde{\sigma}$ with $\theta(\tilde{\sigma}) \in (\theta^* - \frac{\varepsilon}{2}, \theta^*)$.

We will show that these $\mathscr{U}$ and T satisfy the property stated in the definition of unstable equilibria. This completes the proof, because any mixed equilibrium σ^* with $\theta(\sigma^*) = \theta^*$ is in the interior of $\mathscr{U}$.

The following result is useful:

Claim 6. *For any initial point $\sigma \in \mathscr{U}$ with $\theta(\sigma) \neq \theta^*$ and for any solution $\boldsymbol{\sigma} \in S_\sigma^\infty$ to the differential inclusion, there is $t < T$ such that $\boldsymbol{\sigma}(t) \notin \mathscr{U}$.*

Proof. First, consider the case in which $\theta(\sigma) \in (\theta^*, \theta^* + \frac{\varepsilon}{2})$. Pick an arbitrary path $\boldsymbol{\sigma} \in S_\sigma^\infty$. Suppose that $\theta(\boldsymbol{\sigma}(t)) \in (\theta^*, \theta^* + \frac{\varepsilon}{2})$ in some period t. Then since θ^* is unstable, $\dot{\boldsymbol{\sigma}}(t) = \sigma - \boldsymbol{\sigma}(t)$ for some σ such that $\theta(\sigma) \geq \theta^* + \varepsilon$. Hence

$$\lambda_1 \cdot \dot{\boldsymbol{\sigma}}(t) = \lambda_1 \cdot (\sigma - \boldsymbol{\sigma}(t)) \geq k_1 - \lambda_1 \cdot \boldsymbol{\sigma}(t) > 0$$

where the weak inequality follows from $\theta(\sigma) \geq \theta^* + \varepsilon$, and the strict inequality follows from $\theta(\boldsymbol{\sigma}(t)) < \theta^* + \frac{\varepsilon}{2}$. This implies that $\lambda_1 \cdot \dot{\boldsymbol{\sigma}}(t)$ increases whenever $\theta(\boldsymbol{\sigma}(t)) \in (\theta^*, \theta^* + \frac{\varepsilon}{2})$ in the current period t. Hence there is $t < T$ such that $\lambda_1 \cdot \theta(\boldsymbol{\sigma}(t)) \geq \theta^* + \frac{\varepsilon}{2}$, implying $\boldsymbol{\sigma}(t) \notin \mathscr{U}$. A similar argument applies to the case in which $\theta(\sigma) \in (\theta^* - \frac{\varepsilon}{2}, \theta^*)$. $\square$

Now we will show that $\mathscr{U}$ and T satisfy the property stated in the definition of unstable equilibria. Pick an arbitrary $\sigma \in \mathscr{U}$. There are two cases to be considered.

Case 1: $\theta(\sigma) \neq \theta^*$. Pick any action x. For β close to one, a perturbed mixture $\beta\sigma + (1-\beta)\delta_x$ is still in the set $\mathscr{U}$, and $\theta(\beta\sigma + (1-\beta)\delta_x) \neq \theta^*$. Hence from the claim above, starting from this perturbed mixture $\beta\sigma + (1-\beta)\delta_x$, any solution to the differential inclusion must leave the set $\mathscr{U}$ within time T. So this σ satisfies the property stated in the definition of unstable equilibria.

Case 2: $\theta(\sigma) = \theta^*$. Pick an arbitrary pure action $x \in F(\delta_{\theta^*})$. Since θ^* is unstable, $\theta(\delta_x) \neq \theta^*$. So from Proposition 5(iii), for any β sufficiently close to one, $\theta(\beta\sigma + (1-\beta)\delta_x) \in (\theta^* - \frac{\varepsilon}{2}, \theta^*) \cup (\theta^*, \theta^* + \frac{\varepsilon}{2})$. Hence, from the above claim, starting from this perturbed mixture $\beta\sigma + (1-\beta)\delta_x$, any solution to the differential inclusion must leave the set $\mathscr{U}$ within time T. So this σ satisfies the property stated in the definition of unstable equilibria. $\square$

If: Let $\theta^* \in (0,1)$ be such that $\theta(\delta_x) \neq \theta^*$ for each pure action $x \in F(\delta_{\theta^*})$, there is at least one mixed equilibrium σ^* with $\theta(\sigma^*) = \theta^*$, and all such mixed equilibria are unstable. We will show that the model θ^* is unstable.

Pick an arbitrary unstable equilibrium σ^* with $\theta(\sigma^*) = \theta^*$, and let T and $\mathscr{U}$ be as in the definition of unstable equilibria. Then each point $\sigma \in \mathscr{U}$ satisfies property (i) or (ii) in the definition of unstable equilibria. In particular, $\sigma = \sigma^*$ satisfies property (ii), i.e., starting from a perturbed action frequency $\beta\sigma^* + (1-\beta)\delta_x$, any solution to the differential inclusion must leave the neighborhood $\mathscr{U}$ of σ^* within time T. This is so because σ^* is an equilibrium and never satisfies property (i).

As the following lemma shows, this property implies that θ^* is indeed unstable.

Lemma 10. θ^* *is* unstable.

Proof. We prove by contradiction, so suppose that θ^* is not unstable. Then from the upper hemi-continuity of F, there is $\varepsilon > 0$ and a pure action $x \in F(\delta_{\theta^*})$ such that

    (a) $\theta(\delta_x) > \theta^*$ and $x \in F(\delta_\theta)$ for all $\theta \in (\theta^* - \varepsilon, \theta^*)$, or

    (b) $\theta(\delta_x) < \theta^*$ and $x \in F(\delta_\theta)$ for all $\theta \in (\theta^*, \theta^* + \varepsilon)$.

Pick such ε and x. In what follows, we focus on the case in which this action x satisfies property (a). (The proof for the other case is symmetric, and hence omitted.)

Since σ^* is a mixed equilibrium with $\theta(\sigma^*) = \theta^*$ and since there is no pure action x with $\theta(\delta_x) = \theta^*$, there must be two actions x and $\tilde{x}$ such that $x, \tilde{x} \in F(\delta_{\theta^*})$ and $\theta(\delta_x) < \theta^* < \theta(\delta_{\tilde{x}})$. Pick such x and $\tilde{x}$. Pick $\beta \in (0,1)$ close to one so that $\theta(\beta\sigma^* + (1-\beta)\delta_x) \in (\theta^* - \varepsilon, \theta^*)$.

Then consider the following path $\boldsymbol{\sigma}$ which starts from $\beta\sigma^* + (1-\beta)\delta_x$:

$$\dot{\boldsymbol{\sigma}}(t) = \begin{cases} \delta_{x^*} - \boldsymbol{\sigma}(t) & \text{if } \theta(\boldsymbol{\sigma}(t)) < \theta^* \\ \sigma^* - \boldsymbol{\sigma}(t) & \text{if } \theta(\boldsymbol{\sigma}(t)) = \theta^* \end{cases}.$$

In words, on this path, the share of x^* increases until $\theta(\boldsymbol{\sigma}(t))$ hits θ^*, and after that $\boldsymbol{\sigma}(t)$ moves toward the equilibrium σ^*. Clearly this path solves the differential inclusion with the initial value $\beta\sigma^* + (1-\beta)\delta_x$, and in particular, if β is sufficiently close to one, this path never leaves the neighborhood $\mathscr{U}$ of σ^*. This implies that $\sigma = \sigma^*$ does not satisfy property (ii) in the definition of unstable equilibria, which is a contradiction. $\qquad\square$

Proof of Proposition 9. Let $\mathscr{H}^*$ be the set of sample paths which satisfy the property stated in Theorem 1. Pick any sample path $h \in \mathscr{H}^*$. We will show that $\lim_{t\to\infty}\sigma^t(h)$ exists, and this limit is a pure action equilibrium.

Let $\underline{\theta}(\sigma)$ denote the minimal element of $\Theta(\sigma)$, and let $\overline{\theta}(\sigma)$ denote the maximal element of $\Theta(\sigma)$. We divide the proof into two steps.

Step 1. In this step, we will establish the following lemma, which shows that for any sample path in which the share of the lowest action x_1 does not shrink to zero, the share of x_1 actually converges to one.

Lemma 3. *If there is a sample path $h \in \mathscr{H}^*$ satisfying $\limsup_{t\to\infty}\sigma^t(h)[x_1] > 0$, then*

(i) $\theta(x_1) \in [\theta_0, \theta_1]$, *so x_1 is a pure action equilibrium.*

(ii) *For all sample paths $h \in \mathscr{H}^*$ satisfying $\limsup_{t\to\infty}\sigma^t(h)[x_1] > 0$, we have $\lim_{t\to\infty}\sigma^t(h)[x_1] = 1$.*

Note that part (i) directly follows from part (ii). (If there is $h \in \mathscr{H}^*$ such that σ^t converges to σ^*, then σ^* must be an equilibrium.) So we will prove only part (ii). We will start with two preliminary lemmas–the proof of each lemma is provided in the online appendix. The first lemma partially characterizes the motion of the highest KLD minimizer $\overline{\theta}$ for sample paths in which the share of the highest action x_N does not shrink to zero.

Lemma 4. *Pick any sample path $h \in \mathscr{H}^*$ such that $\limsup_{t\to\infty}\sigma^t(h)[x_N] > 0$. Then there is t^* such that for any $t > t^*$ with $\overline{\theta}(\sigma^t) < \theta(x_N)$, we have $\overline{\theta}(\sigma^t) < \theta(x)$ for some $x \in F(\delta_{\overline{\theta}(\sigma^t)})$.*

The next lemma partially characterize the motion of the lowest KLD minimizer $\underline{\theta}$ for sample paths in which the share of the lowest action x_1 does not shrink to zero.

Lemma 5. *Pick any sample path $h \in \mathcal{H}^*$ such that $\limsup_{t \to \infty} \sigma^t(h)[x_1] > 0$. Then there is t^* such that for any $t > t^*$ such that $\overline{\theta}(\sigma^t) < \theta(x_N)$ and such that $\overline{\theta}(\sigma^t) < \theta(x)$ for some $x \in F(\delta_{\overline{\theta}(\sigma^t)})$, we have $x^{t+1} \neq x_N$.*

Now we will prove a key result in the proof: It shows that for any sample path in which the share of the lowest action x_1 does not shrink to zero, the share of the highest action x_N must converge to zero.

Lemma 6. *For any sample path $h \in \mathcal{H}^*$ with $\limsup_{t \to \infty} \sigma^t(h)[x_1] > 0$, we have $\lim_{t \to \infty} \sigma^t(h)[x_N] = 0$.*

Proof. Suppose not, so that there is a sample path h such that $\limsup_{t \to \infty} \sigma^t(h)[x_1] > 0$ and $\limsup_{t \to \infty} \sigma^t(h)[x_N] > 0$. Pick such h. From Lemmas 4 and 5, there is t^* such that for any $t > t^*$ with $x^{t+1} = x_N$, we have $\overline{\theta}(\sigma^t) = \theta(x_N)$. That is, after a long time, the action x_N can be chosen only in a period in which $\overline{\theta}(\sigma^t) = \theta(x_N)$. Since $\limsup_{t \to \infty} \sigma^t(h)[x_1] > 0$, there is $t^{**} > t^*$ such that $x^{t^{**}+1} = x_1$. Then we must have $\overline{\theta}(\sigma^{t^{**}+1}) < \theta(x_N)$, which in turn implies that $x^{t^{**}+2} \neq x_N$. Iterating this argument shows that $x^{t+1} \neq x_N$ for all $t \geq t^{**}$, which contradicts with $\limsup_{t \to \infty} \sigma^t(h)[x_N] > 0$. $\qquad\square$

The next lemma shows that the same result holds for the second-highest action x_{N-1}:

Lemma 7. *For any sample path $h \in \mathcal{H}^*$ with $\limsup_{t \to \infty} \sigma^t(h)[x_1] > 0$, we have $\lim_{t \to \infty} \sigma^t(h)[x_{N-1}] = \lim_{t \to \infty} \sigma^t(h)[x_N] = 0$.*

To prove Lemma 7, we will use the next two lemmas, which are counterparts to Lemmas 4 and 5. The proofs are omitted, as they are similar to those of Lemmas 4 and 5.

Lemma 8. *Pick any sample path $h \in \mathcal{H}^*$ such that $\limsup_{t \to \infty} \sigma^t(h)[x_{N-1}] > 0$. Then there is t^* such that for any $t > t^*$ with $\overline{\theta}(\sigma^t) < \theta(x_{N-1})$, we have $\overline{\theta}(\sigma^t) < \theta(x)$ for some $x \in F(\delta_{\overline{\theta}(\sigma^t)})$.*

Lemma 9. *Pick any sample path $h \in \mathcal{H}^*$ such that $\limsup_{t \to \infty} \sigma^t(h)[x_1] > 0$. Then there is t^* such that for any $t > t^*$ such that $\overline{\theta}(\sigma^t) < \theta(x_{N-1})$ and such that $\overline{\theta}(\sigma^t) < \theta(x)$ for some $x \in F(\delta_{\overline{\theta}(\sigma^t)})$, we have $x^{t+1} \neq x_{N-1}, x_N$.*

Proof. Now we will prove Lemma 7. Suppose not, so that there is $h \in \mathcal{H}^*$ with $\limsup_{t \to \infty} \sigma^t(h)[x_1] > 0$, $\limsup_{t \to \infty} \sigma^t(h)[x_{N-1}] > 0$, and $\limsup_{t \to \infty} \sigma^t(h)[x_N] = 0$. (Note that $\limsup_{t \to \infty} \sigma^t(h)[x_N] = 0$ follows from Lemma 6.) Pick such h. From Lemmas 8 and 9, there is $t^* > 0$ such that for any $t > t^*$ with $x^{t+1} = x_{N-1}$ or $x^{t+1} = x_N$, we have $\overline{\theta}(\sigma^t) \geq \theta(\theta_{N-1})$. At the same

time, since $\limsup_{t\to\infty} \sigma^t(h)[x_1] > 0$ and $\limsup_{t\to\infty} \sigma^t(h)[x_N] = 0$, there is $t^{**} > t^*$ such that $\overline{\theta}(\sigma^{t^{**}}) < \theta(\theta_{N-1})$. Then $x^{t^{**}+1} \neq x_{N-1}, x_N$, which in turn implies $\overline{\theta}(\sigma^{t^{**}+1}) < \theta(\theta_{N-1})$. Iterating this argument, we can show that $x^{t^{**}+1} \neq x_{N-1}, x_N$ for all $t > t^{**}$. But this contradicts with $\limsup_{t\to\infty} \sigma^t(h)[x_{N-1}] > 0$. $\square$

Using the argument similar to the proof of Lemma 7, we can show that for any sample path $h \in \mathcal{H}^*$ with $\limsup_{t\to\infty} \sigma^t(h)[x_1] > 0$, we have $\lim_{t\to\infty} \sigma^t(h)[x_2] = \cdots = \lim_{t\to\infty} \sigma^t(h)[x_N] = 0$. This implies the result of Lemma 3.

Step 2. In this step, we will show that a similar result to Lemma 3 holds for other actions $x_2, \cdots, x_N$: That is, for any sample path in which the share of x_n does not shrink to zero, the share of x_n actually converges to one.

Lemma 10. *If there is a sample path $h \in \mathcal{H}^*$ satisfying $\limsup_{t\to\infty} \sigma^t(h)[x_n] > 0$, then*

(i) $\theta(x_n) \in [\theta_{n-1}, \theta_n]$, so x_n is a pure action equilibrium.

(ii) For all sample paths $h \in \mathcal{H}^*$ satisfying $\limsup_{t\to\infty} \sigma^t(h)[x_n] > 0$, we have $\lim_{t\to\infty} \sigma^t(h)[x_n] = 1$.

This lemma immediately implies the result in Proposition 9, because for any path h, there must be some n such that $\limsup_{t\to\infty} \sigma^t(h)[x_n] > 0$.

We will prove Lemma 10 only for $n = 2$, because the same argument applies to all higher $n > 2$. Since part (i) is an immediate consequence of part (ii), we will prove part (ii) only. That is, we will show that for any sample path in which the share of the second-lowest action does not shrink to zero, its share actually converges to one.

The following lemma is a counterpart to Lemma 5.

Lemma 11. *Pick any sample path $h \in \mathcal{H}^*$ such that $\limsup_{t\to\infty} \sigma^t(h)[x_2] > 0$. Then there is t^* such that for any $t > t^*$ such that $\theta(x_2) \leq \overline{\theta}(\sigma^t) < \theta(x_N)$ and such that $\overline{\theta}(\sigma^t) < \theta(x)$ for some $x \in F(\delta_{\overline{\theta}(\sigma^t)})$, we have $x^{t+1} \neq x_N$.*

Now we show that for any sample path in which the share of the second-lowest action x_2 does not shrink to zero, the share of the highest action x_N converges to zero.

Lemma 12. *For any sample path $h \in \mathcal{H}^*$ with $\limsup_{t\to\infty} \sigma^t(h)[x_2] > 0$, we have $\lim_{t\to\infty} \sigma^t(h)[x_N] = 0$.*

Proof. Suppose not, so that there is a sample path h such that $\limsup_{t\to\infty} \sigma^t(h)[x_2] > 0$ and $\limsup_{t\to\infty} \sigma^t(h)[x_N] > 0$. Pick such h. From Lemma 3, we have $\lim_{t\to\infty} \sigma^t(h)[x_1] = 0$. From

65

Lemmas 4 and 11, there is t^* such that for any $t > t^*$ with $x^{t+1} = x_N$, we have $\overline{\theta}(\sigma^t) = \theta(x_N)$ or $\overline{\theta}(\sigma^t) < \theta(x_2)$. That is, after a long time, the action x_N can be chosen only in a period in which $\overline{\theta}(\sigma^t) = \theta(x_N)$ or $\overline{\theta}(\sigma^t) < \theta(x_2)$. Since $\limsup_{t \to \infty} \sigma^t(h)[x_2] > 0$, there is $t^{**} > t^*$ such that $x^{t^{**}+1} = x_2$. Obviously we have $\overline{\theta}(\sigma^t) < \theta(x_N)$ for all $t > t^{**}$. This means that for any $t > t^* + 1$ with $x^{t+1} = x_N$, we have $\overline{\theta}(\sigma^t) < \theta(x_2)$, i.e., the action x_N can be chosen only in a period in which $\overline{\theta}(\sigma^t) < \theta(x_2)$.

Pick $\varepsilon > 0$ such that $\limsup_{t \to \infty} \sigma^t(h)[x_2] > 2\varepsilon$. Since $\lim_{t \to \infty} \sigma^t(h)[x_1] = 0$ and $\theta(x_2) < \cdots < \theta(x_N)$, there is $t^{***} > t^{**}$ such that for any $t > t^{***}$ and for any σ^t with $\overline{\theta}(\sigma^t) < \theta(x_2)$, we have $\sigma^t(x_2) > 1 - \varepsilon$. That is, after period t^{***}, the highest action x_N can be chosen only in a period in which the past action frequency is concentrated on the second-lowest action x_2 and its share is greater than $1 - \varepsilon$. In other words, the highest action x_N can be chosen only in a period in which its share is extremely low and less than ε. Without loss of generality, assume that $\frac{1}{t^{***}} < \varepsilon$. Then for any period $t > t^{***}$, the share $\sigma^t(x_N)$ of the highest action x_N cannot exceed 2ε, which is a contradiction. $\qquad\square$

Likewise, we can show that for any sample path $h \in \mathcal{H}^*$ with $\limsup_{t \to \infty} \sigma^t(h)[x_2] > 0$, we have $\lim_{t \to \infty} \sigma^t(h)[x_3] = \cdots = \lim_{t \to \infty} \sigma^t(h)[x_N] = 0$. The proof is very similar to that of Lemma 9, and hence omitted. $\square$

Proof of Proposition 10. $\Delta \cup_{\mu \in \Delta\Theta(\sigma)} F_\beta(\mu) \subseteq \Delta \cup_{\mu \in \Delta\Theta(\sigma)} F_0(\mu)$: Let $\sigma \in \Delta \cup_{\mu \in \Delta\Theta(\sigma)} F_\beta(\mu)$. Fix any x such that $\sigma(x) > 0$. Since $\sigma \in \Delta \cup_{\mu \in \Delta\Theta(\sigma)} F_\beta(\mu)$, there exists $\mu_x \in \Delta\Theta(\sigma)$ such that $x \in F_\beta(\mu_x)$. It suffices to show that $x \in F_0(\mu_x)$. Since $x \in F_\beta(\mu_x)$, for any $x' \in X$,

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

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., $x \in F_0(\mu_x)$.

$\Delta \cup_{\mu \in \Delta\Theta(\sigma)} F_0(\mu) = \cup_{\mu \in \Delta\Theta(\sigma)} \Delta F_0(\mu)$: The direction $\supseteq$ holds trivially, so we only establish $\subseteq$. Let $\sigma \in \Delta \cup_{\mu \in \Delta\Theta(\sigma)} F_0(\mu)$. Fix any x, x' such that $\sigma(x) > 0$. Since $\sigma \in \Delta \cup_{\mu \in \Delta\Theta(\sigma)} F_\beta(\mu)$, there exist $\mu_x, \mu_{x'} \in \Delta\Theta(\sigma)$ such that $x \in F_0(\mu_x)$ and $x' \in F_0(\mu_{x'})$. By weak identification and

the fact that μ_x and $\mu_{x'}$ 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 $x'' \in X$,

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

and so $x' \in F_0(\mu_x)$. Since x' is an arbitrary element in the support of σ, we have shown that there is a common belief μ_x under which any action in the support of σ is optimal.

Finally, if there is a unique KLD minimizer $\theta(\sigma)$, then $\Delta\cup_{\mu\in\Delta\Theta(\sigma)} F_\beta(\mu) = F_\beta(\delta_{\theta(\sigma)})$, and, using the fact that, for all x, $B(x,\cdot,\delta_{\theta(\sigma)}) = \delta_{\theta(\sigma)}$ $Q(\cdot \mid x)$-a.s., it is straightforward to see that $F_\beta(\delta_{\theta(\sigma)}) = F_0(\delta_{\theta(\sigma)})$.

B   Online Appendix

Proof of additional results in the appendix

Proof of Lemma 1. Let P^x denote the probability distribution of the histories $h = (x_t, y_t)_{t=1}^\infty$ when the agent chooses x every period.

Claim 7. *For any $\eta > 0$, there is T such that for any action x,*

$$P^x(\forall t \geq T \forall \theta \quad |L_t(\theta) - K(\theta,x)| < \eta) > 0$$

Proof. Pick any $\eta > 0$. From Lemma 2, $\lim_{T\to\infty} P^x(\forall t \geq T \forall \theta \quad |L_t(\theta) - K(\theta,x)| < \eta) = 1$. This implies the result we want. $\qquad\square$

Now we will prove Lemma 1. Let $L_t(\theta,x) = \frac{1}{t\sigma_t(x)}\sum_{\tau=1}^t 1_{\{x_\tau=x\}} \log \frac{q(y_t|x_t)}{q_\theta(y_t|x_t)}$ be the sample average of the likelihood ratio, where the sample is taken from the periods in which the agent chooses x. Note that we have $L_t(\theta) = \sum_{x\in X} \sigma_t(x) L_t(\theta,x)$.

Pick $\eta > 0$ arbitrarily, and pick T as in the above claim. Let $\mathscr{H}$ be the set of histories h such that $|L_t(\theta,x) - K(\theta,x)| < \eta$ for all x and t such that $t\sigma_t(x) > T$. Then there is $\underline{q} > 0$ such that $P^f(\mathscr{H}) > \underline{q}$ for any initial prior μ_0 and any policy f.

Pick an arbitrary $h \in \mathscr{H}$, and let $\xi > 0$ be such that $\left|\log \frac{q(y|x)}{q_\theta(y|x)} - \log \frac{q(\tilde{y}|x)}{q_\theta(\tilde{y}|x)}\right| < \xi$ for all x, θ, y, and $\tilde{y}$. Then we have

$$|L_t(\theta,x) - K(\theta,x)| < \begin{cases} \eta & \text{if } t\sigma_t(x) > T \\ \xi & \text{otherwise} \end{cases}$$

for all x, θ, and t. This implies that $\sigma_t(x)|L_t(\theta,x) - K(\theta,x)| < \max\{\eta, \frac{T\xi}{t}\}$. So for any $t > T^* \equiv \frac{T\xi}{\varepsilon}$, we have $\sigma_t(x)|L_t(\theta,x) - K(\theta,x)| < \eta$. Hence for any $t > T^*$,

$$|L_t(\theta) - K(\theta, \sigma_t)| \leq \sum_{x \in X} \sigma_t(x)|L_t(\theta,x) - K(\theta,x)| < \eta.$$

Since K and T^* are chosen independently of $h \in \mathcal{H}$, this implies the result we want. $\square$

Proof of Lemma 2: We first prove the following claim:

Claim 8. *Under Assumption 5, for each σ and $\tilde{\sigma}$ such that $\theta(\sigma) > \theta(\tilde{\sigma})$, $K(\theta,\sigma)$ is strictly decreasing with respect to θ in $[\theta(\tilde{\sigma}), \theta(\sigma)]$, and $K(\theta,\tilde{\sigma})$ is strictly increasing with respect to θ in $[\theta(\tilde{\sigma}), \theta(\sigma)]$.*

Proof. Pick σ and $\tilde{\sigma}$ as stated. For each $\beta \in [0,1]$, let $\sigma_\beta = \beta\sigma + (1-\beta)\tilde{\sigma}$.

We will prove only that $K(\theta,\sigma)$ is strictly decreasing with respect to θ on $[\theta(\tilde{\sigma}), \theta(\sigma)]$. Suppose not, so that there is θ', $\theta'' \in [\theta(\tilde{\sigma}), \theta(\sigma)]$ such that $\theta' < \theta''$ and $K(\theta',\sigma) \leq K(\theta'',\sigma)$. We consider the following two cases.

Case 1: $K(\theta',\tilde{\sigma}) \leq K(\theta'',\tilde{\sigma})$. In this case, $K(\theta',\sigma_\beta) \leq K(\theta'',\sigma_\beta)$ for all β, so θ'' cannot be the unique minimizer of $K(\theta,\sigma_\beta)$, i.e., $\theta(\sigma_\beta) \neq \theta''$ for all β. But this is a contradiction, because $\theta(\sigma_\beta)$ is continuous in β and $\theta(\sigma_0) \leq \theta'' \leq \theta(\sigma_1)$.

Case 2: $K(\theta',\tilde{\sigma}) > K(\theta'',\tilde{\sigma})$. Let β' be such that $\theta(\sigma_{\beta'}) = \theta'$. Then we have $K(\theta',\sigma_{\beta'}) < K(\theta'',\sigma_{\beta'})$, which is equivalent to

$$\beta'(K(\theta',\sigma) - K(\theta'',\sigma)) < (1-\beta')(K(\theta'',\tilde{\sigma}) - K(\theta',\tilde{\sigma})).$$

Then for all $\beta \geq \beta'$,

$$\beta(K(\theta',\sigma) - K(\theta'',\sigma)) < (1-\beta)(K(\theta'',\tilde{\sigma}) - K(\theta',\tilde{\sigma})),$$

which implies $K(\theta',\sigma_\beta) < K(\theta'',\sigma_\beta)$. So $\theta(\sigma_\beta) \neq \theta''$ for all $\beta \geq \beta'$. But this is a contradiction, because $\theta(\sigma_\beta)$ is continuous in β and $\theta(\sigma_{\beta'}) < \theta'' < \theta(\sigma_1)$. $\square$

Pick an arbitrary σ^*. We will show that the Kullback-Leibler divergence $K(\theta,\sigma^*)$ is single-peaked in $[\underline{\theta}, \overline{\theta}]$. First, consider the case in which $\theta(\sigma^*) = \underline{\theta}$. Let $\tilde{\sigma} = \sigma^*$, and let σ be such that $\theta(\sigma) = \overline{\theta}$. Then from the claim above, $K(\theta,\sigma^*)$ is strictly increasing with respect to θ in $[\underline{\theta}, \overline{\theta}]$, which implies single-peakedness.

Next, consider the case in which $\theta(\sigma^*) < \underline{\theta}$. Let $\tilde{\sigma} = \sigma^*$, and let σ be such that $\theta(\sigma) = \overline{\theta}$. Then from the claim above, $K(\theta,\sigma^*)$ is strictly increasing with respect to θ in $[\theta(\sigma^*), \overline{\theta}]$.

Similarly, letting $\sigma = \sigma^*$ and $\tilde{\sigma}$ be such that $\theta(\tilde{\sigma}) = \underline{\theta}$, the claim above implies that $K(\theta, \sigma^*)$ is strictly decreasing with respect to θ in $[\underline{\theta}, \theta(\sigma^*)]$. Hence $K(\theta, \sigma^*)$ is single-peaked. $\square$

Proof of Lemma 5. The proof is very similar to that of Proposition 2(ii). Pick (θ_n, θ_{n+1}) and h as stated. Pick an arbitrarily small $\eta > 0$. Then pick $\varepsilon > 0$ such that $\theta(\sigma) > \theta_{n+1} - \eta$ for all σ such that $\|\sigma - \tilde{\sigma}\| < \varepsilon$ for some $\tilde{\sigma}$ with $\theta(\tilde{\sigma}) > \theta_{n+1} - \frac{\eta}{2}$.

From Lemma 4, there is $T > 0$ such that given any initial value $\boldsymbol{\sigma}(0)$ with $\theta(\boldsymbol{\sigma}(0)) > \theta_n$ and given any solution $\boldsymbol{\sigma} \in S^\infty_{\boldsymbol{\sigma}(0)}$ to the differential inclusion,

$$\theta(\boldsymbol{\sigma}(t)) > \theta_{n+1} - \frac{\eta}{2} \tag{33}$$

for all $t \geq T$. Pick such T. Also, pick $\tilde{T}$ large enough that (21) holds for any $t > \tilde{T}$ and for any $s \in [0, 2T]$.

By the assumption, there is $t > \tilde{T}$ such that $\theta(\boldsymbol{w}(t)) > \theta_n$. Pick such t. Then from (21), (33), and the definition of ε, we have $\theta(\boldsymbol{w}(t+s)) > \theta_{n+1} - \eta$ for all $s \in [T, 2T]$. Applying the same argument again, we obtain $\theta(\boldsymbol{w}(t+s)) > \theta_{n+1} - \eta$ for all $s \geq T$, which implies that $\liminf_{t \to \infty} \theta(\boldsymbol{w}(t)) \geq \theta_{n+1} - \eta$. Since η can be arbitrarily small, we obtain the result. $\square$

Proof of Lemmas 4 and 5. To prove these lemmas, we start with the following result, which partially characterizes the motion of $\underline{\theta}$ and $\overline{\theta}$ when the action frequency σ follows the differential inclusion.

Lemma 13. *The following results hold:*

(i) Let σ be such that there is θ^* such that $K(\theta^*, \sigma) < K(\theta, \sigma)$ for all $\theta < \theta^*$ and such that $\theta^* \leq \theta(x)$ for all $x \in F(1_{\theta^*})$. Then for any solution $\boldsymbol{\sigma} \in S^\infty_\sigma$ to the differential inclusion starting from this σ, $\underline{\theta}(\boldsymbol{\sigma}(t)) \geq \theta^*$ for all $t > 0$.

(ii) Let σ be such that there is θ^* such that $K(\theta^*, \sigma) < K(\theta, \sigma)$ for all $\theta > \theta^*$ and such that $\theta^* \geq \theta(x)$ for all $x \in F(\delta_{\theta^*})$. Then for any solution $\boldsymbol{\sigma} \in S^\infty_\sigma$ to the differential inclusion starting from this σ, $\overline{\theta}(\boldsymbol{\sigma}(t)) \leq \theta^*$ for all $t > 0$.

Proof. We will prove only part (i); the proof of part (ii) is symmetric. Pick σ, θ^*, and $\boldsymbol{\sigma}$ as stated in part (i). Let x_n be the smallest action in the set $F(\delta_{\theta^*})$.

Step 1: $\underline{\theta}(\boldsymbol{\sigma}(t)) \geq \theta^*$ for all small $t > 0$.

We will first show that the result holds for small t. Since $\underline{\theta}(\sigma) \geq \theta^*$, we have $F(\triangle\Theta(\boldsymbol{\sigma}(0))) \subseteq \{x_n, \cdots, x_N\}$. Then from the upper hemi-continuity of $\Theta(\sigma)$, there is $\bar{t} > 0$ such that $F(\triangle\Theta(\boldsymbol{\sigma}(t))) \subseteq$

69

$\{x_n, \cdots, x_N\}$ for all $t \in [0, \bar{t}]$. We will show that $\underline{\theta}(\boldsymbol{\sigma}(t)) \geq \theta^*$ for all $t \in (0, \bar{t}]$. Pick $t \in (0, \bar{t}]$. Then there is $\sigma_t \in \triangle\{x_n, \cdots, x_N\}$ such that $\boldsymbol{\sigma}(t) = \frac{1}{1+t}(\sigma + t\sigma_t)$. By the assumption, we have

$$K(\theta^*, \sigma) < K(\theta, \sigma) \quad \forall \theta < \theta^*.$$

Also from the single-peakedness assumption and $\theta^* \leq \theta(x_n) < \cdots < \theta(x_N)$,

$$K(\theta^*, \sigma_t) < K(\theta, \sigma_t) \quad \forall \theta < \theta^*.$$

Taken together,

$$K(\theta^*, \boldsymbol{\sigma}(t)) < K(\theta, \boldsymbol{\sigma}(t)) \quad \forall \theta < \theta^*.$$

This in turn implies that $\underline{\theta}(\boldsymbol{\sigma}(t)) \geq \theta^*$, as desired.

Step 2: $\underline{\theta}(\boldsymbol{\sigma}(t)) \geq \theta^*$ for all $t > 0$.

Now we will show that $\underline{\theta}(\boldsymbol{\sigma}(t)) \geq \theta^*$ for all $t > 0$. Suppose not so that there is $t > 0$ such that $\underline{\theta}(\boldsymbol{\sigma}(t)) < \theta^*$. Let t^* be the infimum of such time t. Note that $t^* \geq \bar{t} > 0$.

Since $\underline{\theta}(\boldsymbol{\sigma}(t)) \geq \theta^*$ for all $t \in (0, t^*)$, we have $F(\triangle\Theta(\boldsymbol{\sigma}(t))) \in \{x_n, \cdots, x_N\}$ for all $t \in (0, t^*)$. Hence there is $\sigma_{t^*} \in \triangle\{x_n, \cdots, x_N\}$ such that $\boldsymbol{\sigma}(t^*) = \frac{1}{1+t^*}(\sigma + t^*\sigma_{t^*})$. Then as in Step 1, we can show that $\underline{\theta}(\boldsymbol{\sigma}(t^*)) \geq \theta^*$. We consider the following two cases:

Case 1: $\underline{\theta}(\boldsymbol{\sigma}(t^*)) > \theta^*$. In this case, from the upper hemi-continuity of $\Theta(\sigma)$, there is $\bar{t}'$ such that $\underline{\theta}(\boldsymbol{\sigma}(t^* + t)) > \theta^*$ for all $t \in [0, \bar{t}')$. This contradicts with the definition of t^*.

Case 2: $\underline{\theta}(\boldsymbol{\sigma}(t^*)) = \theta^*$. In this case, as in Step 1, we can show that $\underline{\theta}(\boldsymbol{\sigma}(t^* + t)) > \theta^*$ for all $t \in [0, \bar{t})$. This contradicts with the definition of t^*. □

Proof of Lemma 4: Pick h as stated. Pick $\varepsilon > 0$ such that $\limsup_{t \to \infty} \sigma^t(h)[x_N] > 2\varepsilon$ and such that for each n with $\theta_n > \theta(x_n)$ and for each σ in the 2ε-neighborhood of $\triangle\{x_1, \cdots, x_n\}$, we have $\bar{\theta}(\sigma) < \theta_n$. Pick $T > 0$ such that $\frac{1}{T+1} < \varepsilon$. Pick t^* such that for any $t > t^*$,

$$\sup_{s \in [0, 2T]} \inf_{\boldsymbol{\sigma} \in S^{\infty}_{\boldsymbol{w}(h)[t]}} |\boldsymbol{\sigma}(s) - \boldsymbol{w}(t+s)| < \varepsilon. \tag{34}$$

We will show that for any $t > t^*$ with $\bar{\theta}(\boldsymbol{w}(t)) < \theta(x_N)$, we have $\bar{\theta}(\boldsymbol{w}(t)) < \theta(x)$ for some $x \in F(\delta_{\bar{\theta}(\boldsymbol{w}(t))})$.

Suppose not, so that there is $t > t^*$ such that $\bar{\theta}(\boldsymbol{w}(t)) < \theta(x_N)$ and $\bar{\theta}(\boldsymbol{w}(t)) \geq \theta(x)$ for all $x \in F(\delta_{\bar{\theta}(\boldsymbol{w}(t))})$. Let x_n be the largest action in the set $F(\delta_{\bar{\theta}(\boldsymbol{w}(t))})$. Then we have $\theta(x_n) <$

θ_n. Note also that $x_n \neq x_N$, because otherwise $\overline{\theta}(w(t)) = \theta(x_N)$, which contradicts with the definition of t.

Pick any solution σ to the differential inclusion starting from this $w(t)$. Let $\theta^* = \overline{\theta}(\sigma^t)$. Then from Lemma 13(ii), we have $\overline{\theta}(\sigma(s)) \leq \theta^*$ for all s. This implies that $F(\triangle\Theta(\sigma(s))) \subseteq \{x_1, \cdots, x_n\}$ for all s.

So for each s, there is $\sigma_s \in \triangle\{x_1, \cdots, x_n\}$ such that $\sigma(s) = \frac{1}{1+s}(\sigma(0) + s\sigma_s)$. Then by the definition of T, $d(\sigma(s), \triangle\{x_1, \cdots, x_n\}) < \varepsilon$ for all s. Then from (34),

$$d(w(t+s), \triangle\{x_1, \cdots, x_n\}) < 2\varepsilon \quad \forall s \in [T, 2T]. \tag{35}$$

Now consider a solution σ' to the differential inclusion starting from $w(t+T)$. From (35) and the definition of ε, we have $\overline{\theta}(w(t+T)) < \theta_n$. Then again from Lemma 13(ii), we have $\overline{\theta}(\sigma'(s)) \leq \overline{\theta}(w(t+T)) < \theta_n$ for all s, which in turn implies $F(\triangle\Theta(\sigma'(s))) \subseteq \{x_1, \cdots, x_n\}$ for all s. Then as in the previous argument, we can show that

$$d(w(t+s), \triangle\{x_1, \cdots, x_n\}) < 2\varepsilon \quad \forall s \in [2T, 3T].$$

Iterating this argument, it follows that

$$d(w(t+s), \triangle\{x_1, \cdots, x_n\}) < 2\varepsilon \quad \forall s \geq T.$$

But this is a contradiction, as $\limsup_{t\to\infty} \sigma^t(h)[x_N] > 2\varepsilon$. $\square$

Proof of Lemma 5: Pick h as stated. Pick $\varepsilon > 0$ such that $\limsup_{t\to\infty} \sigma^t(h)[x_1] > 2\varepsilon$ and such that for any n with $\theta(x_n) > \theta_{n-1}$ and for any σ in the 2ε-neighborhood of $\triangle\{x_n, \cdots, x_N\}$, we have $\underline{\theta}(\sigma) > \theta_{n-1}$. Pick T such that $\frac{1}{1+T} < \varepsilon$. Pick t^* such that (34) holds for all $t > t^*$.

We will show that for any $t > t^*$ such that $\overline{\theta}(w(t)) < \theta(x_N)$ and such that $\overline{\theta}(w(t)) < \theta(x)$ for some $x \in F(\delta_{\overline{\theta}(w(t))})$, there is no $\overline{t} > 0$ such that $w(t+\overline{t}) = \frac{1}{1+\overline{t}}(w(t)+\overline{t}\delta_{x_N})$.

Suppose not, so that there are $t > t^*$ and $\overline{t} > 0$ such that $\overline{\theta}(w(t)) < \theta(x_N)$, $\overline{\theta}(w(t)) < \theta(x)$ for some $x \in F(\delta_{\overline{\theta}(w(t))})$, and $w(t+\overline{t}) = \frac{1}{1+\overline{t}}(w(t)+\overline{t}\delta_{x_N})$.

Set $\theta^* = \overline{\theta}(w(t))$. Then by the definition of the Kl minimizer, we have

$$K(\theta^*, w(t)) \leq K(\theta, w(t)) \quad \forall \theta < \theta^*.$$

Also, by the single-peakedness assumption and $\theta^* < \theta(x_N)$,

$$K(\theta^*, \delta_{x_N}) < K(\theta, \delta_{x_N}) \quad \forall \theta < \theta^*.$$

71

Taken together,

$$K(\theta^*, w(t+\bar{t})) < K(\theta, w(t+\bar{t})) \quad \forall \theta < \theta^*.$$

We will consider the following two cases separately:

Case 1: $\theta^* = \overline{\theta}(w(t)) < \theta(x)$ for all $x \in F(\delta_{\theta^*})$. Consider a solution σ to the differential inclusion starting from $w(t+\bar{t})$. Then Lemma 13(i) implies that $\underline{\theta}(\sigma(t)) \geq \theta^*$ for all t, which in turn implies that $F(\triangle\Theta(\sigma(t))) \subseteq \{x_n, \cdots, x_N\}$. where x_n is the smallest action in the set $F(\delta_{\theta^*})$. Then as in the proof of the previous lemma, we can show that

$$d(w(t+\bar{t}+s), \triangle\{x_n, \cdots, x_N\}) < 2\varepsilon \quad \forall s \geq T.$$

But this is a contradiction, as $\limsup_{t \to \infty} \sigma^t(h)[x_1] > 2\varepsilon$.

Case 2: $\theta^* = \overline{\theta}(w(t)) \geq \theta(x)$ for some $x \in F(\delta_{\theta^*})$ and $\theta^* < \theta(x)$ for some $x \in F(\delta_{\theta^*})$. In this case, $\theta^* = \theta_{n-1}$ for some n, and $\theta(x_{n-1}) \leq \theta^* < \theta(x_n)$. Pick such n.

Recall that

$$K(\theta^*, w(t+\bar{t})) < K(\theta, w(t+\bar{t})) \quad \forall \theta < \theta^*.$$

Since $\frac{\partial K(\theta^*, w(t))}{\partial \theta} = 0$ and $\frac{\partial K(\theta^*, \delta_{x_N})}{\partial \theta} < 0$, we have $\frac{\partial K(\theta^*, w(t+\bar{t}))}{\partial \theta} < 0$, which in turn implies that there is $\theta^{**} \in (\theta^*, \min\{\theta_n, \theta(x_n)\})$ such that

$$K(\theta^{**}, w(t+\bar{t})) < K(\theta, w(t+\bar{t})) \quad \forall \theta < \theta^{**}.$$

Consider a solution σ to the differential inclusion starting from $w(t+\bar{t})$. Then Lemma 13(i) implies that $\underline{\theta}(\sigma(t)) \geq \theta^{**}$ for all t, which in turn implies that $F(\triangle\Theta(\sigma(t))) \subseteq \{x_n, \cdots, x_N\}$. The rest of the proof is the same as that for Step 1. $\square$

Attracting Sets Need Not Be Robustly Attracting

The agent has three actions, x_1, x_2, and x_3. Given an action x_k, a consequence y is randomly drawn from $Y = \mathbb{R}^3$ according to the normal distribution $N(e_k, I)$, so the action influences the mean of the consequence y. However, the agent does not recognize that the action influences the consequence. Her model space is the probability simplex $\Theta = \triangle X$, and for each model θ, she believes that the consequence follows the normal distribution $N(\theta, I)$. So given a mixture $\sigma \in \triangle X$, the closest model is $\theta = \sigma$, i.e., $\Theta(\sigma) = \{\sigma\}$ for each σ.

For each degenerate belief δ_θ, the optimal policy is given as follows. Consider the model space Θ, and choose the points $A = (\frac{2}{3}, 0, \frac{1}{3})$, $B = (\frac{1}{3}, \frac{2}{3}, 0)$, $C = (0, \frac{1}{3}, \frac{2}{3})$, and $\sigma^* = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ as in Figure 9. For each model θ in the interior of the triangle $AB\sigma^*$, $F(\delta_\theta) = \{x_2\}$, i.e.,

the optimal policy is x_2 if the belief puts probability one on some model θ in this triangle. Similarly, the optimal action is x_3 for the triangle $BC\sigma^*$, and x_1 for the triangle $CA\sigma^*$. For the point σ^* and the models outside the triangle ABC, all actions are optimal, that is, $F(\delta_\theta) = X$ for these models θ. For all models on the boundary of the triangles, the optimal policy is chosen in such a way that $F(\delta_\theta)$ is upper hemi-continuous with respect to θ. For example, on the line $A\sigma^*$, $F(\delta_\theta) = \{x_1, x_2\}$.
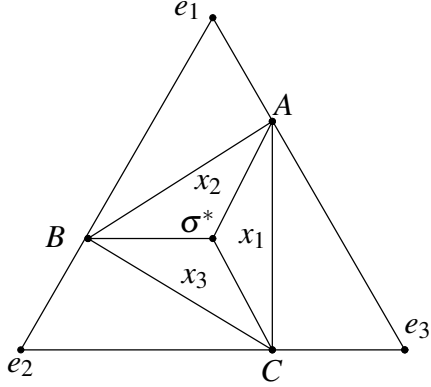

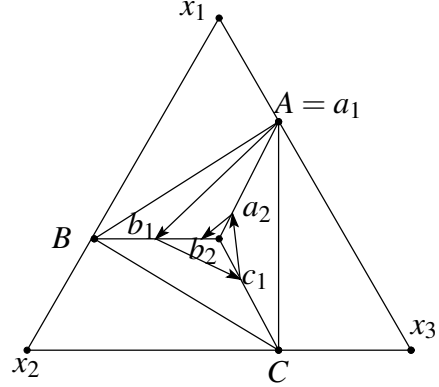Figure 9: Policy $F(\delta_\theta)$ for each model θ

Figure 10: Path starting from a_1.

In this example, the model $\theta = \sigma^*$ is an attracting equilibrium, and its basin of attraction is the interior of the triangle ABC. For example, suppose that the action frequency so far is the point $a_1 = A$, and the action x_2 is chosen today. Then the new action frequency is an interior point of the triangle $AB\sigma^*$, and the agent chooses x_2 until the action frequency hits the point $b_1 = (\frac{1}{3}, \frac{1}{2}, \frac{1}{6})$ on the line $B\sigma^*$. After that, the agent chooses the action x_3 until the action frequency hits the point $c_1 = (\frac{2}{9}, \frac{1}{3}, \frac{4}{9})$ on the line $C\sigma^*$; then the agent chooses the action x_1 until the action frequency hits the point $a_2 = (\frac{5}{12}, \frac{1}{4}, \frac{1}{3})$. From there on, the solution to the differential inclusion takes the path $a_2 b_2 c_2 a_3 b_3 c_3 \cdots$ and converges to σ^*, where

$$a_n = (a_n^1, a_n^2, a_n^3) = \left(1 - \frac{1}{3} - \frac{1}{9c_{n-1}^3}, \frac{1}{9c_{n-1}^3}, \frac{1}{3}\right)$$

$$b_n = (b_n^1, b_n^2, b_n^3) = \left(\frac{1}{3}, 1 - \frac{1}{3} - \frac{1}{9a_n^1}, \frac{1}{9a_n^1}\right)$$

$$c_n = (c_n^1, c_n^2, c_n^3) = \left(\frac{1}{9b_n^2}, \frac{1}{3}, 1 - \frac{1}{3} - \frac{1}{9b_n^2}\right).$$

See Figure 10. Similarly, starting from any interior points of the triangle ABC, any solution σ to the differential inclusion will eventually converge to σ^*.

Now we will modify this example in such a way that the equilibrium σ^* is still attracting

73

but not robustly attracting. Take the points d_0, d_1, $\cdots$ as in Figure 11, that is, d_0 is the intersection point of the line AB and the line passing through σ^* and C, and for each $n \geq 1$, d_n is the intersection point of the line $a_n b_n$ and the line passing through σ^* and C. Then take the sequence $(z_0, z_1, \cdots)$ such that $z_0 = d_1$, $z_1 = (d_1^1, \frac{d_1^2 + d_2^2}{2}, 1 - d_1^1 - \frac{d_1^2 + d_2^2}{2})$, and $z_k = \frac{z_{k-2} + d_2}{2}$ for each $k \geq 2$. Intuitively, $z_0 z_1 \cdots$ is a "jagged bridge" which connects d_1 and d_2, whose step size shrinks as it goes. See Figure 12.
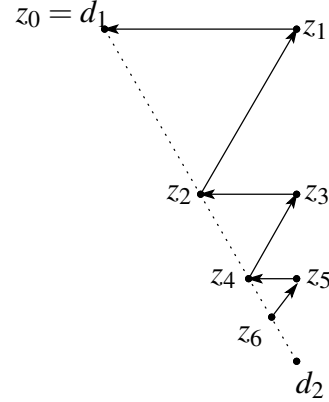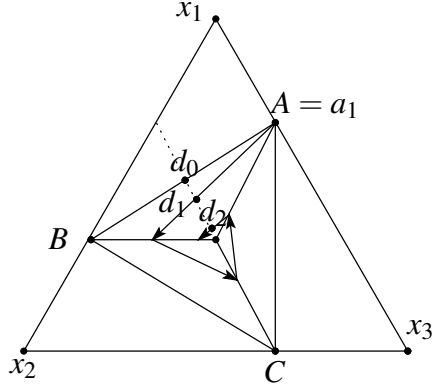

Figure 11: Policy $F(\delta_\theta)$ for each model θ

Figure 12: Jagged path. It does not reach d_2.

Assume that for each model θ on this bridge $z_0 z_1 z_2 \cdots$, the optimal policy is $F(\delta_\theta) = \{x_1, x_2\}$. Then starting from any point on this bridge $z_0 z_1 \cdots$, a solution $\boldsymbol{\sigma}$ to the differential inclusion can move along this bridge and reach the point d_1. However, starting from the point d_2, $\boldsymbol{\sigma}$ cannot move to d_1; this is so because for every large n, z_n is slightly different from d_2, which means that the bridge $z_0 z_1 \cdots$ do not reach the point d_2 exactly. Accordingly, the asymptotic motion of $\boldsymbol{\sigma}$ is the same as before, i.e., as long as the starting point is in the interior of the triangle ABC, $\boldsymbol{\sigma}$ converges to σ^*.

The same is true even if we add more bridges. Suppose that for each n, there is a jagged path from d_n toward d_{n+1}. Even with this change, σ^* is still attracting, for example, starting from the point b_1, $\boldsymbol{\sigma}$ must follow the path $b_1 c_1 a_2 b_2 c_2 \cdots$ and eventually converge to σ^*.

However, adding these bridges significantly changes the solution $\tilde{\boldsymbol{\sigma}}$ to the perturbed differential inclusion. Indeed, starting from the point d_n, $\tilde{\boldsymbol{\sigma}}$ can move to d_{n-1} through the jagged path, because this path is ε-close to the point d_n for any small ε. For the same reason, $\tilde{\boldsymbol{\sigma}}$ can move to d_{n-2}, d_{n-3}, $\cdots$, and can eventually reach the point d_0, which is outside of the basin of σ^*. This implies that σ^* is not robustly attracting with these bridges.

74

Proof of sufficient condition for robustly attracting set

It is obvious that A is attracting, so we will show that the condition stated in the definition of robustly attracting sets is satisfied. Pick $\zeta > 0$ such that $B_{2\zeta}(A)$ is in the set U defined above. Let $C = \overline{B_{2\zeta}(A)} \setminus B_\zeta(A)$. Since C is compact, $(\hat{\sigma} - \sigma) \cdot \nabla V(\sigma) < 0$ is bounded away from zero uniformly. Then Lipchitz-continuity of ∇V ensures that there is $\varepsilon > 0$ such that $(\hat{\sigma} - \tilde{\sigma}) \cdot \nabla V(\tilde{\sigma}) < 0$ for all $\sigma \in C$, $\tilde{\sigma} \in B_\varepsilon(\sigma)$, and $\hat{\sigma} \in F(\triangle \Theta(\sigma))$. This implies that any solution to the ε-perturbed differential inclusion (11) also has a contraction property in the interior of the set C; i.e., if the current action frequency $\tilde{\sigma}$ is an interior point of C and $d(\tilde{\sigma}, C) \geq \varepsilon$, then at the next instant, the action frequency becomes closer to the set A. This immediately implies that A is robustly attracting.

Example: Action frequency converges to non-unitary equilibrium

The consequence space is $Y = \{y_1, y_2, \diamond\}$. We can interpret y_1 and y_2 as states of the world, and $\diamond$ represents a situation where the realization of the state is not observed. There are three actions, x_1, x_2, and $x_\diamond$. The payoffs are $\pi(x_1, y_1) = \pi(x_2, y_2) = 1$, $\pi(x_1, y_2) = \pi(x_2, y_1) = 0$, so this is a problem where the agent wants to match the action x_i with the state y_i. Action $x_\diamond$ is a safe consequence that results in no information about the state, i.e., $Q(\diamond \mid x_\diamond) = 1$, and yields $\pi(x_\diamond, \diamond) = .55$. Action x_2 leads to y_1 for sure and action x_1 leads to y_2 for sure, i.e., $Q(y_1 \mid x_2) = Q(y_2 \mid x_1) = 1$. If the agent knew this information, she would obviously prefer to choose $x_\diamond$. Instead, the agent knows that $x_\diamond$ is a safe action, but incorrectly believes that her choice of x_1 or x_2 does not affect the state, $Q_\theta(y_2 \mid x_1) = Q_\theta(y_2 \mid x_2) = \theta$. We assume that $\theta \in \Theta = \{1/4, 3/4\}$, so that the agent believes that the probability of y_2 is either $1/4$ or $3/4$. Let μ denote the agent's subjective probability that $\theta = 3/4$.

For simplicity, we assume the agent is myopic. Therefore, the agent's optimal policy is $F_0(\mu) = \{x_1\}$ if $\mu < .4$, $F_0(\mu) = \{x_\diamond\}$ if $\mu \in (.4, .6)$, $F_0(\mu) = \{x_2\}$ if $\mu > .6$, with the agent being indifferent between $\{x_1, x_\diamond\}$ at $\mu = .4$ and between $\{x_\diamond, x_2\}$ at $\mu = .6$. The KLD function is

$$K(\theta, \sigma) = \sigma(x_1) \ln \frac{1}{\theta} + \sigma(x_2) \ln \frac{1}{1-\theta} + \sigma(x_\diamond) \ln \frac{1}{1}.$$

Since $K(1/4, \sigma) < (=) > K(3/4, \sigma)$ for $\sigma(x_1) < (=) > \sigma(x_2)$, it follows that

$$\Theta(\sigma) = \begin{cases} \{1/4\} & \text{if } \sigma(x_1) < \sigma(x_2) \\ \{1/4, 3/4\} & \text{if } \sigma(x_1) = \sigma(x_2) \\ \{3/4\} & \text{if } \sigma(x_1) > \sigma(x_2) \end{cases}.$$

In the unique Berk-Nash equilibrium, the agent chooses $x_\diamond$ with probability 1, $\sigma(x_\diamond) = 1$. This corresponds to a situation where the agent stops experimenting with actions x_1 and x_2 and settles for the safe action. Note that this is a Berk-Nash equilibrium because both θ_1 and θ_2 minimize KLD given $\sigma(x_\diamond) = 1$, and $x_\diamond$ is optimal given $\mu \in [.4, .6]$.

There are also a continuum of equilibria that are not Berk-Nash equilibria. These are all the profiles $\sigma^p = (p, p, 1 - 2p)$, $p \in (0, 1/2]$. Note that given σ^p, both θ_1 and θ_2 minimize KLD. By definition of equilibrium, we are free to choose a different belief with support $\{\theta_1, \theta_2\}$ to justify each of the actions. The reason is that these equilibria are not Berk-Nash equilibria is that there is no single belief that supports all three actions.

Starting with initial prior $\mu_0 \in (0, 1)$, the fact that the state is deterministic for a given action implies that the dynamics of this problem can be easily characterized without referring to the results in this paper. In particular, for some priors, the agent's action converges to $x_\diamond$, but for other priors it converges to one of the equilibria that are not Berk-Nash equilibria, $\sigma^{1/2} = (1/2, 1/2, 0)$.[44]

---

[44]In this case, the agent eventually cycles between choosing x_1, observing y_2 and moving the posterior to a region where x_2 is optimal, then observing y_1 and moving the posterior back to the place where x_1 was optimal, and so on. If we relax the assumption that the state is deterministic, then we can show that there are priors such that the action frequency converges with positive probability to the set of non Berk-Nash equilibria, rather than to a single equilibrium.