Limits Points of Endogenous Misspecified Learning *

Drew Fudenberg†  Giacomo Lanzani‡  Philipp Strack §

First posted version: March 10, 2020
This version: April 13, 2020

Abstract

We study how a misspecified agent learns from endogenous data when their prior belief does not impose restrictions on the distribution of outcomes, but can assign probability 0 to a neighborhood of the true model. We characterize the stable actions, which have a very high probability of being the long-run outcome for some initial beliefs, and the are positively attracting actions, which have positive probability of being the long-run outcome for any initial full support belief. A Berk-Nash equilibrium is uniformly strict if the equilibrium action is a strict best response to all the outcome distributions that minimize the Kullback-Leibler divergence from the truth, and uniform if the action is a best response to all those distributions. Uniform Berk-Nash equilibria are the unique possible limit actions under a myopic policy. All uniformly strict Berk-Nash equilibria are stable. They are positively attractive under causation neglect, where the agent believes that their action does not influence the outcome, and under correlation neglect, where the agent believes that the outcome distribution associated with one action does not convey information about the outcomes associated with others. Some results generalize to settings where the agent observes a signal before acting.

*We thank Demian Pouzo, Ignacio Esponda, Ben Golub, Kevin He, Mira Frick, Roberto Corrao, Ryota Iijima, Yuhta Ishi, and Yuichi Yamamato for helpful conversations, and National Science Foundation grant SES 1643517 and the Guido Cazzavillan Scholarship for financial support.

†Department of Economics, MIT
‡Department of Economics, MIT
§Department of Economics, Yale University
1 Introduction

We study the joint evolution of an agent’s actions and beliefs when their action can influence the distribution of outcomes, and their prior may be misspecified in the sense that it assigns probability 0 to a neighborhood of the true model. We assume that the sets of actions and outcomes are finite, but do not restrict the agent’s prior beliefs to a finite number of states or any specific functional form. In this general environment, the agent’s prior is a belief over the set of action-contingent outcome distributions, and the agent is misspecified if they assign probability 0 to a neighborhood of the true distribution of outcomes for some action. The agent’s prior also determines how they perceive the correlation between the outcome distributions induced by different actions, which we show is a key determinant of the long-run outcome of the learning process.

Our results characterize the possible limit points of the agent’s action and their stability properties. First, Proposition 1 shows that if the agent plays myopically and play converges to an action \( a \), then it converges to a Berk-Nash equilibrium (Esponda and Pouzo, 2016), and moreover that if the agent’s prior satisfies some regularity conditions then the limit action must be a uniform Berk-Nash equilibrium. Berk-Nash equilibrium requires that the action is myopically optimal against some belief that is supported over the action-contingent outcome distributions that are closest to the one induced by action \( a \). Here the “closest” distributions are those in the support of the agent’s prior belief that minimize the Kullback-Leibler divergence from the true distribution given that the agent plays \( a \); we call these the “KL-minimizers for \( a \).” The stronger concept of uniform Berk-Nash equilibrium requires that the action is a best response to any mixture over the KL minimizers, as opposed to at least one of them.

We then investigate sufficient conditions for two alternative definitions of what it means for an action to be a long-run outcome. We say that an action is stable if play converges to it with high probability for some open set of initial beliefs. Proposition 2 shows that every uniformly strict Berk-Nash equilibrium is stable, regardless of the agent’s discount factor, where “strict” indicates that the action is the strict myopic best response to the agent’s beliefs, and “uniformly” requires that this is true for all of the closest outcome distributions (as opposed to being true for at least one of them). The converse is not true in general, but Proposition 3 shows that under a minor technical condition the only stable equilibria are the uniform Berk-Nash equilibria.

We say that an action is positively attracting if there is positive probability that it is the limit outcome under every optimal policy for every full-support prior belief. Our setup allows
us to model a number of different forms of misspecified learning. In particular, in the case of *causation neglect*, where the agent mistakenly believes that the distribution of outcomes is the same for all actions, and in subjective bandit problems, where the agent mistakenly believes that the outcomes observed when playing one action are uninformative about the outcome distributions induced by others, we obtain partial converses to Proposition 1: All uniformly strict equilibria are positively attractive, meaning that they have positive probability of being the limit outcome from any starting belief. Moreover, in subjective bandit problems that are *weakly identified* (Esponda and Pouzo (2016)) we can relax uniformly strict to strict.

To prove these results, we first extend Diaconis and Freedman (1990)’s result that Bayesian updating is uniformly consistent to the case of misspecified prior beliefs, a fact that may be of use in future work. We use this extension to guarantee that the agent starts to play the equilibrium action with positive probability. Then, we use the stability result of Proposition 2 to show that with positive probability the agent uses the action forever. We also observe that in a supermodular decision problem, extreme uniformly strict equilibria are positively attractive. There the additional structure of the problem lets us dispense with the first step of the proof.

We also generalize our results to a setting in which the agent observes a signal before taking an action. Here too a limit action must be a Berk-Nash equilibrium. Moreover, if the agents ignore the predictive value of the signals, a behavior that we label *signal neglect*, every uniformly strict Berk-Nash equilibrium is positively attractive.

We illustrate how the sorts of misspecification that can be modeled in our environment correspond to incorrect beliefs about the extensive form of a game, and also with an example of a central bank choosing an exchange-rate policy. The effectiveness of its policy depends on the strength of the economy, summarized by the probability of an expansion in each period and the probability of a speculative attack on the domestic currency. One form of misspecification is that the central bank may not assign positive probability to the correct map from the policy adopted to the probability distribution over outcomes. A second is that the central bank may have causation neglect and think that its policy does not influence the probability distribution over outcomes.

1.1 Related Work

Here we briefly discuss the most relevant prior work. We make more detailed comments about the links to the literature after introducing the formal model.

The statistics literature starting with Berk (1966) studies misspecified Bayesian learning
in environments where the distribution (or stochastic process) of outcomes is exogenous. In many economic applications, actions and associated signal distributions aren’t fixed but change endogenously over time depending on an action taken by the agent (Arrow and Green 1973, Nyarko 1991, Esponda 2008). Esponda and Pouzo (2016) define Berk–Nash equilibrium, which relaxes Nash equilibrium by replacing the requirement that players’ beliefs are correct with the requirement that each player’s belief minimizes the Kullback–Leibler divergence to their observations on the support of their prior. They also analyze dynamics when the payoff function is subject to small random shocks.

Heidhues, Koszegi, and Strack (2018) assume that the agent’s subjective belief is updated according to a linear updating rule. Frick, Iijima, and Ishii (2019b) assume that the agent’s prior is concentrated on a finite set of states, and they make a continuity assumption that our model does not satisfy. Esponda, Pouzo, and Yamamoto (2019) use stochastic approximation techniques to study when the action frequencies converge, as opposed to our focus on whether play converges to a fixed action. Fudenberg, Romanyuk, and Strack (2017) give a complete characterization of limit actions and beliefs for possibly non-myopic learners in a binary prior model with Brownian noise in continuous time. Molavi (2019) studies misspecification in a temporary equilibrium model of macroeconomics; his leading example is where agents mistakenly think that some variables have no impact.¹

All the previous papers consider misspecified Bayesian agents. There is also a literature that studies the long-run outcomes under different learning heuristics. Such heuristics are due to misspecification in the sense that the agent is unable to formulate a probabilistic assessment of the data generating process. Many of these heuristic features some form of neglect of relevant elements of the environment similar to the ones we consider in our Section 4 (see, e.g., Tversky and Kahneman, 1973, Rabin and Schrag, 1999, and Jehiel, 2018).

¹Other recent papers studying misspecification in specific examples of single agent decision problems with endogenous data are Gagnon-Bartsch, Rabin, and Schwartzstein (2018), He (2019), and Heidhues, Kőszegi, and Strack (2018). Misspecification has also been studied in settings of social learning, where agents learn from observations generated by others as in Bohren (2016), and Bohren and Hauser (2018), and Frick, Iijima, and Ishii (2019a).
2 The Model

2.1 Setup

Actions, Utilities and Objective Outcome Distributions  We consider a discrete
time problem: In each period $t \in \{1, 2, 3, \ldots\}$ an agent chooses an action from the finite set
$A$. This choice has two effects. First, each action $a \in A$ induces an objective probability
distribution $p_a^* \in \Delta(Y) \subset \mathbb{R}^{|Y|}$ over the finite set of possible outcomes $Y$. Second, the
action, paired with the realized outcome, determines the flow payoff of the agent via the
utility function $u : A \times Y \to \mathbb{R}$.

Formally, we consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The sample space $\Omega = (Y^\infty)^A$
consists of infinite sequences of action dependent outcome realizations $(x_{a,1}, x_{a,2}, \ldots)_{a \in A}$,
and $x_{a,k}$ determines the outcome when the agent takes the action $a$ for the $k$-th time. $\mathcal{F}$ is
the product sigma algebra and the probability measure $\mathbb{P}$ is the product measure induced
by independent draws from the relevant component of $p^*$. We denote the outcome observed
by the agent in period $t$ after action $a_t$ by $y_t = x_{a_t,k}$, where $k = |\{\tau \leq t : a_\tau = a_t\}|$ is the
number of times the agent has taken action $a_t$ up to and including period $t$.

Subjective Beliefs of the Agent  The agent correctly believes that the map from actions
to probability distributions over outcomes is fixed and depends only on their current action,
but they are uncertain about the distribution each action induces. Let $P = \times_{a \in A} \Delta(Y) \subset
\mathbb{R}^{|Y| \times |A|}$ be the space of all action-dependent outcome distributions, which we call the agent’s
models, and let $p_a \in \Delta(Y)$ denote the $a$-th component of $p \in P$. We endow $P$ with the sup-
norm topology. The agent’s uncertainty is captured by a prior belief $\mu_0 \in \Delta(P)$, where $\Delta(P)$
denotes the metric space of Borel probability measures on $P$ endowed with the Prokhorov
metric, so that it has the topology of weak convergence of measures. Here the support of $\mu_0$
is the set of distributions over outcomes that the agent thinks are possible. We call these
the conceivable outcome distributions, and denote them as $\Theta = \text{supp} \mu_0$. We do not require
that the agent’s model is correctly specified, i.e. that the true outcome distribution $p^*$ is
conceivable. The only restriction that we impose is that if $p$ is conceivable, then no finite
number of observations eliminate it from the support of the agent’s beliefs. Formally, we
will maintain the following assumption:

---

$^2$We endow $A$, as well as any other finite set, with the discrete topology.

$^3$Throughout we denote objective distributions with a superscript $\ast$. 
Assumption 1. For all $p \in \Theta$ and $\varepsilon > 0$, there exists $p' \in \Theta$ with $||p' - p|| < \varepsilon$ such that for all $a \in A$, if $p^*_a(y) > 0$ then $p'_a(y) > 0$.

Our specification allows the agent’s subjective uncertainty to be correlated across actions. For example, under causation neglect, the agent has a belief about action-contingent distributions that is perfectly correlated: they are certain that every action generates the same outcome distribution.

Updating of Subjective Beliefs We assume throughout that the agent updates their beliefs using Bayes rule. Denote by $\mu_t(\cdot | (a^t, y^t))$ the subjective belief the agent obtains using Bayes rule after action sequence $a^t$ and signal sequence $y^t$. Since the agent’s prior has support $\Theta$, any posterior belief $\mu_t(\cdot | (a^t, y^t))$ lies in the set $\Delta(\Theta)$. Moreover from Bayes rule we have that for every $C \subseteq \Theta$

$$
\mu_t(C | (a^t, y^t)) = \frac{\int_{p \in C} \prod_{\tau=1}^{t} p_{ac}(y_\tau) d\mu_0(p)}{\int_{p \in P} \prod_{\tau=1}^{t} p_{ac}(y_\tau) d\mu_0(p)}.
$$

(Bayes Rule)

We sometimes suppress the dependence of the posterior belief on the realized sequence and just write $\mu_t$.

Behavior of the Agent A (pure) policy $\pi : \bigcup_{t=0}^{\infty} A^t \times Y^t \to A$ specifies an action for every history $(a_1, y_1, a_2, y_2, \ldots, a_t, y_t)$, and an initial action $a_1$. We assume that the agent’s objective is to maximize the expected discounted value of per-period utility with discount factor $\beta \in [0, 1)$, and restrict to optimal policies. Throughout, we denote by $a_{t+1} = \pi(a^t, y^t)$ the action taken in period $t$, where $(a^t, y^t) = (a_s, y_s)_{s=1}^t$ is the sequence of realized actions and outcomes. Together, the probability measure $P$ and a policy $\pi$ induce a probability measure $P_\pi$ on $(a_\tau, y_\tau)_{\tau=1}^{\infty}$.

Standard results guarantee that in this setting there is an optimal policy $\pi$ that depends on the history only through the agent’s beliefs, that is,

$$
\mu_t(\cdot | (a^t, y^t)) = \mu_t(\cdot | (a^{t'}, y^{t'})) \implies \pi(a^t, y^t) = \pi(a^{t'}, y^{t'}).
$$
In what follows, we restrict attention to policies that satisfy this restriction.

For every belief \( \nu \in \Delta(P) \) we denote by \( \nu_a \) the marginal distribution associated with action \( a \), i.e. \( \nu_a(C) = \int_{p \in C} d\nu(p) \) for all \( C \subseteq \Delta(\Delta(Y)) \). We denote by \( \mathbb{E}_{p_a} [f(y)] = \sum_{y \in Y} f(y) p_a(y) \) the expectation of \( f : Y \to \mathbb{R} \) under the outcome distribution \( p_a \in \Delta(Y) \). \( A^m(\nu) \) denotes the set of myopic best replies to the belief \( \nu \), i.e.,

\[
A^m(\nu) = \arg\max_{a \in A} \int_{\Delta(Y)} \mathbb{E}_{p_a}[u(a, y)] d\nu_a(p_a). \tag{1}
\]

We say that the agent acts myopically if for all \( (a^t, y^t) \in \bigcup_{\tau=0}^{\infty} A^\tau \times Y^\tau \) we have \( \pi(a^t, y^t) \in A^m(\mu_t(\cdot \mid a^t, y^t)) \).

### 2.2 Forms of Misspecification

Our model encompasses many sorts of misspecified learning. Here we summarize the cases we analyze. We give detailed economic examples of these biases and their implications in Section 6.

#### 2.2.1 Causation Neglect

An agent with causation neglect believes that the realized outcome is not affected by the chosen action. This corresponds to the case in which the prior assigns positive probability only to distributions over outcomes that are independent of the chosen action.

**Definition 1** (Causation Neglect). We say that an agent has causation neglect if for every \( a, a' \in A \) and every \( p \in \Theta \), we have \( p_a = p_{a'} \).

Note that this definition does not say whether the action really does influence the distribution or not; if it does not then it is not a mistake to “neglect” it.

An agent with causation neglect can be misspecified in two ways: Their belief that the action doesn’t change the distribution of outcomes might be incorrect, and the support of their prior belief might not include the true distribution over outcomes for some actions (i.e. there exists \( a \in A \) with \( p_a^* \notin \{p_a \in \Delta(Y) : p \in \Theta\} \)).

An agent with causation neglect updates their belief about the outcome distribution as if they faced an i.i.d. environment. This allows us to use a novel extension of the Diaconis and Freedman (1990) result about uniform consistency with misspecified beliefs to guarantee that the beliefs will concentrate on the conceivable outcome distributions closest to the empirical average. We use this result show that if \( a \) is a uniformly strict Berk-Nash equilibrium, there
is a sequence of outcomes such that the best reply to the induced belief is \( a \), and such that with positive probability the agent sticks to action \( a \) forever. This means that under causation neglect, all uniformly strict Berk-Nash equilibria are positively attractive. Notice that under correlation neglect generically all the Berk-Nash equilibria are uniformly strict.

2.2.2 Subjective Bandit Problems

The other extreme case encompassed by our setup is where the agent thinks that they face a bandit problem, that is that the probability distributions over outcomes induced by different actions are subjectively independent. This corresponds to the case where the agent’s prior \( \mu_0 \) is a product measure.

**Definition 2** (Bandit Problem). We say that an agent faces a **subjective bandit problem** if
\[
\mu_0 = \otimes_{a \in A} \mu_{0,a} \in (\Delta(\Delta(Y)))^A.
\]
Each \( \mu_{0,a} \in \Delta(\Delta(Y)) \) is the agent’s prior about the distribution over outcomes induced by action \( a \).

We use our extension of Diaconis and Freedman \cite{DiaconisF1990} to show that uniformly strict Berk-Nash equilibria are positively attractive in this setting as well, provided that the agent is sufficiently patient.\footnote{The proof shows that if \( b \) is a uniformly strict Berk-Nash equilibrium and the agent is very patient, then there is positive probability that the agent’s beliefs eventually give \( b \) the highest Gittins index. Note that the agent’s discount factor is irrelevant under causation neglect since there the agent does not think there is any information value in experimenting with other actions.}

2.2.3 One Dimensional Parametric Decision Problems

In one-dimensional parametric decision problems, the agent’s uncertainty is summarized by a parameter \( \gamma \in \mathbb{R} \). This restriction makes it easier to rule out cycles and guarantee convergence. The parameter determines the distribution over outcomes through a function \( \phi(\gamma) \) which maps parameters to action-dependent outcome distributions.

**Definition 3** (One-Dimensional Parametric Decision Problems). The decision problem is **one-dimensional parametric** if there exists \( \Gamma \subseteq \mathbb{R} \) and a function \( \phi: \Gamma \to P \) such that \( \Theta \subseteq \{\phi(\gamma): \gamma \in \Gamma\} \). A one-dimensional decision problem is **supermodular** if the actions \( A \) can be ordered such that \( (\gamma, a) \mapsto E_{\phi(\gamma)_a}[u(a, y)] \) is supermodular.\footnote{Formally, if \( a > a' \) and \( \gamma > \gamma' \) then \( E_{\phi(\gamma)_a}[u(a, y)] + E_{\phi(\gamma')_a'}[u(a', y)] \geq E_{\phi(\gamma')_a}[u(a, y)] + E_{\phi(\gamma)_a'}[u(a', y)] \).}

Esponda, Pouzo, and Yamamoto \cite{EspondaPY2019} consider one-dimensional decision problems in which for every action \( a \) the set of action-contingent outcome distributions that minimize the
Kullback-Leibler divergence is a singleton, and the true data generating process corresponds to a value of $\gamma$ that may be outside the support of the agent’s prior. They show that the action process converges almost surely. They also show that, if the decision problem is supermodular and all pure Berk-Nash equilibria are uniformly strict, play converges to one of the Berk-Nash equilibria.

Heidhues, Kôszegi, and Strack (2018) show that in supermodular decision problems where the outcomes are real numbers and $\phi$ is an additive shift, whenever there is a unique Berk-Nash equilibrium play converges there from any starting belief. Our Example 3 shows that their result does not hold in our more general structure: the unique uniformly strict Berk-Nash equilibrium may not be positively attractive. We show that when a stronger version of supermodularity is imposed, our positive attractiveness results can extend to extremal uniformly strict Berk-Nash equilibria.

### 2.2.4 Binary Prior

Another assumption that is commonly made to facilitate the analysis of misspecified models is that of a binary prior, that is $|\Theta| = 2$. Fudenberg, Romanyuk, and Strack (2017) characterizes the long-run behaviour in the case of a binary prior when the outcome is the sum of the chosen action and a Brownian motion. Bohren (2016) and Bohren and Hauser (2018) analyze misspecified binary-prior models in the context of social learning.

### 2.2.5 Finite Support

A final frequently-made assumption when studying misspecified learning is that the support of the prior is finite. Our general setup encompasses this case as well, which allows us to highlight an important difference between environments with finite or infinite support. With a finite-support prior, if behavior converges to an action $a$, $a$ is a uniform Berk-Nash equilibrium. That is, $a$ is a best reply to all the outcome distributions that minimize the Kullback-Leibler divergence from $p_n^\ast$. However, we show by example that non uniform Berk-Nash equilibria can be limit points when the support of the prior is infinite.

Frick, Iijima, and Ishii (2019b) assumes a finite prior, but our setup does not nest theirs as it does not satisfy their condition which implicitly requires a continuum of actions. However, we share their interest in the stability properties of equilibria.
2.2.6 Signal Neglect

Here we suppose that each period the agent observes a signal \( s \in S \) before taking an action \( a \in A \). The signal may convey information about the outcome distribution, and it may also directly enter the payoff function.

We allow the agent to be uncertain about outcome distributions induced by various signals and actions. Let \( P = (\Delta(Y))^{A \times |S|} \subset \mathbb{R}^{Y \times |A| \times |S|} \) be the space of all signal and action dependent outcome distributions. The agent’s belief is a probability measure \( \mu \) over \( P \), where \( p_{s,a}(y) \) denotes the probability under \( p \in P \) of outcome \( y \) after observing signal \( s \) playing action \( a \).

**Definition 4** (Signal Neglect). An agent has signal neglect if for every \( s, s' \in S \) and every \( a \in A \), and \( p \in \Theta \), we have \( p_{s,a} = p_{s',a} \).

As an example, consider a seller who chooses whether to undertake an investment (e.g., hiring an additional worker for the day) after observing the number of people in the market that day. The agent realizes that the signal is payoff relevant, but they do not realize that when fewer consumers show up, a higher fraction of them buy. We show that this form of misspecification will result in underinvestment when market attendance is high. We discuss this example in more detail in Section 6.

3 Limit Points and Berk-Nash Equilibria

We will be interested in when the agent’s actions converge, and what the possible limit points are. Note that these are different questions than whether the agent’s beliefs converge: Beliefs can oscillate when actions are fixed, as in Berk’s example where there the agent doesn’t have an action choice, and conversely actions can oscillate with fixed beliefs if the agent is indifferent. Thus, the agent’s actions might converge without their beliefs converging. Intuitively, if two states explain the observed data equally well on average, the log-likelihood ratio between these two states is a random walk and thus oscillates between assigning high probability to each of the two states. Conversely, if the agent is indifferent between multiple actions at the limit belief, their actions might not converge even though their beliefs do.

---

8More generally the agent might mistakenly think that some components of the signal vector have no effect on outcomes, as in Molavi (2019)

9The possibility that beliefs oscillate with fixed actions is the driving force behind the uniformity requirement in several of our results, see e.g., Proposition 1(ii)
Formally, the action process converges to action \( a \) if there exists a time period \( T \in \mathbb{N} \) such that \( a_t = a \) for all later time periods \( t > T \). We say that the action process converges to \( a \) with positive probability (resp. with probability 1) under policy \( \pi \) if there is a measurable set \( C \subseteq A^x \times Y^x \) with \( \mathbb{P}_\pi[C] > 0 \) (resp. with \( \mathbb{P}_\pi[C] = 1 \)) such that \( a_t \) converges to \( a \) in \( C \). Note that a given prior may be consistent with multiple optimal policies, and which policy is used can influence whether the action process converges and if so to which points.

The concept of Berk-Nash Equilibria (Esponda and Pouzo, 2016) will play a key role in our analysis. Intuitively, a Berk-Nash equilibrium is an action \( a \) such that there exists a belief that makes the action optimal, and which gives positive probability only to the conceivable outcome distributions that best match the objective outcome distribution \( p^*_a \).

Formally, given two distributions over outcomes \( q, q' \in \Delta(Y) \) we define

\[
H(q, q') = -\sum_{y \in Y} q(y) \log q'(y).
\]

Recall that for every \( p, p_a \) denotes the outcome distribution corresponding to action \( a \). Denote the set of conceivable outcome distributions for action \( a \) that best match \( p^*_a \) by

\[
\hat{\Theta}_a(p) = \arg\min_{p_a \in \Theta} H(p^*_a, p_a) \subset \Delta(Y).
\]

And for each action \( a \), let

\[
\hat{\Theta}(a) = \{ p \in \Theta : p_a \in \hat{\Theta}_a(a) \} = \arg\min_{p \in \Theta} H(p^*_a, p_a) \subset P
\]

(2)
denote the set of action-contingent outcome distributions that minimize the Kullback-Leibler divergence between distributions in the support of \( \mu_0 \) and the true distribution \( p^*_a \) given that \( a \) is played.\(^{10}\) Note that the elements of \( \hat{\Theta}(a) \) are elements of \( P \), and so specify an outcome distribution for each action \( a' \in A \), even though which elements of \( P \) are included only depends on the distribution \( p^*_a \) induced by \( a \) and on the outcome distributions that \( \mu \) thinks \( a \) might induce. We call this the set of “KL-minimizers for action \( a \) and belief \( \mu \).”

Note also that if \( p^* \in \Theta \) then \( \hat{\Theta}_a(a) = \{ p^*_a \} \) and each minimizing \( p \) has \( p_a = p^*_a \). In particular this is true if \( \mu \) has full support.

Following Berk (1966), we expect the agent’s beliefs to concentrate on the KL minimizers for action \( a \) and belief \( \mu_0 \) if an agent with prior belief \( \mu_0 \) always plays \( a \). This motivates

\(^{10}\)The Kullback-Leibler divergence between \( p^*_a \) and \( p_a \) is given by \(-\sum_{y \in Y} p^*_a(y) \log(p^*_a(y)) + H(p^*_a, p_a)\) so any \( p_a \) that minimizes \( H(p^*_a, p_a) \) also minimizes the Kullback-Leibler divergence between \( p^*_a \) and \( p_a \).
Esponda and Pouzo (2016)’s notion of a Berk-Nash equilibrium. We will use several variations of this concept to capture various senses in which an action is or is not a long-run outcome of the agent’s learning process.

**Definition 5 (Berk-Nash Equilibrium).**

(i) Action \( a \in A \) is a Berk-Nash equilibrium if for some belief \( \nu \in \Delta(\hat{\Theta}(a)) \), \( a \) is myopically optimal given \( \nu \), i.e. \( a \in A^m(\nu) \).

(ii) Action \( a \) is a strict Berk-Nash equilibrium if for some belief in \( \nu \in \Delta(\hat{\Theta}(a)) \), \( a \) is the unique myopically optimal action, i.e. \( \{a\} = A^m(\nu) \).

(iii) Action \( a \) is a uniform Berk-Nash equilibrium if for all \( q \in \hat{\Theta}_a(a) \) there exists a \( \nu \in \Delta \left( \{p \in \hat{\Theta}(a) : \forall y \in \text{supp} \ p^*_a, p_a(y) = q(y)\} \right) \) such that \( a \in A^m(\nu) \).

(iv) Action \( a \) is a uniformly strict Berk-Nash equilibrium if for every belief in \( \nu \in \Delta(\hat{\Theta}(a)) \), \( a \) is the unique myopically optimal action, i.e., \( \{a\} = A^m(\nu) \).

Playing action \( a \) induces distribution \( p^*_a \) over the observed outcomes, and the set \( \hat{\Theta}_a(a) \) includes all the outcome distributions that best match \( p^*_a \). Uniformity requires that for each \( p_a \in \hat{\Theta}_a(a) \) there exists a belief about the outcome distributions induced by the other actions that justifies \( a \) as a best reply.

**Remark 1.** In settings where the KL-minimizer is unique, the uniformity requirement has no bite. However, in some applications it is natural to have multiple KL minimizers. For example, suppose that the agent is randomly matched with one of \( n \) possible opponents, and that the outcome \( y \) is the action selected by the opponent. Then the elements of \( \Theta \) correspond to the strategies of the different opponents, and the KL-divergence between them is often equal. More generally, we should expect that uniformity has bite in all the setting with some structural symmetry property due to the uniform distribution of some underlying variable (here corresponding to the identity of the matched opponent).

Observe that the difference between Berk-Nash equilibrium and uniform Berk-Nash equilibrium disappears in the correctly specified case, where both concepts coincide with self-confirming equilibrium.

**Assumption 2 (K-positivity).** The prior \( \mu_0 \) is K-positive if there exists \( K > 0 \) such that for every \( p \in \Theta \) and \( \varepsilon > 0 \) we have \( \mu_0(B_\varepsilon(p)) \geq K\varepsilon \).

Assumption 2 is a generalization to our misspecified setting of the \( \phi \)-positivity assumption of Diaconis and Freedman (1990). It is always satisfied if the prior has finite support.

\(^{11}\)Esponda, Pouzo, and Yamamoto (2019) call this a strict equilibrium.
The following result motivates our focus on Berk-Nash equilibria and more particularly on uniform Berk-Nash equilibria.

**Proposition 1 (Limit Actions are Berk-Nash Equilibria).** Suppose the agent uses an optimal myopic policy.

(i) If actions converge to a $a \in A$ with positive probability, then $a$ is a Berk-Nash equilibrium.

(ii) If the prior is $K$-positive and actions converge to a $a \in A$ with positive probability, then $a$ is a uniform Berk-Nash equilibrium.

Proposition 1(i) says that limit actions must be Berk-Nash equilibria. In outline, this follows from the fact that if actions converge to an action then eventually the agent always plays that action, and Berk (1966)’s result that the agent’s beliefs converge to the set of KL minimizers when their observations are a sequence of i.i.d. signals. Esponda and Pouzo (2016), Esponda, Pouzo, and Yamamoto (2019), and Frick, Iijima, and Ishii (2019b) obtain similar results in related settings. Proposition 1(i) is most similar to Proposition 1 of Esponda, Pouzo, and Yamamoto (2019) and would follow from it if we made slightly stronger assumptions. Example 6 shows that this conclusion fails without Assumption 1.

The second part of Proposition 1’s shows that limit actions must be a uniform Berk-Nash equilibrium when the prior is $K$-positive. When $a$ is not a uniform Berk-Nash equilibrium, there is a KL minimizer $p_a \in \hat{\Theta}_a(a)$ such that $a$ is not a myopic best reply. The proof shows that since all the minimizers fit the true outcome generating process $p_a^*$ equally well, the log-likelihood ratio of any two of them is a random walk. We then combine the Central Limit Theorem with the Kochen-Stone Lemma to prove that for every $p_a \in \hat{\Theta}_a(a)$, this random walk will eventually visit a region where the log-likelihood of the outcome distributions that coincide with $p_a$ conditional on $a$ is arbitrarily high, so the agent stops playing $a$.

Example 8 shows that Assumption 2 is needed. Here the agent’s prior has countable support, and assigns vanishingly low probability to distributions that are close to one of the KL minimizers.

We do not know of any results similar to Proposition 1 (ii). Indeed, the only other equilibrium refinement we know of that tests for optimality against all beliefs in a non-singleton set is Fudenberg and He (2020) which also has a foundations in learning theory.

---

12Specifically, our Assumption 1 is weaker than their Assumption 2 (iii). Esponda and Pouzo (2016)’s similar result assumed that the utility function is subject to i.i.d. shocks as in Fudenberg and Kreps (1993); we do not require those shocks here.

13The Kochen-Stone lemma is a sharpening of the Borel-Cantelli lemma.

14Thus the argument extends Berk (1966)’s example where two different conceivable probability distributions are the best explanations of the observed data, and the log-likelihood ratio is a random walk.

15Fudenberg and He (2020) studies a steady-state model where the agents are correctly specified Bayesians;
4 Sufficient Conditions for Long-Run Persistence

Proposition 1 shows that play can only converge to a given action \( a \) if that action is a Berk-Nash equilibrium. This section gives sufficient conditions for two different senses of what it means for an action to be a plausible long run outcome.

4.1 Stability

We say that action \( a \) is “stable” if play is very likely to converge to \( a \) starting from some beliefs in a neighborhood of a KL-minimizer for \( a \). For \( \nu \in \Delta(\Theta) \), let \( B_\varepsilon(\nu) = \{ \nu' \in \Delta(\Theta) | d(\nu', \nu) \leq \varepsilon \} \) be the set of beliefs over conceivable distributions that are within \( \varepsilon \) of \( \nu \).

As a last piece of notation, define the set \( \hat{\Theta}(a) \) as all outcome distributions such that the marginal distribution with respect to action \( a \) is at most \( \varepsilon \) away from a KL minimizer:

\[
\hat{\Theta}(a) = \{ p \in \Theta : \text{there exists } q \in \hat{\Theta}(a) \text{ with } ||q_a - p_a|| \leq \varepsilon \}.
\]

**Definition 6 (Stability).**

(i) A Berk-Nash equilibrium \( a \) is **stable for discount factor** \( \beta \) if there is a belief \( \nu \) such that for every \( \kappa \in (0, 1) \), there exists an \( \varepsilon > 0 \) such that for all prior beliefs in \( B_\varepsilon(\nu) \) the action prescribed by any optimal policy converges to \( a \) with probability larger than \( (1 - \kappa) \).

(ii) A Berk-Nash equilibrium \( a \) is **uniformly stable for discount factor** \( \beta \) if for every \( \kappa \in (0, 1) \), there is an \( \varepsilon > 0 \) such that for all prior beliefs \( \nu \in \Delta(\Theta) \) such that \( \nu(\hat{\Theta}(a)) > 1 - \varepsilon \) the action prescribed by any optimal policy converges to \( a \in A \) with probability greater than \( 1 - \kappa \).

Once again, the uniformity requirement has bite only if the agent is misspecified: if \( p^* \in \Theta \), the actions that are stable are uniformly stable, and they coincide with strict self-confirming equilibria.

The next proposition shows that if a Berk-Nash equilibrium is uniformly strict then convergence to that equilibrium is arbitrarily likely for some beliefs:

**Proposition 2.** Every uniformly strict Berk-Nash equilibrium is uniformly stable for any discount factor \( \beta \).

---

it does not study the dynamics away from the steady state.
Frick, Iijima, and Ishii (2019b) obtain a similar conclusion in a model where the agent’s prior has finite support. This conclusion is sharper than showing there are beliefs under which there is positive probability of converging to the equilibrium, as in Proposition 4 of Esponda, Pouzo, and Yamamoto (2019).

The proof of Proposition 2 first derives a neighborhood of action-dependent outcome distributions that are close to the Kullback-Leibler minimizers such that if the beliefs assign a sufficiently high probability to that neighborhood, the optimal action is the uniformly strict Berk-Nash equilibrium $a$. That such a neighborhood exists for a myopic policy follows from the definition of uniformly strict Berk-Nash equilibrium. Under a non-myopic policy, since beliefs are not degenerate, some actions may have an experimentation value. However, when the beliefs are sufficiently concentrated around the minimizers, the value of any alternative action cannot be much higher then its value against the most favorable minimizer, and since $a$ is a uniformly strict Berk-Nash equilibrium, this value is strictly lower than that of $a$. Then we combine an observation of Frick, Iijima, and Ishii (2019b) together with a generalization of the arguments in Fudenberg and Levine (1992) to the misspecified case, to argue that a transformation of the odds ratio of this neighborhood is a positive supermartingale, under the outcome distribution induced by action $a$. Finally, we use the Dubins’ inequality to guarantee that if the probability initially assigned to the neighborhood is sufficiently high, that probability never crosses the threshold that makes action $a$ suboptimal.

Proposition 2 is in contrast to the non-convergence example of Heidhues, Koszegi, and Strack (2018), who study a model with a continuum of actions and a convex, real-valued set of states. In that model actions that are sufficiently near the strict best response incur arbitrarily small losses, and are best responses to nearby beliefs. As we explain in Section 7 it is not clear what the right definition of uniform stability for this setting.

The next example shows that Proposition 2 does not extend to strict Berk-Nash equilibria that are not uniformly strict.

**Example 1** (A strict but not stable Berk-Nash equilibrium). Suppose there are two actions $a$ and $b$, and two outcomes $Y = \{0, 1\}$. We identify the elements of $\Delta(Y)$ with the probability they assign to outcome 1. Let $p_a^* = \frac{1}{2}$, and suppose that the agent believes that the outcome distribution does not depend on the action, so they have correlation neglect, and let $\Theta = \left[0, \frac{1}{2}\right] \cup \left[\frac{3}{4}, 1\right]$. Suppose that $A^m(\delta_{3/4}) = \{a\}$, and $A^m(\delta_{1/4}) = \{b\}$.

Here $a$ is a strict Berk-Nash equilibrium supported by the belief $\delta_{3/4}$, but it is not uniformly strict. Now fix $\nu \in \hat{\Theta}(a)$. If $\nu = \delta_{1/4}$ it is immediate that the action process cannot converge to a starting from belief $\nu$. Suppose instead that $\nu = i\delta_{1/4} + (1 - i)\delta_{3/4}$ with $i \in [0, 1)$. If we let
\[ \nu_\varepsilon = (1 + \varepsilon(1 - i)) \delta_{1/4} + \varepsilon(1 - i) \delta_{3/4 + \varepsilon}, \text{ then } \argmin_{p \in \text{supp} \nu_\varepsilon} H (p^*_a, p_a) = \{ \frac{1}{3} \} \text{ and } a \not\in A^m \left( \delta_{1/4} \right). \]

Note also that as \( \varepsilon \) goes to 0, \( \nu_\varepsilon \) converges to \( \delta_{3/4} \). Yet by Proposition 2 the probability of converging to a starting from belief \( \nu_\varepsilon \) is 0 for all \( \varepsilon \). Since \( \nu \) was an arbitrary element of \( \hat{\Theta}(a) \), this proves that \( a \) is not stable.

The next result is a partial converse to Proposition 2. It shows a key difference between misspecified and correctly specified environments. In misspecified settings, more than one outcome distribution may be the best explanation of \( p^*_a \). If \( a \) is not a uniform Berk-Nash equilibrium, it is not a best reply for some of the KL-minimizers, and there is positive probability that the agent’s belief will eventually concentrate on a minimizer that leads the agent to switch to an action \( a' \neq a \).

**Proposition 3.** Suppose that \( a \) is not a uniform Berk-Nash equilibrium. If for every \( p \in \hat{\Theta}(a) \) and \( \varepsilon > 0 \) there exists \( q \in B_\varepsilon (p) \), \( q \not\in \hat{\Theta}(a) \), then \( a \) is not stable under a myopic policy.

If \( a \) is not a Berk-Nash equilibrium, the result immediately follows from Proposition 1. When \( a \) is a Berk-Nash equilibrium but not a uniform Berk-Nash equilibrium, there is at least one \( p \in \hat{\Theta}(a) \) for which the best reply is not \( a \). Suppose that the agent’s initial belief \( \nu \) assigns a very low positive probability to this \( p \), and the remaining probability to distributions \( p' \) that are close to distributions in \( \hat{\Theta}(a) \) but are not in \( \hat{\Theta}(a) \). Then, \( \{ p \} = \argmin_{p' \in \text{supp} \nu} H (p^*_a, p'_a) \), and Proposition 1 implies that since \( a \not\in A(\delta_p) \), the probability of converging to \( a \) starting with belief \( \nu \) is zero.

**A Characterization of Berk-Nash Equilibria** Combining the results of this section with Proposition 1 gives a sharp relationship between equilibria and stable outcomes.

**Corollary 1.** If \( \text{supp} p^*_a = Y \) and \( \beta = 0 \), the following are equivalent:

(i) \( a \in A \) is a uniform Berk-Nash equilibrium.

(ii) There exists a belief \( \nu \in \Delta(\Theta) \) with \( \hat{\Theta}_a(a) \subseteq \text{supp} \nu_a \) and an optimal policy such that actions converge to \( a \in A \) with positive probability.

---

16 We could have defined stability with respect to a myopic policy instead of all of them. Under the assumptions of the next proposition, actions that are not uniform Berk-Nash equilibria would fail the stability test even under that more permissive definition.

17 The hypothesis of Proposition 3 excludes finitely supported priors, but in this case Proposition 1 guarantees that the probability of converging to a non-uniform equilibrium is 0. A fortiori, this means that non-uniform equilibria are not stable.
Thus, for myopic agents, uniform Berk-Nash equilibria are the (unique) outcomes that can arise as long-run behavior if the agent starts with a belief that does not rule out any of the conceivable outcome distributions that best fit the true outcome generating process.

**Corollary 2.** For every discount factor $\beta \in [0,1)$ the following are equivalent:

(i) $a \in A$ is a uniformly strict Berk-Nash equilibrium.

(ii) $a \in A$ is uniformly stable for $\beta$.

Theorem 1 of Frick, Iijima, and Ishii (2019b) characterizes the locally stable beliefs for finite-support priors. Because the associated actions and preferences are left implicit it is difficult to compare their results with ours, but as we pointed out earlier, actions can converge even when beliefs do not.

**Example 2** (A Stable Berk-Nash Equilibrium that is not uniformly strict). To see that in Corollary 2 we cannot replace uniformly stable with stable, suppose there are 2 actions $a$ and $b$ that induce the same distribution on $Y = \{0,1\}$. The agent has an independent prior with full support marginal distribution. Suppose also that $a$ is the unique best response to any belief has positive probability on 1, while both $a$ and $b$ are best responses when sure the outcome will be 0. Here if the agent assigns positive probability to every outcome distribution and the true one is not degenerate on outcome 0, then $a$ is not uniformly strict but it is stable.

### 4.2 Positive Attractiveness

The previous section gave sufficient conditions for an action to be played in the long-run with high probability for *some* prior beliefs. Another natural notion of $a$ being a long-run outcome is that for *every* prior belief there is strictly positive probability that the agent’s action converges to $a$.

We say that an action is *positively attracting* if there is positive probability that it is the limit outcome under every optimal policy.

**Definition 7** (Positively Attracting). The action $a \in A$ is **positively attracting** if for every optimal policy $\pi$

\[
P_\pi \left[ \lim_{t \to \infty} a_t = a \right] > 0.
\]

In later sections we give sufficient conditions for uniformly strict Berk-Nash equilibria to be positively attracting. Benaim and Hirsch (1999) prove a similar result for the linearly stable Nash equilibria of stochastic fictitious play, which is based on correctly-specified
Bayesian learning, but positive attractiveness has not previously been studied in the setting of this paper.\footnote{The Bayesian foundation of fictitious play assumes that the players believe that the environment is stationary and have Dirichlet priors. Away from a steady state the players are thus misspecified, but when the system converges to a steady state the stationarity assumption is asymptotically correct. In our setting, “substantial” misspecification can persist even when behavior converges.}

These arguments rely on Lemma in the appendix, which shows that beliefs about the outcome distribution concentrate around the distributions that best fit the empirical frequency of outcomes. Importantly, this result applies pathwise, and does not require that either actions or empirical frequencies converge. It is based on arguments made in Diaconis and Freedman (1990) for the case of agents with full support beliefs.

### 4.2.1 Causation Neglect

When the agent has causation neglect, they believe that the distribution over outcomes is the same for all actions. Proposition\footnote{The Bayesian foundation of fictitious play assumes that the players believe that the environment is stationary and have Dirichlet priors. Away from a steady state the players are thus misspecified, but when the system converges to a steady state the stationarity assumption is asymptotically correct. In our setting, “substantial” misspecification can persist even when behavior converges.} gives a necessary condition for the convergence of beliefs and actions. Example\footnote{The Bayesian foundation of fictitious play assumes that the players believe that the environment is stationary and have Dirichlet priors. Away from a steady state the players are thus misspecified, but when the system converges to a steady state the stationarity assumption is asymptotically correct. In our setting, “substantial” misspecification can persist even when behavior converges.} shows that this condition is not sufficient to ensure convergence, even when there is a unique Berk-Nash equilibrium, and this equilibrium is uniformly strict.

**Example 3.** In this example the agent has three actions and their prior has support \{p^1, p^2, p^3\}. The details are in the following table.

<table>
<thead>
<tr>
<th>a</th>
<th>a = 1</th>
<th>a = 2</th>
<th>a = 3</th>
<th>H(p^*_a, \cdot)</th>
<th>A^m(\delta(\cdot))</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>1 2 3</td>
<td>1 2 3</td>
<td>1 2 3</td>
<td>a = 1 a = 2 a = 3</td>
<td></td>
</tr>
<tr>
<td>u</td>
<td>1 0 0</td>
<td>0 1 0</td>
<td>0 0 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p^*</td>
<td>0.1 0.9 0</td>
<td>0.9 0.1 0</td>
<td>0.1 0.1 0.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p^1</td>
<td>0.5 0.3 0.2</td>
<td>0.5 0.3 0.2</td>
<td>0.5 0.3 0.2</td>
<td>1.15 0.74 2.03</td>
<td>a = 1</td>
</tr>
<tr>
<td>p^2</td>
<td>0.3 0.5 0.2</td>
<td>0.3 0.5 0.2</td>
<td>0.3 0.5 0.2</td>
<td>0.74 1.15 2.03</td>
<td>a = 2</td>
</tr>
<tr>
<td>p^3</td>
<td>0.1 0.1 0.8</td>
<td>0.1 0.1 0.8</td>
<td>0.1 0.1 0.8</td>
<td>2.3 2.3 0.64</td>
<td>a = 3</td>
</tr>
</tbody>
</table>

In this example \(a = 3\) is the only Berk-Nash equilibrium and is uniformly strict. However, if the agent takes an action \(a \in \{1, 2\}\) then the subjective likelihood assigned to \(p^3\) goes down and thus play never converges to \(a = 3\) if the prior assigns sufficiently low probability to \(p^3\).
Proposition 4. Suppose that the agent has causation neglect. If \( a \) is a uniformly strict Berk-Nash equilibrium such that \( p^*_a \) is absolutely continuous with respect to \( p^*_{a'} \) for all \( a' \in A \), then it is positively attracting.

The proposition’s assumption of causation neglect guarantees that the uncontingent empirical distribution is a sufficient statistic for the agent’s beliefs. To prove the result, we first use Lemma 6 to show that on every path of outcome realizations, these beliefs concentrate around the distributions that minimize the KL divergence from the empirical frequency. We use this concentration to show there is a finite sequence of outcomes that has positive probability and leads the agent to play \( a \). Since \( a \) is a uniformly strict Berk-Nash equilibrium, if beliefs concentrate around the minimizers, \( a \) becomes the unique best reply. While using \( a \), the relative probability assigned to distributions in \( \hat{\Theta} \) increases in expectation, so we can combine Dubins’ upcrossing inequality with the fact that \( a \) is the unique myopic best reply to beliefs concentrated in \( \Theta(a) \) to show that with positive probability the agent will stick to action \( a \) forever.

Corollary 3. Suppose that the agent has causation neglect, and that for all actions \( a \in A \), \( \text{supp} p^*_a = Y \). Then every uniformly strict Berk-Nash equilibrium is positively attracting.

Esponda, Pouzo, and Yamamoto (2019) shows that for every uniformly strict Berk-Nash equilibrium \( a \), there exists at least one prior with support equal to \( \Theta \) under which the policy converges to \( a \) with positive probability. Frick, Iijima, and Ishii (2019b) provide sufficient conditions for there to be probability 1 that the system converges to a specific Berk-Nash equilibrium from any initial belief. In Proposition 4 we show that under causation neglect, every uniformly strict Berk-Nash equilibrium has positive probability of being the limit behavior starting from every initial prior. This relaxes the full support condition of Esponda, Pouzo, and Yamamoto (2019), and unlike Frick, Iijima, and Ishii (2019b) does not impose conditions that imply global convergence to a specific outcome.

4.2.2 Subjective Bandit Problems

Recall that in a subjective bandit problem, the agent’s prior \( \mu_0 \) is a product measure \( \mu_0 = \times_{a \in A} \mu_a \). Each \( \mu_a \in \Delta (\Delta (Y)) \) is the belief about the distribution over outcomes induced by action \( a \).

Proposition 5. In a subjective bandit problem there exists a \( \hat{\beta} < 1 \) such that if \( \beta \geq \hat{\beta} \) then under any optimal policy \( \pi \), every uniformly strict Berk-Nash equilibrium is positively attracting.
The proof uses the fact that patient agents experiment with actions that they believe might give them a higher payoff. The conclusion of the proposition is false for myopic agents even in the correctly specified case, where the Berk-Nash equilibria correspond to the self-confirming equilibria, and with probability 1 the agent may always play whichever action is myopically optimal given their initial beliefs.

The requirement of a uniformly strict Berk-Nash equilibrium is very demanding in a perceived bandit setting, as the Kullback-Leibler divergence between the true and subjective outcome distributions induced by action \( b \) does not constrain the “off-path” beliefs about the consequences of other actions, and very optimistic off-path beliefs can make some other action a better reply.

However, if the equilibrium is \textit{weakly identified}, we can relax uniformly strict to strict.

\textbf{Definition 8.} A Berk-Nash equilibrium \( a \) is \textit{weakly identified} (Esponda and Pouzo, 2016) if for all \( p, p' \in \hat{\theta}(a) \) we have \( p_a = p'_a \).

Weak identification guarantees that once behavior stabilizes on action \( a \), there is no additional updating about the relative likelihood of the KL minimizing outcome distributions. Under causation neglect, weak identification is a relatively strong condition, as it requires that the KL minimizer is unique. Weak identification is significantly weaker in subjective bandits, as it only requires the existence of a unique conceivable outcome distribution \( q_a \) that best matches \( p^*_a \), without imposing any restrictions on what the agent believes about the consequences of other actions.

\textbf{Proposition 6.} For every (perceived) bandit problem there is a \( \bar{\beta} < 1 \) such that if the discount factor \( \beta \geq \bar{\beta} \), then under any optimal policy \( \pi \), every weakly identified strict Berk-Nash equilibrium is positively attracting.

\subsection{Strongly Supermodular environments}

\textbf{Definition 9.} We say that the decision problem is \textbf{strongly supermodular} if we can strictly order the space of actions \( (A, >) \), outcomes \( (Y, >) \), and the set of conceivable distributions \( (\Theta, >) \) so that:

1. \( u \) is strictly supermodular in \( a \) and \( y \);

2. if \( p, p' \in \Theta \) and \( p > p' \), then for all \( a \in A \), and \( y \in Y \) we have \( p_a(\{y' : y' > y\}) > p'_a(\{y' : y' > y\}) \).

19
Proposition 7. In a strongly supermodular decision problem, if $p^*_a$ (resp. $p^*_\bar{a}$) has full support, and the highest action $\bar{a}$ (resp. the the lowest action $a$) is a uniformly strict Berk-Nash equilibrium, then $\bar{a}$ (resp. $a$) is positively attracting.

Strong supermodularity implies that for each action there is a set of outcome realizations, the highest $y$’s, that if observed supports the use of action $\bar{a}$. A finite number of such realizations will induce the agent to use action $\bar{a}$, and from that moment our Proposition 2 guarantees that there is positive probability that the agent will use $a$ forever.

5 Signals

Suppose each period the agent observes a signal $s$ from a finite set $S$ before taking an action. Thus the analog of an action in the previous sections is now a strategy, namely a map $\sigma : S \rightarrow A$ from signals to actions. Signals may be payoff relevant, so now utility is a map $u : A \times Y \times S \rightarrow \mathbb{R}$, and signals may also be useful for predicting the outcome distributions, i.e., now $p_{a,s} \in \Delta(Y)$ depends both on this period action and signal.

Here we expand the probability space of our basic model in the obvious way: The sample space $\Omega = S^\infty \times (Y^\infty)^A$ consists of infinite sequences of signal and action dependent outcome realizations $(s_k, x_{a,s',k})_{k \in \mathbb{N}, a \in A, s' \in S}$ and $x_{a,s',k}$ determines the outcome when the agent takes the action $a$ for the $k$-th time after $s$.\(^{19}\)

To complete the model we also need to specify the distribution of signals. We focus here on the case where the distribution of $s$ is fixed (iid) with distribution $\zeta$, which is known to the agent, as in Esponda and Pouzo (2016).\(^{20}\)

Subjective Beliefs of the Agent The agent correctly believes that the map from actions and signals to probability distributions over outcomes is fixed, but they are uncertain about the distribution each signal and action pair induces. Let $P \times |S|$ be the space of all signal and action dependent outcome distributions. The agent’s uncertainty is captured by a prior belief $\mu_0 \in \Delta(P)$, with again $\Theta = \text{supp}_{\mu_0}$. The only restriction is the generalization of Assumption 1.

Assumption 3. For all $p \in \Theta$ and $\varepsilon > 0$, there exists $p' \in \Theta$ with $||p' - p|| < \varepsilon$ such that for all $a \in A, s \in S$, if $p^*_a(y) > 0$ then $p'_{a,s}(y) > 0$.

\(^{19}\)We specify the details of the sample space and its probability measure in Appendix 8.3.

\(^{20}\)Note that we need to allow for a continuum of signals for payoff perturbations to generate continuous best response distributions, as assumed by Esponda and Pouzo (2016) and Frick, Iijima, and Ishii (2019b).
Denote by $\mu_t(\cdot \mid (s^t, a^t, y^t)) \in \Delta(P)$ the agent’s subjective belief obtained using Bayes rule after observing the sequence of signals and outcomes $(s^t, y^t)$ when taking the actions $a^t$, so that
\[
\mu_t(C \mid (s^t, a^t, y^t)) = \frac{\int_{p \in C} \prod_{t=1}^{\tau} p_{a_t^t, s_t^t}(y_t) d\mu_0(p)}{\int_{p \in P} \prod_{t=1}^{\tau} p_{a_t^t, s_t^t}(y_t) d\mu_0(p)}.
\] (3)

The relevant set of “closest beliefs to the truth” is now
\[
\hat{\Theta}(\sigma) = \arg\min_{p \in \Theta} \sum_{s \in S} \zeta(s) H \left( p_{\sigma(s), s}^*, p_{\sigma(s), s} \right).
\] (4)

We use this modified definition of the minimizers to extend the definition of Berk-Nash equilibrium and uniformly strict Berk-Nash equilibrium to this more general setting.

**Definition 10** (Berk-Nash Equilibrium).

(i) Strategy $\sigma$ is a **Berk-Nash equilibrium** if there exists a belief $\nu \in \Delta(\hat{\Theta}(\sigma))$ such that $\sigma$ is myopically optimal given $\nu$.

(ii) Strategy $\sigma$ is a **uniformly strict Berk-Nash equilibrium** if $\sigma$ is the unique myopic best reply to any belief in $\nu \in \Delta(\hat{\Theta}(\sigma))$.

**Proposition 8.** Suppose the agent uses an optimal myopic policy. Then if the strategy prescribed by the policy converges to $\sigma$ with positive probability, then $\sigma$ is a Berk-Nash equilibrium.

The proof associates each action contingent outcome distribution $p$ with the joint distribution $p_{\sigma} \in \Delta(S \times Y)$ over signals and outcomes that is induced by $p$ and the use of strategy $\sigma$. We then use Berk’s Theorem for the belief process in this augmented environment to argue that the policy converges to $\sigma$, then beliefs concentrate around $\hat{\Theta}(\sigma)$. We finally use the 0-1 Law to show that if $\sigma$ is not optimal with respect to some beliefs concentrated in $\hat{\Theta}(\sigma)$, almost surely the optimal myopic policy will prescribe an action different from the one dictated by $\sigma$, a contradiction.

We are interested cases where the distribution $p_{a, s}$ on outcomes depends on both $a$ and $s$, but the has “signal neglect,” and mistakenly thinks that the distribution of $y$ given $a$ is independent of $s$. Here the only reason why the signals matter for the agent is that they may enter into the payoff function directly.\(^{21}\) This is very similar to the case where the agent estimates the same distribution for all actions, but here for each $a$ that is played

\(^{21}\)If the signals are indeed uncorrelated with the outcomes, then this is a model of learning with payoff perturbations as in Fudenberg and Kreps [1993], except that those models typically assume that the signals have a density that is absolutely continuous with respect to Lebesgue measure. If the agent has signal neglect
infinitely often, Berk-Nash equilibrium requires that beliefs minimize the divergence with the signal-independent belief

$$\hat{\Theta}(\sigma) = \arg\min_{\mu \in \text{supp}_p} H(\bar{p}_\sigma, p),$$

where $\bar{p}_\sigma = \sum_s \zeta(s)p_{\sigma(s),s}(y)$.

Example 7 illustrates the long-run biases that can be induced by signal neglect. There, we study the choice of a seller who receives a signal about the market attendance in the current period, and can decide whether to undertake an investment that may boost sales, with the outcome $y$ the fraction of market participants who buy. The seller does not realize that when more consumers show up, a lower fraction of them buy, and we show that this can lead to persistent underinvestment when the market attendance is high.

**Definition 11** (Positively Attracting). The strategy $\sigma \in A^S$ is positively attracting if for every optimal policy $\pi$

$$\mathbb{P}_\pi \left[ \lim_{t \to \infty} \pi(s^{t+1}, a^t, y^t) = \sigma(s^{t+1}) \right] > 0.$$ 

**Proposition 9.** Under causation and signal neglect, if $\text{supp} p_{a,s}^* = Y$ for all $(a, s) \in A \times S$, then any uniformly strict Berk-Nash equilibrium $\sigma$ is positively attracting.

The proof structure is very similar to Proposition 4 because with signal neglect we can apply our extension of the Diaconis and Freedman (1990) result to the uncontingent empirical distribution.

### 6 Extensions

**6.1 Learning in Large Population Games**

Our framework can be used as the building block for steady-state learning in large population games. We now highlight how to map our framework to that setting, and why the biases we consider are relevant there. Consider a finite game with a set of players roles $I$, and action space $A = \times_{i \in I} A_i$. There is a continuum of agents in each player role $i \in I$. Every period, the agents are matched uniformly at random to play the game. Agents discount the future at rate $\hat{\beta}$, and they have a probability $\gamma$ of surviving into next period.

and in addition believes that the signals are not directly payoff relevant, signal neglect reduces to replacing $p^*$ with the compound distribution over outcomes. With this change all the previous propositions continue to hold.
In a steady state, the problem faced by an agent in population $i$ is formally equivalent to the one we have considered in the previous sections: the agent correctly believes they are facing a stationary environment, and they realize that they do not affect the next period distribution of opponents’ strategies. Here, the finite set of actions $A$ is $A_i$, $Y = A_{-i}$, $u$ is the payoff function of player $i$, and the actual discount factor of the agent is $\beta = \hat{\beta} \gamma$, and for simplicity, we assume that at the end of the period the agent perfectly observes the strategy of the opponents.

The objective probability distribution $p^* \in P = \Delta(Y)^A$ is determined by the steady-state behavior of the opponents. Causation neglect corresponds to the bias of an agent who thinks they are playing a simultaneous-move game, when in reality the opponents observe the agent’s choice before moving. Subjective bandit problems arise when the agent has independent beliefs about the responses to different strategies. In games of incomplete information, the agent has signal neglect when they think the game has independent private values.

In all of these situations, our results help characterize the possible limit actions. Of course, extensive-form games may not have strict equilibria, so some of our propositions will not apply, but it may be possible to extend some conclusions to equilibria that are on-path strict in the sense of Fudenberg and He (2020). Also, games need not have pure-strategy equilibria, but it may be possible to apply our methods to setting where each agent plays deterministically and different agents in the same player role chose different actions.\footnote{Alternatively we could enrich the model with payoff perturbations as in Fudenberg and Kreps (1993) and Esponda and Pouzo (2016).}

### 6.2 Markov Decision Problems

If the agent’s action influences the signal, then the true model is that of a Markov decision problem. Even if the agent ignores this, the evolution of their beliefs and actions becomes more complicated. And if the agent is aware of it, and tries to solve a Markov decision problem as in Esponda and Pouzo (2019) then the problem is yet more complex. We hope to have more to say about this in future work.

### 6.3 Other Examples

**Example 4** (Central Bank Policy). A central bank decides between three actions:

- $f$: keep a flexible exchange rate with the dollar,
c: peg the currency to the dollar imposing some controls to the capital, or

$ g$: peg the currency without imposing controls.

The outcome has three binary components, $y = (y^e, y^s, y^{cf})$, where $y^e$ says whether the economy is in expansion in the current period, $y^s$ whether there is a speculative attack on the currency, and $y^{cf}$ if there is capital flight. The central bank believes that the three components are distributed according to independent binomial distributions that do not depend on the action, with

$$p(y) = \eta_{y^e} \eta_{y^s} \eta_{y^{cf}} (1 - \eta_{y^e})^{1-y^e} (1 - \eta_{y^s})^{1-y^s} (1 - \eta_{y^{cf}})^{1-y^{cf}}.$$  

The payoff function is given by

$$u(f, y) = y^e$$
$$u(c, y) = \frac{3}{2} y^e - y^s$$
$$u(g, y) = 2y^e - \frac{y^{cf}}{2} - \frac{y^s}{2}.$$  

This specification of the payoffs captures the fact that a fixed exchange rate facilitates trade and foreign investment when the currency isn’t attacked. In this setting, Proposition 4 guarantees that if $p^*$ has full support, every uniformly strict Berk-Nash equilibrium has positive probability of arising as the limit outcome.

Example 5 (Uniform Berk-Nash that isn’t Positively Attracting). Now maintain the same action set and utility function for the central bank, and suppose that it does not have causation neglect, and that its prior gives positive probability only to four action-contingent outcome distributions $\bar{p}_1, \bar{p}_2, \bar{p}_1, \bar{p}_2$ where

$$\bar{p}_2^1 (y^e = 1) = \bar{p}_1^2 (y^e = 1) = \frac{2}{3}; \quad \bar{p}_1^1 (y^e = 1) = \bar{p}_2^2 (y^e = 1) = 1$$
$$\bar{p}_2^2 (y^e = 1) = \bar{p}_1^1 (y^e = 1) = \frac{1}{3}; \quad \bar{p}_1^1 (y^e = 1) = \bar{p}_2^2 (y^e = 1) = \frac{2}{3}$$

$$\forall p \in \Theta, a \in A, \quad p_a (y^{cf} = 1) = 1.$$  

\footnote{See e.g., Frankel, Schmukler, and Servén (2004).}
and

\[ p^2_f (y^s = 1) = p^2_f (y^s = 1) = p^1_f (y^s = 1) = p^2_f (y^s = 1) = \frac{1}{10} \]

\[ p^1_c (y^s = 1) = p^2_c (y^s = 1) = \frac{9}{10} \]

\[ p^1_c (y^s = 1) = p^1_c (y^s = 1) = \frac{2}{10} \]

The true data generating process is such that

\[ p^*_f (y^e) = \frac{1}{2} \quad p^*_c (y^e) = \frac{5}{6} \quad p^*_f (y^e) = \frac{1}{10} \quad p^*_c (y^e) = \frac{2}{10} \quad \text{and} \quad \forall p \in \Theta, \quad p^*_g (y^e = 1) = 1. \]

In this case, pegging the currency to the dollar is a uniformly strict Berk-Nash equilibrium, but it is not positively attractive: For every discount factor \( \beta \), if the prior assigns sufficiently high probability to \( \bar{p}^2 \) and \( \bar{p}^2 \), the probability of converging to \( s \) under an optimal policy is 0. To see why, note that when the currency is pegged to the dollar, \( \bar{p}^2 \) is indistinguishable from \( \bar{p}^1 \), and \( \bar{p}^2 \) is indistinguishable from \( \bar{p}^1 \). Therefore, the bank doesn’t update their beliefs, and remains convinced that a pegged currency is highly susceptible to a speculative attack, so it maintains a flexible exchange rate.

**Example 6** (Role of Assumption 1). Suppose there are two actions \( a \) and \( b \), and two outcomes \( Y = \{0, 1\} \), and let \( u(a, 0) = u(b, 1) = 1 - u(a, 1) = 1 - u(b, 0) \). We identify the elements of \( \Delta(Y) \) with the probability they assign to outcome 1. Let \( p^*_a = \frac{2}{3} \) and \( p^*_b = 1 \). Suppose that the agent believes that the outcome distribution does not depend on the action, so they have correlation neglect, and let \( \Theta = \{ \frac{1}{3}, 1 \} \). Here \( b \) is the unique Berk-Nash equilibrium, and it is uniformly strict. However, if the prior assigns a sufficiently high probability to \( 1/3 \), the agent will start playing \( a \), and with positive probability they will observe outcome 0 in the first period. But after this observation, the posterior assigns probability one to \( p = 1/3 \) and the actions will converge to \( a \).

**Example 7** (Signal Neglect). A seller in a physical marketplace can hire one shop assistant to work for the day \( a_H \) or not hire anyone \( a_N \). The outcome \( y \in Y \) is the percentage of consumers in the marketplace that buy the good, with two possibilities, \( y_h > y_l \).

Before choosing whether to hire, the agent observes the the number of people at the market that day \( s \in \{s_h, s_l\} \), with \( s_h > s_l \). The payoff function is \( u(a, y, s) = sy - 1_{c=a_H} \). The seller realizes that the signal is payoff relevant, but falsely believes that it does not provide any
information about the outcome. The agent is uncertain about how useful it is to hire a shop assistant, and in particular they do not know whether hiring is ineffective, i.e., for all \( a \in A, y \in Y \), \( p_a(y) = 1/2 \), or if it is not, i.e., \( p_{ah}(y_H) = 3/4 \) and \( p_{an}(y_H) = 1/4 \).

The fraction of consumers who buy varies with the signal: On days with fewer consumers, the ones that actually come to the market are more likely to purchase the good. Formally:

\[
p_{s_h,a_H}(y_H) = 1/2, \quad p_{s_h,a_N}(y_H) = 1/4, \quad p_{s_l,a_H}(y_H) = 3/4, \quad p_{s_l,a_N}(y_H) = 1/2.
\]

Let \( s(y_H - y) \) < 1 < \( s(y_H - y) \), so that it is not optimal to hire a shop assistant after \( s_L \), and it is optimal to hire an assistant after \( s_H \). Then the only Berk-Nash equilibrium is that the shop assistant is never hired.\(^{24}\)

7 Conclusion

In many economically relevant settings it seems plausible that agents misunderstand some aspects of the world. For this reason it is important to understand what beliefs such agents will develop and how they will behave. This paper provides sharp characterizations of what actions arise as the long-run outcomes of misspecified learning. We show that all uniformly strict Berk Nash equilibria are stable, and that under a mild condition only uniform Berk Nash equilibria can be stable. Moreover we show that when the support of the agent’s prior is finite, play can only converge to uniform Berk-Nash equilibria.\(^{25}\) Our work thus suggests uniformity should be imposed as a refinement of Berk-Nash equilibrium. We then provide the first sufficient conditions for an action to be positively attracting under misspecified learning. Here we highlight the role played by the correlation that the agent perceives between the outcome distributions associated with different actions.

When the agent has a finite number of possible actions or stage-game strategies, as we have assumed in this paper, an equivalent definition of uniformly strict Berk-Nash equilibrium is an action \( a \) that is the unique best response to every belief in a neighborhood of the

\(^{24}\)If the agent followed the optimal strategy, they would observe the same frequency of sales in days with \( s = s_H \) and with the shop assistant hired as in days with \( s = s_L \) and without the shop assistant: \( p_{s_h,a_H}(y_H) = 1/2 = p_{s_l,a_N}(y_H) \). This holds because the shop assistant offsets the lower per-customer demand on days with high attendance. However, this observation supports the belief that the shop assistant is useless. Since the myopic best reply to \( \delta_p \) is to never hire the shop assistant, by Proposition 8 this suboptimal action is the only possible limit action.

\(^{25}\)Note that the uniformity issue that we address cannot arise in a correctly specified model, where the agent always learn the outcome distribution induced by the equilibrium action. Note also that our results do not imply that actions converge.
KL-minimizers for $a$. With infinitely many actions and continuous payoff functions, actions that are sufficiently near the strict best response incur arbitrarily small losses, and are best responses to nearby beliefs. Here the two definitions of uniformly strict Berk-Nash equilibrium are not equivalent. Indeed, as shown by an example in Heidhues, Koszegi, And Strack, 2018, some Berk-Nash equilibria that are uniformly strict Berk-Nash in the sense of Definition 5(iv) may not be positively attractive. However, we conjecture that the positive attractiveness result continues to hold under the alternative definition.

In future work we hope to extend our analysis to Markov decision problems, as in Esponda and Pouzo (2019) and to misspecified learning in multiplayer games, as in Eyster and Rabin (2005), Jehiel (2005), and Jehiel and Koessler (2008).

8 Appendix

The structure of the appendix is as follows. Section 8.1 establishes some preliminary lemmas. Section 8.2 contains the results of the main text for the models that do not have signals, except for the first part of Proposition 1 which will follow immediately from the more general Proposition 8. Finally, Section 8.3 contains some technical details about the probability space and the proofs of the results for the model with signals.

8.1 Preliminary lemmas and definitions

Lemma 1. For every $a \in A$ and $\varepsilon > 0$, $\hat{\Theta}(a)$, $\hat{\Theta}_a(a)$, $\varepsilon(a)$, and $\Delta(\hat{\Theta}(a))$ are compact.

Proof. Compactness of $\hat{\Theta}(a)$ follows from the generalization of Weierstrass Theorem to lower-semicontinuous functions (see, e.g., Theorem 2.43 in Aliprantis and Border, 2013). Since the projection map is continuous, and $\hat{\Theta}_a(a)$ is the projection of $\hat{\Theta}(a)$, $\hat{\Theta}_a(a)$ is compact as well. Since $\hat{\Theta}_a(a)$ is closed, it immediately follows that $\hat{\Theta}(a)$ is closed as well, henceforth compact. Given the compactness and separability of $\hat{\Theta}(a)$, $\Delta(\hat{\Theta}(a))$ is compact by, e.g., Theorem 6.4 in Parthasarathy (2005).

For every $p \in P$ and every policy $\pi$ denote as $E_{p,\pi}$ the probability distribution over action and outcome sequences that is induced by policy $\pi$ under outcome distribution $p$. For convenience, for every policy $\pi$ and initial belief $\nu$ we will work with the agent’s normalized value throughout, which is

$$V(\pi, \nu) = \frac{\int_p E_{p,\pi} \left[ \sum_{t=1}^{\infty} \beta^{t-1} u(a_t, y_t) \right] d\nu(p)}{1 - \beta}.$$
We first bound the difference between the value of using action $a$ and the value of any other action in terms of their expected utility given that beliefs are concentrated around the outcome distributions $\hat{\Theta}(a)$ that minimize the Kullback-Leibler divergence from the correct distribution $p^*_a$ induced by $a$.

Define the set of beliefs over conceivable distributions that assign at least probability $1 - \varepsilon$ to $\hat{\Theta}^\varepsilon(a)$ as

$$M_{\varepsilon,a} = \{\nu \in \Delta(\Theta) : \nu(\hat{\Theta}^\varepsilon(a)) \geq 1 - \varepsilon\}.$$ 

**Lemma 2.** If $a \in A$ is a uniformly strict Berk-Nash equilibrium, for every optimal policy $\pi$, there exists an $\hat{\varepsilon} > 0$ such that for all $\varepsilon < \hat{\varepsilon}$

$$\nu \in M_{\varepsilon,a} \implies \pi(\nu) = a.$$ 

**Proof.** Let $\pi^a$ denote the policy that prescribes to always play $a$. Define $G(\varepsilon)$ as the minimal gain from playing $a$ forever instead of using (one of) the best policy $\tilde{\pi}$ that does not play $a$ at a prior belief $\nu$ in $M_{\varepsilon,a}$

$$G(\varepsilon) = \min_{\tilde{\pi}} \min_{\pi \neq a, \nu \in M_{\varepsilon,a}} (V(\pi^a, \nu) - V(\tilde{\pi}, \nu)).$$

Notice that by Tychonoff’s Theorem, the fact that the set of histories is countable, and the finiteness of the actions, the space of the policy functions endowed with the product topology is compact. Since the subset of policy functions that do not prescribe $a$ at the initial history is closed, this subset is compact as well. Moreover, given that $\beta \in (0,1)$, the value function is continuous at infinity, and therefore $V(\pi^a, \nu) - V(\cdot, \nu)$ is a continuous function of the policy. Notice also that since $E_{p,\pi} [\sum_{t=1}^{\infty} \beta^{t-1} u(\pi^a, \nu, \pi^a, \nu)$ is continuous in $p$, $V(\pi^a, \cdot) - V(\tilde{\pi}, \cdot)$ is continuous in $\nu$. Therefore, given that $\varepsilon \to M_{\varepsilon,a}$ is a continuous and compact valued correspondence, we can conclude by the Maximum Theorem that $G$ is continuous in $\varepsilon$. Since $a$ is a uniformly strict Berk-Nash equilibrium, $G(0) > 0$, and there is an $\hat{\varepsilon}$ such that if $\varepsilon \leq \hat{\varepsilon}$, $G(\varepsilon) > 0$. This implies that for any optimal policy $\pi$ it must be such that $\nu \in M_{\varepsilon,a}$ implies that $\pi(\nu) = a$, which proves the lemma.

This next Lemma extends an argument of Fudenberg and Levine (1992) to take into account misspecification.
Lemma 3. Let \( p, p', p^* \in \Delta (Y) \), and \( l \in (0, 1) \) be such that

\[
\sum_{y \in Y} p^*(y) \left( \frac{p(y)}{p'(y)} \right)^l > 1.
\]  \hspace{1cm} (5)

Then, there is \( \varepsilon' > 0 \) such that for all \( \nu \in \Delta (\Delta (Y)) \), if we let

\[
\nu(C \mid y) = \frac{\int_{y \in C} q(y)d\nu(q)}{\int_{y \in \Delta (Y)} q(y)d\nu(q)},
\]

then

\[
\sum_{y \in Y} p^*(y) \left[ \left( \frac{\nu(B_{\varepsilon'}(p) \mid y)}{\nu(B_{\varepsilon'}(p') \mid y)} \right)^l \right] \geq \left( \frac{\nu(B_{\varepsilon'}(p))}{\nu(B_{\varepsilon'}(p'))} \right)^l.
\]

Proof. Without loss of generality, we can assume that \( \nu(B_{\varepsilon'}(p')) > 0 \). Define \( G : \mathbb{R}_+ \rightarrow \mathbb{R} \) as

\[
G(\varepsilon) = \min_{\nu' \in \Delta (B_{\varepsilon'}(p')), \nu \in \Delta (B_{\varepsilon'}(p))} \sum_{y \in Y} p^*(y) \left( \frac{\int_{B_{\varepsilon'}(p)} \bar{q}(y)d\bar{\nu}(\bar{q})}{\int_{B_{\varepsilon'}(p')} \bar{q}(y)d\bar{\nu}'(\bar{q})} \right)^l.
\]

By the Maximum Theorem, the compactness of \( \Delta (B_{\varepsilon'}(p')) \) and \( \Delta (B_{\varepsilon'}(p)) \) (see, e.g, Theorem 6.4 in Parthasarathy, 2005) and the fact that \( G(0) > 1 \) by equation (5), there is \( \varepsilon' > 0 \) such that for all \( \nu' \in \Delta (B_{\varepsilon'}(p')) \), \( \bar{\nu} \in \Delta (B_{\varepsilon'}(p')) \)

\[
\sum_{y \in Y} p^*(y) \left( \frac{\int_{B_{\varepsilon'}(p)} \bar{q}(y)d\bar{\nu}'(\bar{q})}{\int_{B_{\varepsilon'}(p')} \bar{q}(y)d\bar{\nu}'(\bar{q})} \right)^l \geq 1.
\]  \hspace{1cm} (6)

Then,

\[
\sum_{y \in Y} p^*(y) \left( \frac{\nu(B_{\varepsilon'}(p) \mid y)}{\nu(B_{\varepsilon'}(p') \mid y)} \right)^l = \sum_{y \in Y} p^*(y) \left( \frac{\int_{B_{\varepsilon'}(p)} \nu(B_{\varepsilon'}(p))\bar{q}(y)d\nu(q)\frac{\nu(q)}{\nu(B_{\varepsilon'}(p))}}{\int_{B_{\varepsilon'}(p')} \nu(B_{\varepsilon'}(p'))\bar{q}(y)d\nu'(q)\frac{\nu(q)}{\nu(B_{\varepsilon'}(p'))}} \right)^l
\]

\[
= \sum_{y \in Y} p^*(y) \left( \frac{\int_{B_{\varepsilon'}(p)} \bar{q}(y)d\frac{\nu(q)}{\nu(B_{\varepsilon'}(p))}}{\int_{B_{\varepsilon'}(p')} \bar{q}(y)d\nu'(q)\frac{\nu(q)}{\nu(B_{\varepsilon'}(p'))}} \right)^l \left( \frac{\nu(B_{\varepsilon'}(p))}{\nu(B_{\varepsilon'}(p'))} \right)^l
\]

\[
\geq \left( \frac{\nu(B_{\varepsilon'}(p))}{\nu(B_{\varepsilon'}(p'))} \right)^l.
\]
where the inequality follows from equation \[6\].

The next lemma extends Lemma 3 of Frick, Iijima, and Ishii (2019b) to show that there exists a uniform \(l\) such that all of the KL-minimizer distributions dominate all the distributions that are \(\varepsilon\) away from the minimizers accordingly to the divergence of order \(l\).

**Lemma 4.** Fix an action \(a\) and \(\varepsilon > 0\). There exists an \(\overline{l} > 0\) such that for all \(l \leq \overline{l}\) for every KL minimizer \(q \in \hat{\Theta}(a)\) and every outcome distribution \(p'\) that is at least \(\varepsilon\) away from any KL minimizer, that is, such that \(p' \notin \hat{\Theta}^\varepsilon(a)\)

\[
f_i(q, p') = \sum_{y \in Y} p^*_a(y) \left( \frac{q_a(y)}{p'_a(y)} \right)^i > 1.
\]

**Proof.** As noted by Frick, Iijima, and Ishii (2019b) in their Lemma 3, a simple continuity argument shows that (i) for each KL minimizer \(q \in \hat{\Theta}(a)\) and every outcome distribution \(p' \notin \hat{\Theta}(a)\) there exists an \(l(q, p')\) such that \(f_i(q, p') > 1\) for all \(l \leq l(q, p')\) and (ii) for all \(q, q' \in \Theta\), if \(\hat{l} > l\) and \(f_i(q, q') \leq 1\), then \(f_i(q, q') = 1\). We will now prove that there exists a uniform \(l\) that works for every \(q \in \hat{\Theta}(a)\) and \(p' \in \hat{\Theta}^\varepsilon(a)\).

Suppose by way of contradiction that there was no \(\overline{l} > 0\) such that for all \(l \leq \overline{l}\) \(f_i(q, p') > 1\) for all \(q \in \hat{\Theta}(a)\) and \(p'\) with \(p'_a \notin \hat{\Theta}^\varepsilon(a)\). Then, we can define a sequence \((q_n, p'_n)\) such that \(f_{\overline{l}}(q_n, p'_n) \leq 1\), and sequential compactness of \(\hat{\Theta}(a) \times \{p \in \Delta(\Theta) : p_a \notin \hat{\Theta}^\varepsilon(a)\}\) guarantees that this sequence has an accumulation point \((q, p')\). However, for, \(n > \frac{1}{l(q, p')}\), \(f_{\overline{l}}(q_n, p'_n) \leq 1\) implies \(f_{l(q, p')}(q_n, p'_n) \leq 1\), but then continuity of \(f_{l(q, p')}\) leads to a contradiction with \(f_{l(q, p')}(q, p') > 1\).

Given two outcome distributions \(q, q' \in \Delta(Y)\) and \(\varepsilon > 0\), let

\[
U_\varepsilon(q, q', \alpha) = \{q'' \in \Delta(Y) : ||\alpha q + (1 - \alpha)q' - q''|| = \varepsilon\}
\]

denote the sphere of radius \(\varepsilon\) around \(\alpha q + (1 - \alpha)q' - q''\).

**Lemma 5.** Fix \(q \in \Delta(Y)\) with \(\text{supp} q \subseteq \text{supp} p^*_a\) and a compact set \(C \subseteq \Delta(Y)\) such that there exists \(q' \in C\) with \(\text{supp} p^*_a \subseteq \text{supp} q'\). Then there exists a \(K > 0\) such that for every \(f' \in U_\varepsilon(q, p^*_a, \alpha)\) with \(\text{supp} f' \subseteq \text{supp} p^*_a\) then

\[
\min_{q' \in C} H ((1 - \alpha)p^*_a + \alpha q, q') - H ((1 - \alpha)p^*_a + \alpha q, q') - \min_{q' \in C} H (f', q') + H (f', q) \leq K\varepsilon.
\]
Proof. First, notice that by the Maximum Theorem, the set

\[ C := \bigcup_{\varepsilon \in [0, 1]} \bigcup_{\alpha \in [0, 1]} \bigcup_{f' \in U_q} \min_{q' \in C} H(f', q') \]

is compact, so there is a \( K_1 \) such that \(-\min_{y \in \text{supp} \mu_a} \min_{q' \in U} H'(y) < K_1\).

Then we have that

\[
\begin{align*}
\min_{q' \in C} H((1 - \alpha)p_a + \alpha q', q') - H((1 - \alpha)p_a + \alpha q, q) - \min_{q' \in C} H(f', q') + H(f', q) \\
\leq \min_{q' \in C} H((1 - \alpha)p_a + \alpha q, q') - \min_{q' \in C} H(f', q') - \varepsilon \min_{y \in \text{supp} \mu_a} \log q(y) \\
\leq K_1 \varepsilon - \varepsilon \min_{y \in \text{supp} \mu_a} \log q(y),
\end{align*}
\]

and if we define \( K = K_1 - \min_{y \in \text{supp} \mu_a} \log q(y) > 0 \), it satisfies the requirement of the statement.

The following lemma is about the concentration of beliefs. The lemma considers the beliefs about the outcome distribution corresponding to a given action, i.e., to elements of \( \Delta_{\Delta(Y)} \) as opposed to elements of \( \times_{a \in A} \Delta(Y) \subset \mathbb{R}^{Y \times |A|} \), so we will lighten notation by working in this smaller space.

Let \( \chi \in \Delta(\Delta(Y)) \) be a belief over probability distributions on \( Y \), and let

\[ Q_{\varepsilon, \chi}(\tilde{q}) = \left\{ q' \in \Delta(Y) : \exists q'' \in \arg\min_{q \in \text{supp} \chi} H(\tilde{q}, q), ||q' - q''||_{\infty} < \varepsilon \right\} \]

be the distributions in \( \text{supp} \chi \) that are within \( \varepsilon \) of a distribution \( q'' \) that minimizes the Kullback-Leibler divergence with the given \( \tilde{q} \).\(^{26}\) We will show that beliefs about the outcome distribution induced by any action \( a \) concentrate at an exponential rate around \( Q_{\varepsilon, \text{supp} \mu_a} \), the distributions that best fit the empirical frequency of outcomes generated by \( a \). Importantly, this result does not require that either actions or empirical frequencies converge. It is based on arguments made in Diaconis and Freedman (1990), who considered agents with full support beliefs. It will be important in what follows that these results apply pathwise.

**Lemma 6.** Let \( \chi_0 \in \Delta(\Delta(Y)) \) and suppose that for every \( t \in \mathbb{N} \), \( C \subseteq \Delta(Y) \), and sequence of outcomes \( y^t \in Y^t \)

\[
\chi_t(C \mid y^t) = \frac{\int_{q \in C} \prod_{\tau=1}^{t} q(y_{\tau}) d\chi_0(q)}{\int_{q \in \Delta(Y)} \prod_{\tau=1}^{t} q(y_{\tau}) d\chi_0(q)},
\]

\(^{26}\)Note that the argmin in this definition need not be continuous because \( \text{supp} \chi \) is not required to be convex.
Then there are a function \( g : \Delta (Y) \times [0, 1] \to \mathbb{R}_+ \) that is upper-semicontinuous in the second argument and a \( K > 0 \) such that for every \( \varepsilon > 0 \)

\[
p'(y) = \sum_{t=1}^{\tau} \frac{1_{y_t = y}}{t} \implies \frac{\chi_t (Q_{\varepsilon, \chi_0} (p') \mid y^t)}{1 - \chi_t (Q_{\varepsilon, \chi_0} (p') \mid y^t)} \geq \left[ \chi_0 \left( Q_{\frac{\varepsilon K}{g(p', \varepsilon)}, \chi_0} (p') \right) \right]^2 e^{tg(p', \varepsilon)}.
\]

**Proof.** Let \( p'(y) = \sum_{t=1}^{\tau} \frac{1_{y_t = y}}{t} \) and fix \( \varepsilon > 0 \). To ease notation, in this proof for every \( \tilde{\varepsilon} > 0 \), we let \( Q(\tilde{\varepsilon}) = Q_{\varepsilon, \chi_0} (p') \). Define

\[
g (p', \varepsilon) = \frac{\min_{p \in Q(\varepsilon)} H (p', p) - \min_{p \in \text{supp} \chi_0} H (p', p)}{2} > 0.
\]

Since \( g \) is upper-semicontinuous in the second argument (see, e.g., Lemma 17.30 in Aliprantis and Border, 2013), for every \( \hat{p} \in \arg\min_{p \in \text{supp} \chi_0} H (p', p) \) there exists an \( \varepsilon_{\hat{p}} \leq \varepsilon \) such that \( p \in B_{\varepsilon_{\hat{p}}} (\hat{p}) \) and \( q \notin Q(\tilde{\varepsilon}) \) implies

\[
\min_{p \notin Q(\varepsilon)} H (p', p) - H (p', p) \geq g (p', \varepsilon).
\]

By the Maximum Theorem, \( \arg\min_{p \in \text{supp} \chi_0} H (p', p) \) is compact, and it is covered by the collection

\[
\bigcup_{\hat{p} \in \arg\min_{p \in \text{supp} \chi_0} H (p', p)} B_{\varepsilon_{\hat{p}}}^{\text{int}} (\hat{p}),
\]

and therefore it admits a finite subcollection. Let \( \varepsilon' \) be the minimum of the \( \varepsilon_{\hat{p}} \) indexing the subcollection. By the same argument of Lemma 5, there is \( K > 0 \) such that \( \varepsilon' > \varepsilon K \). From the definition of \( \chi_t \), we have that for all \( y^t \) such that the corresponding empirical distribution is \( p' \),

\[
\frac{\chi_t (Q(\varepsilon) \mid y^t)}{1 - \chi_t (Q(\varepsilon) \mid y^t)} = \frac{\int_{Q(\varepsilon)} \sum_{y \in Y} q(y)^{tp'(y)} (1 - q(y))^{t(1-p'(y))} d\chi_0(q)}{\int_{\text{supp} \chi_0 \setminus Q(\varepsilon)} \sum_{y \in Y} q(y)^{tp'(y)} (1 - q(y))^{t(1-p'(y))} d\chi_0(q)} \geq \frac{\int_{Q(\varepsilon')} \exp(-t H (p', q)) d\chi_0(q)}{\exp(-t \min_{p \notin Q(\varepsilon)} H (p', p))} = \chi_0 (Q(\varepsilon')) \int_{Q(\varepsilon')} \exp(t \min_{p \notin Q(\varepsilon)} H (p', p) - t H (p', q)) d\chi_0(q) \geq \chi_0 (Q(\varepsilon'))^2 e^{tg(p', \varepsilon)}.
\]
8.2 Proof of results stated in the text

Proof of Proposition 1. Proposition 1 (i) follows immediately from Proposition 8. For Proposition 1 (ii), the key idea is to consider the empirical frequency of the different outcomes.

If \( a \) is not a uniform Berk-Nash equilibrium, there is \( q \in \hat{\Omega}_a(a) \) such that if \( \text{supp} \nu \subseteq \{ p \in \hat{\Omega}(a) : \forall y \in \text{supp} p_a^*, p_a(y) = q(y) \} \), then \( a \notin A^m(\nu) \). We fix such a \( q \) throughout this proof.

Claim 1. There exists \( \varepsilon > 0 \) such that if

\[
\frac{\nu \left( \{ p \in \hat{\Omega}(a) : \forall y \in \text{supp} p_a^*, |p_a(y) - q(y)| < \varepsilon \} \right)}{1 - \nu \left( \{ p \in \hat{\Omega}(a) : \forall y \in \text{supp} p_a^*, |p_a(y) - q(y)| < \varepsilon \} \right)} > \frac{1 - \varepsilon}{\varepsilon},
\]

then \( a \notin A^m(\nu) \).

Proof of Claim 1. Let

\[
G(\nu) = \int_{\Delta(Y)} \mathbb{E}_{p_a} [u(a, y)] \, d\nu_a(p_a) - \max_{b \in A} \int_{\Delta(Y)} \mathbb{E}_{p_b} [u(a, y)] \, d\nu_b(p_b)
\]

and notice that \( G \) is continuous by the Maximum Theorem, and \( a \in A^m(\nu) \) if and only if \( G(\nu) \geq 0 \). Suppose by way of contradiction that for every \( n \) there exists a \( \nu_n \) such that

\[
\frac{\nu_n \left( \{ p \in \hat{\Omega}(a) : \forall y \in \text{supp} p_a^*, |p_a(y) - q(y)| < 1/n \} \right)}{1 - \nu_n \left( \{ p \in \hat{\Omega}(a) : \forall y \in \text{supp} p_a^*, |p_a(y) - q(y)| < 1/n \} \right)} \geq \frac{1 - 1/n}{1/n}
\]

and \( a \in A^m(\nu) \), that is, \( G(\nu_n) \geq 0 \). Because \( \Delta(\Omega) \) is sequentially compact, \( (\nu_n)_{n \in \mathbb{N}} \) has a converging subsequence of \( (\nu_{n_i})_{i \in \mathbb{N}} \to \nu^* \). Thus \( \{ p \in \hat{\Omega}(a) : \forall y \in \text{supp} p_a^*, p_a(y) = q(y) \} \) = 1 and \( G(\nu^*) \geq 0 \), which would imply that \( a \in A^m(\nu^*) \), a contradiction.

Now fix such an \( \varepsilon \) and for every \( \alpha \in (0, 1) \), let

\[
f_\alpha = (1 - \alpha)p_a^* + \alpha q.
\]

Linearity of \( H \) in its first argument implies that for every \( \alpha \in (0, 1) \),

\[
\arg\min_{p_a : p \in \Theta} H(f_\alpha, p_a) = \{ q \}.
\]
Moreover, let \( g \) be defined as in the proof of Lemma 6. We have

\[
2g \left( (1 - \alpha)p_a^* + \alpha q, \varepsilon \right)
\]

\[
\geq \inf_{q' \in \Delta(Y) \setminus B_\varepsilon(q)} \sum_{y \in Y} [(1 - \alpha)p_a^*(y) + \alpha q(y)] \log q'(y) - \sum_{y \in Y} [(1 - \alpha)p_a^*(y) + \alpha q(y)] \log q(y)
\]

\[
\geq (1 - \alpha) \inf_{q' \in \Delta(Y) \setminus B_\varepsilon(q)} \sum_{y \in Y} p_a^*(y) [\log q'(y) - \log q(y)]
\]

\[
+ \alpha \inf_{q' \in \Delta(Y) \setminus B_\varepsilon(q)} \sum_{y \in Y} q(y) [\log q'(y) - \log q(y)]
\]

\[
\geq 0 + \alpha \inf_{q' \in \Delta(Y) \setminus B_\varepsilon(q)} \sum_{y \in Y} q(y) [\log q'(y) - \log q(y)] \geq 2\alpha \varepsilon^2,
\]

where the first inequality follows from the definition of \( g \) and the fact that the RHS minimizes over a larger set, the second inequality follows from concavity of the minimum, the third from the fact that \( q \) is a KL minimizer, and the fourth from Corollary 3.5 and Proposition 4.7 in Diaconis and Freedman (1990).

Let \( I = \max_{\alpha \in [0,1]} g ((1 - \alpha)p_a^* + \alpha q, \varepsilon) \). For every \( t \in \mathbb{N} \), let \( \alpha_t = 2t^{-\frac{1}{2}} \). If the empirical frequency is \( f_{a_t} \) after \( t \) periods, and only action \( a \) has been used, then from Lemma 6 and part (ii) of Assumption 2

\[
\frac{\mu_t \left( \{ p \in \hat{\Theta}(a) : \forall y \in \text{supp} \ p_a^* \mid p_a(y) - q(y) \mid < \varepsilon \} \right)}{1 - \mu_t \left( \{ p \in \hat{\Theta}(a) : \forall y \in \text{supp} \ p_a^* \mid p_a(y) - q(y) \mid < \varepsilon \} \right)} \geq \frac{\mu_0 \left( \{ p \in \hat{\Theta}(a) : \forall y \in \text{supp} \ p_a^* \mid p_a(y) - q(y) \mid < \varepsilon \} \right) \exp \left( t\alpha_t\varepsilon^2 \right)}{\frac{K\varepsilon}{I} \exp \left( \frac{1}{2} t^2 \hat{K} \varepsilon^2 \right)}.
\]

By Lemma 5 there exists an \( \varepsilon > 0 \) and a \( \hat{K} > 0 \) such that if the empirical frequency is \( f_t \) after \( t \) periods and \( |f_{a_t} - f_t| < \varepsilon t^{-\frac{1}{2}} \) then

\[
\min_{q' \in \Delta(Y) \setminus B_\varepsilon(q)} \sum_{y \in Y} f_t (\log q'(y) - \log q(y)) \geq t^{-\frac{1}{2}} \hat{K}
\]

and therefore

\[
\frac{\mu_t \left( \{ p \in \hat{\Theta}(a) : \forall y \in \text{supp} \ p_a^* \mid p_a(y) - q(y) \mid < \varepsilon \} \right)}{1 - \mu_t \left( \{ p : \mid p_a - q \mid < \varepsilon \} \right)} \geq \frac{K\varepsilon}{I} \exp \left( \frac{1}{2} t^2 \hat{K} \varepsilon^2 \right).
\]

Fix an outcome \( y^1 \in \text{supp} \ p_a^* \), and let \( \tilde{f}_t \) be the empirical frequency of the other \( |\text{supp} \ p_a^*| - 1 \) outcomes in the support of \( p_a^* \). Denote by \( \tilde{p}_a t \) the true probabilities of the same \( |\text{supp} \ p_a^*| - 1 \)
outcomes. By Assumption 2, $\tilde{f}_t \cdot t - \tilde{p}_a^* t$ is a $|\text{supp } p^*_a| - 1$ dimensional random walk under the
distribution $\tilde{p}_a^*$. The increments of this random walk have expected value of 0. The covariance
matrix for the increments is given by $\Sigma_{y,y'} = -2\tilde{p}_a^*(y)\tilde{p}_a^*(y')$ if $y \neq y'$ and $2\tilde{p}_a^*(y)(1 - \tilde{p}_a^*(y))$ if
$y = y'$\textsuperscript{27}. Fix an arbitrary $y \neq y^0$. If we let $D$ be the identity matrix in part M35 of Theorem
2.3 of Berman and Plemmons [1994], for every $y' \in Y$, we have that

$$2\tilde{p}_a^*(y')(1 - \tilde{p}_a^*(y')) = 2\tilde{p}_a^*(y') \sum_{y \neq y'} \tilde{p}_a^*(y) > 2\tilde{p}_a^*(y') \sum_{y \neq y',y^0} \tilde{p}_a^*(y)$$

proving that the matrix is diagonal dominant and therefore not singular. By the Central
Limit Theorem $(\tilde{f}_t - \tilde{p}_a^*)\sqrt{t}$ converges to a Normal random variable with mean 0 and covariance
matrix $\Sigma_{y,y'}$. Let $F_t = B_{\sqrt{t}} \left( \tilde{p}_a^* + \frac{1}{\sqrt{t}} (q - \tilde{p}_a^*) \right)$. We have that

$$\mathbb{P} \left[ \tilde{f}_t \in F_t \right] = \mathbb{P} \left[ \sqrt{t}(\tilde{f}_t - \tilde{p}_a^*) \in B_\varepsilon (q - \tilde{p}_a^*) \right]$$

Taking the limit $t \to \infty$ yields that

$$\lim_{t \to \infty} \mathbb{P} \left[ \tilde{f}_t \in F_t \right] = \mathbb{P} \left[ \tilde{Z} \in B_\varepsilon (q - \tilde{p}_a^*) \right]$$

where $\tilde{Z}$ is a random variable that is Normally distributed with mean $\tilde{0}$ and covariance
matrix $\Sigma_{y,y'}$. Consequently, if we denote as $E_t$ the event that $f_t \in F$, it follows that

$$\sum_{t=1}^{\infty} \mathbb{P} [E_t] = \infty.$$

\textsuperscript{27}To see this, the covariance between $1_y$ and $1_{y'}$ is given by:

$$\tilde{p}_a^*(y) (1 - E(1_y)) (0 - E(1_{y'})) + \tilde{p}_a^*(y') (0 - E(1_y))(1 - E(1_{y'})) + (1 - \tilde{p}_a^*(y') - \tilde{p}_a^*(y)) (0 - E(1_y))(0 - E\tilde{p}_a^*(1_{y'}))$$

$$= \tilde{p}_a^*(y) (1 - \tilde{p}_a^*(y)) (-\tilde{p}_a^*(y')) + \tilde{p}_a^*(y') (1 - \tilde{p}_a^*(y)) \tilde{p}_a^*(y') + (1 - \tilde{p}_a^*(y') - \tilde{p}_a^*(y)) (-\tilde{p}_a^*(y')) (1 - \tilde{p}_a^*(y'))$$

$$= -\tilde{p}_a^*(y) \tilde{p}_a^*(y') (1 - \tilde{p}_a^*(y')) (1 - \tilde{p}_a^*(y')) + \tilde{p}_a^*(y') \tilde{p}_a^*(y) (1 - \tilde{p}_a^*(y') - \tilde{p}_a^*(y))$$

$$= -\tilde{p}_a^*(y) \tilde{p}_a^*(y') (1 - \tilde{p}_a^*(y') - \tilde{p}_a^*(y') + 1 + \tilde{p}_a^*(y') + \tilde{p}_a^*(y) = -2\tilde{p}_a^*(y) \tilde{p}_a^*(y').
We furthermore have that
\[
\liminf_{t \to \infty} \frac{\sum_{s=1}^{t} \sum_{r=1}^{t} \mathbb{P} [E_s \text{ and } E_r]}{\left( \frac{1}{t} \sum_{s=1}^{t} \mathbb{P} [E_s] \right)^2} = \liminf_{t \to \infty} \frac{\frac{1}{t} \sum_{s=1}^{t} \sum_{r=1}^{t} \mathbb{P} [E_s \text{ and } E_r]}{\left( \frac{1}{t} \sum_{s=1}^{t} \mathbb{P} [E_s] \right)^2} \\
\leq \liminf_{t \to \infty} \frac{\frac{1}{t} \sum_{s=1}^{t} \mathbb{P} [E_s]}{\left( \frac{1}{t} \sum_{s=1}^{t} \mathbb{P} [E_s] \right)^2} = \lim_{t \to \infty} \frac{1}{\mathbb{P} [E_t]} \\
= \frac{1}{\mathbb{P} \left[ \hat{Z} \in B_\varepsilon (q - p^*_a) \right]}. 
\]

It thus follows from the Kochen-Stone lemma (see Kochen, Stone, et al. [1964] or Exercise 2.3.20 in Durrett [2008]) that
\[
\mathbb{P} \left[ \bigcap_{t=1}^{\infty} \bigcup_{s=t}^{\infty} E_s \right] \geq \mathbb{P} \left[ \hat{Z} \in B_\varepsilon (q - p^*_a) \right] > 0.
\]

The event \( \bigcap_{t=1}^{\infty} \bigcup_{s=t}^{\infty} E_s \) is invariant under finite permutations of the increments \( (1_{y_t=y^1}, \ldots, 1_{y_t=y|\text{supp } p^*_a|-1} - p^*_a) \) with different time indices, so the Hewitt–Savage zero–one law (see, e.g., Theorem 8.4.6 in Dudley [2018]) implies that the probability of the event \( \bigcap_{t=1}^{\infty} \bigcup_{s=t}^{\infty} E_s \) must equal zero or one. As the probability is strictly positive it must equal one.

This implies that \( f_t \in F_t \) infinitely often with probability 1. It follows that the agent will eventually take an action different from \( a \):
\[
\mathbb{P} [a_t \neq a \text{ for some } t] = 1.
\]

Thus the action can not converge to \( a \) with positive probability.

---

**Example 8.** To see that Proposition 1 (ii) does not hold without Assumption 2, modify Example 1 as follow: Keep the same action space \( \{a, b\} \), outcome space \( Y = \{0, 1\} \), suppose the agent correctly believes that the action has no impact on the outcome distribution, and that \( p^* = \frac{1}{2} \).
Now suppose that the agents assigns positive probabilities to the following countable set:

\[
\left\{ \frac{3}{4} \right\} \cup \left\{ \frac{1}{4} - \frac{1}{n^2} : n \geq 3 \right\},
\]

where distributions are indexed by the probability that they assign to outcome 1. Note that \( \frac{1}{4} \) is in \( \Theta \) even though it doesn’t exactly correspond to any of the agent’s conceivable models. Let \( p(n) = \frac{1}{4} - \frac{1}{n^2} \).

Suppose that the agent’s utility function is given by

\[
u(a, 0) = 0 = u(b, 1), u(a, 1) = 1, u(b, 0) = 4/5.
\]

Then \( b \) is not preferred to \( a \) for any beliefs with \(^{28}\)

\[
\nu \left( \{3/4\} \right) > 1/2,
\]

and it is strictly preferred to \( a \) if \(^{29}\)

\[
\nu \left( \{3/4\} \right) < 1/3.
\]

Then \( a \) is a Berk-Nash equilibrium but not a uniform Berk-Nash equilibrium, yet play can converge to it with positive probability from a prior \( \mu_0 \) we specify below.

\(^{28}\) Notice that the outcome distribution most favorable to action \( b \) and least favorable to action \( a \) is \( p(3) = 1/4 - 1/9 = 5/36 \). Therefore, if \( \nu_4 (\{3/4\}) > 1/2 \),

\[
\int_{\Delta(Y)} \mathbb{E}_p [u(a, y)] d\nu(p) \geq \sum_{n=3}^{\infty} p(n) u(a, 1) \nu(\{p(n)\}) + \frac{3}{4} u(a, 1) \nu(\{3/4\}) \geq \frac{5}{36} u(a, 1)(1 - \nu(\{3/4\})) + \frac{3}{4} u(a, 1) \nu(\{3/4\}) > 4/9
\]

and

\[
\int_{\Delta(Y)} \mathbb{E}_p [u(b, y)] d\nu(p) \leq \sum_{n=3}^{\infty} (1 - p(n)) u(b, 0) \nu(\{p(n)\}) + \frac{1}{4} u(b, 0) \nu(\{3/4\}) \leq \frac{31}{36} u(b, 0)(1 - \nu(\{3/4\})) + \frac{1}{4} u(b, 0) \nu(\{3/4\}) < 4/9.
\]

\(^{29}\) If \( \nu_4 (\{3/4\}) < 1/3 \),

\[
\int_{\Delta(Y)} \mathbb{E}_p [u(a, y)] d\nu(p) \leq \sum_{n=3}^{\infty} p(n) u(a, 1) \nu(\{p(n)\}) + \frac{3}{4} u(a, 1) \nu(\{3/4\}) \leq \frac{1}{4} u(a, 1)(1 - \nu(\{3/4\})) + \frac{3}{4} u(a, 1) \nu(\{3/4\}) < \frac{5}{12}
\]

and

\[
\int_{\Delta(Y)} \mathbb{E}_p [u(b, y)] d\nu(p) \geq \sum_{n=3}^{\infty} (1 - p(n)) u(b, 0) \nu(\{p(n)\}) + \frac{1}{4} u(b, 0) \nu(\{3/4\}) \geq \frac{3}{4} u(b, 0)(1 - \nu(\{3/4\})) + \frac{1}{4} u(b, 0) \nu(\{3/4\}) = \frac{7}{15}.
\]
Notice that
\[ 1 \leq p^* (1) \left( \frac{3}{4} \right)^{l_n} + p^* (0) \left( \frac{1}{4} \right)^{l_n} = \frac{1}{2} \left( \frac{3}{4} \right) + \frac{1}{2} \left( \frac{1}{4} + \frac{1}{n^2} \right) \]
where
\[ l_n = \frac{\log \left( 1 - \frac{1}{n^2} \right)}{\log \left( \frac{1}{1 - \frac{1}{n^2}} + \log 3 \right)} \]

Then by Dubins’ inequality (see, e.g., page 27 of Neveu, 1975), for all \( K_1 \) and \( K_2 \) there exists \( C_n \leq \frac{\frac{1}{2}}{\sum_{n=3}^{\infty} \frac{1}{n^2}} \) such that if \( \mu_0 (p(n)) \leq C_n \) and \( \mu_0 \left( \frac{3}{4} \right) > \frac{1}{2} \), the probability that
\[ \limsup_t \frac{\mu_t (p(n))}{\mu_t \left( \frac{3}{4} \right)} > \frac{1}{n^2} \]
\( K_1 \) is smaller then \( \frac{1}{n^2} \) \( K_2 \). Let \( \mu_0 (p(n)) = C_n \) and \( \mu_0 \left( \frac{3}{4} \right) = 1 - \sum_{n=3}^{\infty} C_n > \frac{1}{2} \), \( K_2 < \frac{1}{\sum_{n=3}^{\infty} \frac{1}{n^2}} \) and \( K_1 < \frac{1}{2 \sum_{n=3}^{\infty} \frac{1}{n^2}} \). By the union bound with probability
\[ 1 - K_2 \sum_{n=3}^{\infty} \frac{1}{n^2} > 0 \]
we have that
\[ \limsup_t \sum_{n=3}^{\infty} \frac{\mu_t (p(n))}{\mu_t \left( \frac{3}{4} \right)} \leq \sum_{n=3}^{\infty} \limsup_t \frac{\mu_t (p(n))}{\mu_t \left( \frac{3}{4} \right)} \leq K_1 \sum_{n=3}^{\infty} \frac{1}{n^2} < \frac{1}{2} \cdot \]

**Example 9.** This example shows that part (ii) of Assumption \( 2 \) is also needed for Proposition \( 1(ii) \). We modify example 1, keeping the same action space \( \{a, b\} \), and let the outcome space be \( Y = \{0, 1, 2\} \). Now suppose the agent correctly believes that the action has no impact on the outcome distribution, and that \( p^* (0) = 1 \). The prior \( \mu_0 \) of the agent is uniform over \( \Theta = \{p', p''\} \),

with
\[ p' (0) = \frac{1}{2} = p' (1) \quad p'' (0) = \frac{1}{2} = p'' (2) \cdot \]
Suppose that the utility \( u(a, 0) = u(a, 2) = u(b, 0) = u(b, 1) = 0, u(a, 1) = 1 \), and \( u(a, b) = 1/2 \) and the agent is myopic. Then \( a \) is not a uniform Berk-Nash equilibrium, since \( A (\delta_{p'}) = b \), and \( p'' \in \Theta (a) \). However, if the agent starts with prior \( \mu_0 \) they use action \( a \), observe
outcome 0 with probability 1, and never update their beliefs. Thus they play action a forever.

**Proof of Proposition 2.** Consider a uniformly strict Berk-Nash equilibrium a, an optimal policy π and κ ∈ (0, 1). By Lemma 2 there exists an ε such that if ν(Θε(a)) ≥ 1 − ε, then π(ν) = a.

Recall that for every l ∈ (0, 1), the function fi : P × P → P is defined by

\[ f_i(\bar{p}, p') = \sum_{y \in Y} p_a^*(y) \left( \frac{\bar{p}_a(y)}{p'_a(y)} \right)^l. \]

By Lemma 4 compactness of Θa(a), and since fi is lower semicontinuous in its first argument, there exists ε′ ∈ (0, ε) such that ̂Θε′(a) implies that fi(̂Θ, p') > 1 for all p' with p' ∉ ̂Θε′(a). Let K = (1 − ε′)/ε. Then

\[ \left( \frac{1 - \nu(\hat{\Theta}^\varepsilon(a))}{\nu(\hat{\Theta}^{\varepsilon}(a))} \right)^l < K \implies \frac{1 - \nu(\hat{\Theta}^\varepsilon(a))}{\nu(\hat{\Theta}^{\varepsilon}(a))} < \frac{\varepsilon}{1 - \varepsilon} \]

\[ \implies \nu(\hat{\Theta}^\varepsilon(a)) > 1 - \varepsilon \implies \pi(\nu) = a. \]

By Lemma 1, ̂Θε′(a) is compact, and therefore it admits a finite cover of

\[ \{ p \in \Theta : \|q_a - p_a\| \leq \varepsilon \}_{i=1}^n \]

where q.a ∈ ̂Θε′(a).

Let ε be such that ν(̂Θε′(a)) > 1 − ε implies that

\[ \left( \frac{1 - \mu_a(\hat{\Theta}^\varepsilon(a))}{\mu_a(\hat{\Theta}^{\varepsilon}(a))} \right)^l < K \frac{(1 - \kappa)}{n}. \]

Then if the agent starts with a belief μ0 such that ν0(Θ(a)) > ̂ε, A(μ0) = {a}. Moreover, by Lemma 3 Dubins’ inequality (see, e.g., page 27 of Neveu, 1975) and the union bound, there is a probability (1 − κ) that the positive supermartingale

\[ \left( \frac{1 - \mu_a^t(\hat{\Theta}^\varepsilon(a))}{\mu_a^t(\hat{\Theta}^{\varepsilon}(a))} \right)^l \]

39
never rises above $K$, so the action played is always $a$, and $\bar{\epsilon}$ satisfies the requirement of the statement.

**Proof of Proposition 3.** Let $\nu \in \Delta \left( \hat{\Theta}(a) \right)$ and fix any $\epsilon > 0$. We construct an initial belief $\nu_\epsilon$ that is $\epsilon$ close to $\nu$ but such that the action process never converges to $a$. The first step is to observe that since $a$ is not a uniform Berk-Nash equilibrium, there is KL-minimizer $\hat{p}$ for which $a$ is not optimal. Next, we associate each outcome distribution in the support of $\nu$ with an outcome distribution that is $\epsilon$ close to it but that it is not a KL minimizer, and we define $\bar{\nu}$ on the set of associated distributions. Finally, $\nu_\epsilon$ is defined as the linear combination between the Dirac measure on $\hat{p}$ and $\bar{\nu}$. Therefore, the unique KL-minimizer in the support of $\nu_\epsilon$ is $\hat{p}$, and we can use Proposition 1 to rule out convergence to $a$.

Since $a$ is not a uniform Berk-Nash equilibrium, there exists $\hat{p} \in \hat{\Theta}(a)$ with $a \notin A_m(\delta_\rho)$. Let $(C_{\epsilon,i})_{i=1}^n$ be a finite collection of open balls of radius $\epsilon$ in $\Delta \left( \Delta_Y^A \right)$ that covers $\hat{\Theta}(a)$ and such that for each $C_{\epsilon,i} \cap \hat{\Theta}(a) \neq \emptyset$. For every $C_{\epsilon,i}$, choose $q_{\epsilon,i} \in C_{\epsilon,i} \setminus \hat{\Theta}(a)$ whose existence follows from the assumption of the proposition.

Define $\Phi_\epsilon : \hat{\Theta}(a) \to 2^\Theta$ as $\Phi_\epsilon(p) = \{q_{\epsilon,i} : p \in C_{\epsilon,i}\}$.

Therefore, for every $q_{\epsilon,i}$, $\Phi_\epsilon^{-1}\{q_{\epsilon,i}\} = \{C_{\epsilon,i} \cap \hat{\Theta}(a)\}$ and the correspondence $\Phi_\epsilon$ is Borel measurable. Moreover, since $(C_{\epsilon,i})_{i=1}^n$ covers $\hat{\Theta}(a)$, $\Phi_\epsilon$ is nonempty and closed valued, so it has a measurable selection $\phi_\epsilon$ by the Kuratowski Selection Theorem (see, e.g., Theorem 18.13 in Aliprantis and Border, 2013).

Define $\nu_\epsilon = \epsilon \delta_\hat{p} + (1 - \epsilon) \bar{\nu}_\epsilon$ where

$$\bar{\nu}_\epsilon(C) = \nu \left( \phi_\epsilon^{-1}(C) \right).$$

Then $\nu_\epsilon \to \nu$, but $\text{argmin}_{p' \in \text{supp } \nu_\epsilon} H(p'_{a}, p'_{a}) = \{\hat{p}\}$, so by Proposition 1, the probability of converging to $a$ starting from belief $\nu_\epsilon$ is 0.

**Proof of Corollary 1.**

(i) $\Rightarrow$ (ii) Let $q \in \hat{\Theta}_a(a)$. Since $a$ is a uniform Berk-Nash equilibrium, there is $\nu_q \in \Delta(\Theta)$ with $\text{supp } \nu_q = \{p : p = \hat{\Theta}_a(a) : p_a = q\}$ such that $a \in A^m(\nu_q)$. Choose an arbitrary $\rho \in \Delta \left( \hat{\Theta}_a(a) \right)$ with $\text{supp } \rho = \hat{\Theta}_a(a)$, and for every $C \subseteq \Theta$, let

$$\nu(C) = \int_{\hat{\Theta}_a(a)} \nu_q(C) \, d\rho(q).$$
supp $\nu_a = \bigcup_{q \in \hat{\Theta}_a(a)} \text{supp } \nu^q_a = \hat{\Theta}_a(a)$. Moreover, by linearity of expected utility, $a \in A^m(\nu)$, as well, and therefore there is a policy $\pi$ that prescribes $a$ when the belief is $\nu$.

$(ii) \implies (i)$ If $\nu \in \Delta(\Theta)$ with $\text{supp } \nu_a = \hat{\Theta}_a(a)$, we have that

$$\arg\min_{p_a \in \text{supp } \nu} H(p^*_a, p_a) = \hat{\Theta}_a(a),$$

and

$$\{p \in \text{supp } \nu : p_a \in \hat{\Theta}_a(a)\} \subseteq \hat{\Theta}(a)$$

where these two sets are respectively the counterparts of $\hat{\Theta}_a(a)$ and $\hat{\Theta}(a)$ where the prior has been replaced by the initial belief $\nu$. But then, $a$ is not a uniform Berk-Nash equilibrium with respect to a prior equal to $\nu$ as well, and by part $(ii)$ of Proposition 1 the probability of convergence to $a$ starting from $\nu$ is $0$.

**Proof of Corollary 2.** $(i) \implies (ii)$ This part is just Proposition 2.

$(ii) \implies (i)$ If $a$ is not a uniformly strict Berk-Nash equilibrium, there exists $p \in \hat{\Theta}(a)$ such that $a \notin A^m(\delta_p)$. But then if we let $\nu = \delta_p$ we have that $\nu \left( \hat{\Theta}(a) \right) = 1$ and the agent will never play $a$.

**Proof of Proposition 4.** Since under causation neglect the agent believes that the action does not change the distribution over outcomes, every $p \in \Theta$ can be identified as an element of $\Delta(Y)$ and every belief $\nu \in \Delta(\Theta)$ can be identified as an element of $\Delta(\Delta(Y))$.

Consider a uniformly strict Berk-Nash equilibrium $a$. By the generalization of Weierstrass Theorem to lower-semicontinuous functions (see, e.g., Theorem 2.43 in Aliprantis and Border, [2013], $\hat{\Theta}(a)$ is a compact set. Therefore, $\Delta(\hat{\Theta}(a))$ is compact (see, e.g, Theorem 6.4 in Parthasarathy, [2005]).

To ease notation, in this proof for every $\varepsilon > 0$ and $q \in \Delta(Y)$ we let $Q_\varepsilon(q) = Q_{\varepsilon, \mu_0, a}(q)$. Denote as $\overline{Q}_\varepsilon(p^*_a)$ the closure of $Q_\varepsilon(p^*_a)$. By Proposition 2 there exists $\varepsilon' > 0$ such that if $\varepsilon' > \varepsilon$ and $\nu \left( \overline{Q}_\varepsilon(p^*_a) \right) > 1 - \varepsilon$ implies $A^m(\nu) = \{a\}$ the probability of playing $a$ forever starting from belief $\nu$ is larger than $1/2$.

By the Maximum Theorem, the correspondence $Q_\varepsilon$ is upper-hemicontinuous, so we can find a sequence of outcome realizations $y^t$ such that if $\hat{p}_t$ is the corresponding empirical frequency, it is sufficiently close to $p^*_a$ to have

$$Q_{\varepsilon'/2}(\hat{p}_t) \subseteq Q_{\varepsilon'}(p^*_a).$$
Therefore by Lemma 6, there exists a time period \( T \) such that for all \( t' > T \), if the empirical frequency \( \hat{p}_t \) of \( p_\alpha \), the agent assigns a relative probability higher than \( K \) to an \( \varepsilon' \)-Ball around \( \bar{p} \)

\[
\frac{\mu_\nu(Q_{\varepsilon'}(p_\alpha))}{1 - \mu_\nu(Q_{\varepsilon'}(p_\alpha))} > \frac{\mu_\nu(Q_{\varepsilon'/2}(\hat{p}_t))}{1 - \mu_\nu(Q_{\varepsilon'/2}(\hat{p}_t))} > \frac{K}{2}.
\]

Notice by replicating the outcome realizations \( y' \) sufficiently many time, we have a sequence of outcomes \( y'' \) such that the empirical frequency \( \hat{p}_{t'} = \hat{p}_t \) and \( t' > T \). Since \( p_\alpha \) is absolutely continuous with respect to to \( p_\alpha' \) for all \( \alpha' \in A \), the previous sequence of outcomes has positive probability, and after this outcome sequence the agent plays \( a \). By Lemma 3 and the law of iterated expectations, conditional on \( a \) being played \( \left( \frac{1 - \mu_\nu(Q_{\varepsilon'/2}(\hat{p}_t))}{\mu_\nu(Q_{\varepsilon'/2}(\hat{p}_t))} \right)^T \) is a positive supermartingale.

Then, by Dubins’ upcrossing lemma, there is a positive probability that this positive supermartingale never rises above \( 1/K^4 \), and therefore \( a \) is never stopped to be played. 

**Proof of Proposition 5.** Let \( b \) be a uniformly strict Berk-Nash equilibrium supported by the belief \( \nu \), and to ease notation, denote the prior \( \mu_0 \) as \( \mu \) throughout this proof. To focus on the additional difficulties involved in the bandit case, we prove the result in the case in which

\[
\arg\min_{p_b, p_\epsilon \in \Theta} H(p_b, p_\epsilon)
\]

is a singleton. The extension to multiple minimizers involves some additional continuity-compactness argument that mimic the ones in Proposition 2. Without loss of generality, we can take \( \nu \) to be such that for all \( a \in A \setminus \{b\} \), \( \nu_a = \delta_{p(a)} \) for some \( p(a) \in \Delta(Y) \) such that \( p_\alpha(y) > 0 \) implies \( p(a)(y) > 0 \). Assume for all \( a \in A \) and \( y \in Y \) such that \( p(a)(y) \) is rational, the general case with \( p(a)(y) \) possibly irrational involving a tedious additional continuity argument.

Let \( \{y(b)_i\}_{i=1}^{N_b} \) be a sequence of outcomes such that \( \hat{p}_{N_b}(y^{N_b}) = p(b) \), and let \( \{y(b)_i\}_{i=1}^{\infty} \) be its infinite repetition. By Lemma 6 for every \( \varepsilon > 0 \), there exists \( K_\varepsilon \) such that for all \( t > K_\varepsilon \), \( \mu_b (B_{\varepsilon}(p_b) | y(b)_t) > 1 - \varepsilon_b \).

Let \( \bar{\beta} \in (0, 1) \) and \( (\varepsilon_\alpha)_a \in \mathbb{R}_+^A \) be such that if \( \beta > \bar{\beta} \) and the belief \( \nu \) is such that \( \nu_b \in \{\mu_b\} \cup \{\mu_b : y(b)_i\} \cup \{\nu'_b : \nu_b'(B_{\varepsilon_b}(p_b)) > 1 - \varepsilon_b\} \), for all \( \alpha' \neq b \), \( \nu_{\alpha'}(B_{\varepsilon_b}(p_{\alpha'})) > 1 - \varepsilon_{\alpha'} \), then the highest Gittins index is the one of action \( b \). Their existence is guaranteed by \( \{a\} = A^m(\nu) \) and the definition of Gittins index. For each \( \beta > \bar{\beta} \), let \( \varepsilon_\beta < \varepsilon_b \) be such that if \( \nu_b (B_{\varepsilon_b}(p_b)) > (1 - \varepsilon_\beta) \) then the probability of converging to play action \( a \) is larger than \( \frac{1}{2} \) under any optimal policy given the discount factor \( \beta \), whose existence is guaranteed
by Proposition 2.

By Lemma 6 for every $a \neq b$ there exists a finite number $n_a$ such that after $n_a$ observations $\nu_a (B_{\varepsilon_a} (p_a) \mid \hat{p}_{na} = p (a)) > 1 - \varepsilon_a$. For every $a \neq b$, let $\bar{n}_a \geq n_a$ and $\{y (a_i)\}_{i=1}^{na}$ be a sequence of realizations such that $\hat{p}_{na} (a) = p (a)$ and $n_a$. Finally, let $n_b = K_{\beta}$. Then, the array $\{y (a_i)\}_{i=1}^{na} \subset \bar{a}_b$ has positive probability, the agent starts to play $a$ after at most $\sum_{a \in A} n_a$ periods, and with probability $\frac{1}{2}$ continues to play $a$ forever.

Proof of Proposition 7. We prove the statement for $\bar{a}$, the other is analogous. Denote the optimal policy used by the agent as $\pi$. Since the environment is strongly supermodular and $\bar{a}$ is a uniformly strict Berk-Nash equilibrium, by Proposition 2 there exists $\bar{p} \in \Theta$ and $K \in (0, 1)$ such that if $\nu (\{p : p > \bar{p}\}) > K$, then the probability that $a$ is going to be used forever is larger than $\frac{1}{2}$. Denote the highest outcome as $\bar{y}$. Since the environment is strongly supermodular, for every action $b \in A$,

$$\frac{\mu_{t+1} (\{p : p > \bar{p}\} \mid (a', y') \cup (b, \bar{y}))}{1 - \mu_{t+1} (\{p : p > \bar{p}\} \mid (a', y'), \beta)} > \frac{\mu_t (\{p : p > \bar{p}\} \mid (a', y'))}{1 - \mu_t (\{p : p > \bar{p}\} \mid (a', y'), \beta)}.$$ 

Therefore, there exists a finite number $n_b (b)$ such that if $a_t = b$ and $y_t = \bar{y}$ for all $t \leq n_b (b)$, then

$$\mu_t (\{p : p > \bar{p}\} \mid (a', y') \cup (b, \bar{y})) > K.$$ 

Consider the event $E$ that for all $b \in A$ and $t \leq n_b (b)$, $x_{t,b} = \bar{y}$. This event has strictly positive probability $P_\pi [E]$, and after some $\hat{T} \leq \sum_{b \neq \bar{a}} (n (b) - 1) + 1$, the policy of the agent prescribes action $\bar{a}$. After $\hat{T} + n (\bar{a})$,

$$\forall \tau \leq \hat{T} + n (\bar{a}), \forall y \in Y \quad P [x_{\tau, \bar{a}} = y | E] = P [x_{\tau, \bar{a}} = y].$$

Therefore, by Proposition 2 the probability of converging to $\bar{a}$ is at least $\frac{P_\pi [E]}{2}$. 

8.3 Extensions to Signals

We start by specifying the details of the sample space $\Omega = \mathcal{S}^\infty \times (Y^\infty)^A$. Formally, we consider the probability space $(\Omega, \mathcal{F}, P)$, where $\mathcal{F}$ is the discrete sigma algebra and the probability measure $P$ is the product measure induced by independent draws (across signal, actions, and time) according to $p^*$.

We denote the outcome observed by the agent in period $t$ after action $a_t$ by $y_t = x_{a_t, s_t, k}$, where $k$ is the number of times the agent has taken action $a_t$ after signal $s_t$ up and including
period \( t \). A (pure) policy \( \pi : \bigcup_{t=0}^{T} S_{t+1} \times A_{t} \times Y_{t} \rightarrow A \) specifies an action for every history \((s_{1}, a_{1}, y_{1}, s_{2}, a_{2}, y_{2}, \ldots, s_{t}, a_{t}, y_{t}, s_{t+1})\), and an initial action \( a_{1} \). Throughout, we denote by \( a_{t+1} = \pi(s_{t+1}, a', y') \) the action taken in period \( t \) where \((s_{t+1}, a', y')\) is a sequence of realized signals, actions, and outcomes.

**Proof of Proposition 8.** Since there is a finite number of actions, if behavior converges to \( \sigma \), there exists a finite period \( T \) such that \( a_{t}(s_{k}, x_{a_{s_{i}}}k)_{k \in \mathbb{N}, a_{s_{i}} \in S} = \sigma(s_{i}) \) for all later time periods \( t > T \). Consider an arbitrary period \( T \in \mathbb{N} \), strategy \( \sigma \in A^{S} \), and initial sequence of signals, actions, and outcomes \((s^{T}, a^{T}, y^{T})\). We will show that almost surely, if

\[
a_{t}(s_{k}, x_{a_{s},k})_{k \in \mathbb{N}, a_{s} \in S} = \sigma(s_{i}) \quad \forall \ t > T \}
\]

is a Berk-Nash equilibrium.

The initial sequence \((s^{T}, a^{T}, y^{T})\) corresponds to the event

\[
E = \{(s_{k}, x_{a_{s},k})_{k \in \mathbb{N}, a_{s} \in S} : \forall t \leq T, y_{t} = x_{a_{t},s_{t}}(\tau \leq t \quad a_{\tau} = a_{t})\}.
\]

Let \( K_{s} = |\{\tau \leq T : (s_{\tau}, a_{\tau}) = (s, \sigma(s))\}| \) be the number of times action \( \sigma(s) \) has been played after having observed signal \( s \) up to and including period \( T \).

Next, we define an ancillary stochastic process on our probability space, \((\nu_{t,\sigma})_{t \geq T} \).

The stochastic process is recursively defined as follows:

\[
\nu_{t,\sigma} = \mu_{0} \left( |(s^{T}, a^{T}, y^{T})| \right) \]

\[
\nu_{t+1,\sigma}(C) = \frac{\int_{p \in C} P_{s_{t+1},\sigma(s_{t+1})}(x_{\sigma(s_{t+1}),s_{t+1},N_{t+1},t})d\nu_{t,\sigma}(p)}{\int_{p \in P} P_{s_{t+1},\sigma(s_{t+1})}(x_{\sigma(s_{t+1}),s_{t+1},N_{t+1},t})d\nu_{t,\sigma}(p)},
\]

where \( N_{s,t} = K_{s} + |\{\tau : T < \tau \leq t : s_{\tau} = s\}| \). Here \( \nu_{t,\sigma} \) would be the agent’s period \( t \) posterior belief after having observed the initial sequence \((s^{T}, a^{T}, y^{T})\) and having followed strategy \( \sigma \) in all periods \( \tau \in (T, t] \). In general, the process \((\nu_{t,\sigma})_{t \geq T} \) will not coincide the evolution of the agent’s beliefs \((\mu_{t})_{t \geq T}\): the agent may choose an action different from the one prescribed by strategy \( \sigma \), leading to different observations and posteriors.

We will prove a statement about \((\nu_{t,\sigma})_{t \geq T} \) that holds almost everywhere in \( E \), and then we argue that \( E \) can partitioned in two subsets, the sequences along which the agent stops to follow \( \sigma \) at some period \( t > T \), and the sequences along which \( \sigma \) is played forever. In the former subset, clearly we do not have convergence to \( \sigma \) after \((s^{T}, a^{T}, y^{T})\). In the latter, the posterior belief \((\mu_{t})_{t \geq T} \) will coincide with \((\nu_{t,\sigma})_{t \geq T} \), and therefore it will inherit its properties.\(^{30}\)

\(^{30}\)Notice that the evolution of \( \nu_{t,\sigma} \) only depends on the sequences \((s_{T+k}, x_{\sigma(s'),s',K_{s+k}})_{k \in \mathbb{N}, s' \in S} \).
Claim 2. For every neighborhood $C$ of \arg\min_{p \in \text{supp}\, \mu_T} \sum_{s \in S} \zeta(s) H \left( p_{s,\sigma(s),s}^*, p_{\sigma(s),s} \right)$,

$$\mathbb{P} \left[ \lim_{k \to \infty} \nu_{k,\sigma}(C) = 1 \mid E \right] = 1.$$ 

Proof of Claim 2. To prove the claim, notice that the probability of the sequence $\left( s_{T+k}, x_{\sigma(s'),s',K_{s'+k}} \right)_{k \in \mathbb{N}, s' \in S}$ conditional on $E$ is the product measure induced by independent draws. Let $Z = S \times Y$. If the strategy of the agent were exogenously fixed at $\sigma$ for every period $t \geq T$, the signal and outcome pair $(s_t, y_t)$ would be i.i.d with probability $p_{\sigma}^*(s, y) : = \zeta(s) p_{s,\sigma(s)}^*(y)$, whereas $p \in \text{supp}\, \mu_T$ would assign $p_{\sigma}(s, y) : = \zeta(s) p_{s,\sigma(s)}(y)$.

Since (1) $\text{supp}\, \mu_T$ is a compact set, (2) $\mu_T$ is a Borel probability measure on $\text{supp}\, \mu_T$, and assumptions (ii), (iii), and (iv) of Berk (1966) are satisfied given the absolute continuity assumption and the finite number of outcomes, it follows from Berk’s Theorem that there exists $E'$ with $\mathbb{P} \left[ E' \mid E \right] = 1$ such that for every neighborhood $C$ of

$$\arg\min_{p \in \text{supp}\, \mu} H \left( p_{\sigma}^*, p_{\sigma} \right) = \arg\min_{p \in \text{supp}\, \mu} \left( - \sum_{(s,y) \in S \times Y} \zeta(s) p_{s,\sigma(s)}^*(y) \log \zeta(s) p_{s,\sigma(s)}(y) \right),$$

we have $\lim_{k \to \infty} \nu_{k,\sigma}(C) = 1$ on $E'$.

Given our absolute continuity hypothesis, $\text{supp}\, \mu_T = \Theta$, and this in turns implies that

$$\arg\min_{p \in \text{supp}\, \mu_T} \sum_{s \in S} \zeta(s) H \left( p_{s,\sigma(s),s}^*, p_{\sigma(s),s} \right) = \hat{\Theta}(\sigma).$$

This proves that if the agent starts to follow $\sigma$ from $(s_T, a_T, y_T)$, then

$$1 = \lim_{k \to \infty} \nu_{k,\sigma}(C) = \mu_t(C) \text{ a.s.}$$

for every neighborhood $C$ of $\hat{\Theta}(\sigma)$.

By the 0-1 Law, $E'$ can be chosen such that if $\left( s_{T+k}, x_{\sigma(s'),s',K_{s'+k}} \right)_{k \in \mathbb{N}, s' \in S} \in E'$, for all $\bar{s} \in S$, $s_t = \bar{s}$ infinitely many times. Suppose that $\sigma$ is not a Berk-Nash equilibrium. Since
$P$ is compact, $\Delta(P)$ is also compact. Moreover, since $P$ is separable, $\Delta(P)$ is metrizable by the Levy-Prokhorov metric and thus sequentially compact. Hence the sequence $(\nu_{k,\sigma})_{k\in\mathbb{N}}$ has an accumulation point $\nu \in \Delta(\hat{\Theta}(\sigma))$. Since $\sigma$ is not a Berk-Nash, there exists $\bar{s} \in S$ such that $\sigma(\bar{s}) \notin A^m(\nu, \bar{s})$, and from the definition of weak convergence of measure and Berk-Nash equilibrium, there exists $\tau'$ such that for $\tau \geq \tau'$

$$\int_{\Delta(P)} \mathbb{E}_{p_{\bar{s},\sigma(\bar{s})}}[u(s, \sigma(\bar{s}), y)] d\nu_{\tau,\sigma}(p) < \max_{a \in A} \int_{\Delta(P)} \mathbb{E}_{p_{s,a}}[u(s, a, y)] d\nu_{\tau,\sigma}(p).$$

Since $s_t = \bar{s}$ infinitely often on $E'$, $s_\tau = \bar{s}$ for some $\tau \geq \tau'$, and the agent will play an action $a_s = \sigma(\bar{s})$. This contradicts convergence to $\sigma$ after $(s^T, a^T, y^T)$. Since both $\sigma$ and $(s^T, a^T, y^T)$ were chosen arbitrarily, the result holds.

**Proof of Proposition 9.** Under causation and signal neglect, $\Theta \subseteq \Delta(\Delta(Y))$. Consider a uniformly strict Berk-Nash equilibrium $\sigma$. By an obvious extension of Lemma 1 to the case with signals, $\Delta\left(\hat{\Theta}(\sigma)\right)$ is compact. Similarly, since $S$ is compact and $\sigma$ is the unique optimal best reply strategy at the beliefs in $\Delta\left(\hat{\Theta}(\sigma)\right)$, Lemma 2 can be extended to guarantee that there exists $\varepsilon \geq 0$ such that if

$$\nu\left(Q_{\varepsilon}(\bar{p}_\sigma)\right) \geq (1 - \varepsilon)$$

then the myopic best reply to $\nu$ is $\sigma$. By the same argument of the proof of Proposition 2 there exists an $l \in (0, 1)$ and $\varepsilon' \in (0, \varepsilon)$, such that if $p \in Q_{\varepsilon'}(\bar{p}_\sigma)$ and $p' \notin Q_{\varepsilon}(\bar{p}_\sigma)$ then $f_l(p, p') \geq 1$.

Using the Maximum Theorem again we can find a sequence of outcome realizations $y^t$ such that if $\hat{p}_t$ is the corresponding empirical frequency, it is sufficiently close to $\bar{p}_\sigma$ to have

$$Q_{\varepsilon/2}(\hat{p}_t) \subseteq Q_{\varepsilon}(\bar{p}_\sigma).$$

Therefore by Lemma 6, there exists a time period $T$ such that for all $t' > T$, if the empirical frequency $\hat{p}_{t'} = \hat{p}_t$, the agent assigns a relative probability higher than $K$ to an $\varepsilon$ Ball around $\bar{p}$

$$\frac{\mu_{t'}(Q_{\varepsilon}(\bar{p}_\sigma))}{1 - \mu_{t'}(Q_{\varepsilon}(\bar{p}_\sigma))} \geq \frac{\mu_{t'}(Q_{\varepsilon/2}(\bar{p}_\sigma))}{1 - \mu_{t'}(Q_{\varepsilon'}(\bar{p}_\sigma))} > 2 \frac{(1 - \varepsilon)}{\varepsilon}.$$
for all \((a, s) \in A \times S\), the previous sequence of outcomes has positive probability, and after this outcome sequence the agent plays \(\sigma\). By Lemma 3 and the law of iterated expectations, conditional on \(a\) being played \(\left(\frac{1-\mu_t^e(Q_{e/2}(\hat{p}_t))}{\mu_t^e(Q_{e/2}(\hat{p}_t))}\right)^t\) is a positive supermartingale.

Then, by Dubins’ upcrossing lemma, there is a positive probability that this positive supermartingale never rises above \(\hat{\epsilon}\left(\frac{\hat{\epsilon}}{1-\hat{\epsilon}}\right)\), that in turns imply that \(\mu_t^e(Q_{e/2}(\hat{p}_t))\) never goes below \((1 - \hat{\epsilon})\) and therefore \(\sigma\) is always played after the sequence \(y_t^t\).

References


