Limit Points of Endogenous Misspecified Learning*

Drew Fudenberg† Giacomo Lanzani‡ Philipp Strack§

First posted version: March 10, 2020

This version: January 22, 2021

Abstract

We study how an agent learns from endogenous data when their prior belief is misspecified. We show that only *uniform Berk-Nash equilibria* can be long-run outcomes, and that all *uniformly strict Berk-Nash equilibria* have an arbitrarily high probability of being the long-run outcome for some initial beliefs. When the agent believes the outcome distribution is exogenous, every uniformly strict Berk-Nash equilibrium has positive probability of being the long-run outcome for any initial belief. We generalize these results to settings where the agent observes a signal before acting.

Keywords: Misspecified learning, Bayesian consistency, Berk-Nash equilibrium

*We thank Aislinn Bohren, Alex Wolitzky, Annie Liang, Ben Golub, Cuimin Ba, Demian Pouzo, Glenn Ellison, Ignacio Esponda, Harry Pei, Kevin He, Mira Frick, Roberto Corrao, Ryota Iijima, Yuhta Ishii, and Yuichi Yamamato for helpful conversations, and National Science Foundation grants SES 1643517 and 1951056, the Guido Cazzavillan Scholarship, and Sloan foundation 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 data generating process. Given the complexity of the real world, such misspecification is plausible in many settings, and has been studied in a wide range of applications.

We consider a general environment with finite actions and outcomes and – unlike most past work – do not restrict the agent's prior belief to have a finite support or any specific functional form. In this environment, the agent's prior is a belief over the set of action-contingent outcome distributions, so the agent is misspecified if they assign probability 0 to a neighborhood of the true map from actions to distribution over outcomes. The agent's prior 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 actions and their stability properties. First, Theorem 1 shows that regardless of the agent's discount factor, if play converges to an action a, that action is a *uniform Berk-Nash equilibrium*. Uniform Berk-Nash equilibrium, which we introduce in this paper, is a refinement of Berk-Nash equilibrium (Esponda and Pouzo (2016)). Berk-Nash equilibrium requires that the action is myopically optimal against some belief in the support of the prior that minimizes the Kullback-Leibler (KL) divergence between the subjective and true outcome distributions given that the agent plays a— that is, the action must be a best response to a "KL minimizer". Uniform Berk-Nash equilibrium strengthens this by requiring that the action is a best response to *any* beliefs with support on these KL minimizers. Intuitively, limit points correspond to myopic optimization even when the agent is not myopic because play will not converge until the agent no longer perceives an "experimentation value" from non-myopic play; the intuition for the uniformity requirement is that when play converges, the agent's beliefs oscillate over all of the KL-minimizing beliefs.

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 arbitrarily high probability for some open set of initial beliefs. Theorem 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 KL-minimizing

outcome distributions (as opposed to being true for at least one of them).

We say that an action is *positively attractive* if there is positive probability that it is the limit outcome under every optimal policy for *every* full-support prior belief. When the agent believes (either rightly or wrongly) that the distribution of outcomes is the same for all actions, or in a "subjective bandit problem" where the agent believes that the outcomes observed when playing one action are uninformative about the outcome distributions induced by other actions, we obtain partial converses to Theorem 1: All uniformly strict Berk-Nash equilibria are positively attractive. 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 establish in Appendix A.3 that with probability one beliefs concentrate exponentially fast around the KL minimizers.[1] We use this concentration result to guarantee that the agent starts to play the equilibrium action with positive probability. We then use the stability result from Theorem 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. In this setting, 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 (potentially) payoff relevant signal before taking an action. Here too a limit action must be a uniform Berk-Nash equilibrium. Moreover, if the agents ignore the predictive value of the signals, i.e., the signals are *subjectively uninformative*, every uniformly strict Berk-Nash equilibrium is positively attractive.

We illustrate our findings in three economic examples: a monopolist that is misspecified about the demand function, a central bank choosing an exchange-rate policy, and a seller that observes a signal and then decides whether to make an investment.

Related Work

Given an objective data generating process, a model is a KL-minimizer if it maximizes the expected likelihood assigned to a randomly drawn outcome over the support of the agent's prior. Berk (1966) shows that the beliefs of an agent asymptotically concentrate on the set of KL minimizers when the data generating process is exogenous. In many economic applications, actions and associated signal distributions are not fixed, but change endogenously over time depending on an action taken by the agent, so the agent's misspecification has

[1]This result is in the spirit of Diaconis and Freedman (1990), which assumes a full-support prior and thus rules out misspecification.

implications for what they observe and thus for their long-run beliefs. Arrow and Green (1973) gives the first general framework for this problem, and Nyarko (1991) points out that the combination of misspecification and endogenous observations can lead to cycles.

There has been a surge of theoretical work on misspecified learning since the seminal work of Esponda and Pouzo (2016), which defines Berk–Nash equilibrium. This is a relaxation of Nash equilibrium (for games as as well as decision problems) that replaces the requirement that players' beliefs are correct with the requirement that each player's belief minimizes the KL divergence between their belief and their observations over the support of their prior. Esponda and Pouzo (2016) shows that Berk-Nash equilibrium is a necessary property for limit points when the payoff function is subject to small i.i.d. random shocks as in Fudenberg and Kreps (1993), and that it is sufficient if in addition the agent is willing to incur asymptotically negligible optimization losses. Esponda and Pouzo (2019) generalizes Berk-Nash equilibrium to Markov decision problems.

Fudenberg, Romanyuk, and Strack (2017) and Bohren and Hauser (2020) provide necessary and sufficient conditions for actions to converge when the support of the agent's prior contains only two points.[2] Heidhues, Kőszegi, and Strack (2018) and He (2019) provide conditions for global convergence of play of a non-myopic agent in a environments with additively separable payoffs that satisfy strong supermodularity restrictions, where the Berk-Nash equilibrium is unique. Heidhues, Kőszegi, and Strack (2021) establishes convergence to a Berk-Nash equilibrium in environments with a normal prior and normal signals. 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.

The most closely related papers are Esponda, Pouzo, and Yamamoto (2019) (henceforth EPY) and Frick, Iijima, and Ishii (2020) (henceforth FII). EPY uses stochastic approximation to determine when the agent's *action frequency* converges in an environment with finitely many actions and fairly general priors. We provide a sharper characterization of when play converges to a single action in the long run, but our results do not characterize the long-run distribution when this convergence does not occur. Corollary 2 in the Appendix combines our results with theirs to derive new results about the limiting action frequencies. FII provides conditions for local and global convergence of the agent's beliefs without explicitly modelling the agent's actions when the agent's prior has finite support.[3]

[2]Bohren and Hauser (2020) considers myopic agents in discrete time; Fudenberg, Romanyuk, and Strack (2017) analyzes a continuous time model with Brownian noise without assuming myopia.

[3]Neither model nests the other. FII assumes finite priors, and impose a continuity assumption that our model can but need not satisfy. Conversely, we rule out the continuum of actions assumed by FII.

Our paper complements the literature on long-run behavior in misspecified models in three ways: First, we establish that without the asymptotically vanishing payoff perturbations of Esponda and Pouzo (2016), play never converges to a non-uniform Berk-Nash equilibrium. This uniformity refinement has no analog in FII because it is with respect to the optimality of actions. Second, we introduce conditions under which an action has positive probability of being the long-run outcome from any initial belief. Finally, we provide the first necessary and sufficient conditions for the choices of forward-looking misspecified agents to converge to a myopic best reply to their beliefs.[4]

Misspecified agents are featured in work in a wide range of fields. There are many examples in behavioral economics, such as the "law of small numbers," the "hot-hand fallacy," the winner's curse, and the link between overconfidence and prejudice.[5] Macroeconomists have been interested in misspecified learning both in the form of misspecified least-squares predictions as well as more sophisticated models of updating and inference.[6] In organizational economics, misspecification has been used to explain e.g. the role of corporate culture and the low rate and low number of minority inventors. In public economics, misspecification helps explain over or under reaction to changes in tax schedules. And in political economy, misspecification has been used to explain the recurrence of populism and political polarization.[7] There is also a related literature on misspecified social learning.[8]

In addition to papers that consider misspecified Bayesian agents, there is a literature that studies the long-run outcomes under learning heuristics that might be used when people are unable to formulate a probabilistic assessment of the data generating process. Many of these heuristics feature a form of neglect of the relevant elements of the environment, similar to the ones we consider in our Section 4, e.g. Tversky and Kahneman (1973), Rabin and Schrag (1999), and Jehiel (2018). More recently, Gagnon-Bartsch, Rabin, and Schwartzstein (2018), Fudenberg and Lanzani (2020), and He and Libgober (2020) analyze various processes that can lead agents to change or expand the set of models they consider possible.

[4]Theorem 4 of Esponda and Pouzo (2016) shows that Berk-Nash is necessary under weak identification and payoff perturbations. Other work either assumes myopia or don't obtain convergence to myopic best reply.
[5]See Kagel and Levin (1986), Rabin and Vayanos (2010), Heidhues, Kőszegi, and Strack (2019).
[6]Bray (1982), Bray and Savin (1986), Cho and Kasa (2015), Cho and Kasa (2017), Molavi (2019).
[7]See Gibbons, LiCalzi, and Warglien (2019) and Bell et al. (2019) for organizational economics, Rees-Jones and Taubinsky (2020) and Morrison and Taubinsky (2019) for public economics, and Levy, Razin, and Young (2020) and Eliaz and Spiegler (2018) for political economy.
[8]See Bohren (2016), Bohren and Hauser (2020), Frick, Iijima, and Ishii (2019), Gagnon-Bartsch (2016), and Mailath and Samuelson (2020).

2 The Model

2.1 Setup

Actions, Utilities and Objective Outcome Distributions We study a sequence of choices made by a single agent. In each period $t \in \{1, 2, 3, \ldots\}$ the 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.[9] Second, the action, paired with the realized outcome, determines the flow payoff of the agent via the utility function $u : A \times Y \to \mathbb{R}$.

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, and let $p_a \in \Delta(Y)$ denote the a-th component of $p \in P$. We endow P with the sup-norm topology, and denote by $B_\varepsilon(p)$ the ball of radius ε around $p \in P$.[10]

The agent's uncertainty is captured by a prior belief $\mu_0 \in \Delta\big(P\big)$, where $\Delta\big(P\big)$ denotes the metric space of Borel probability measures on P endowed with the topology of weak convergence of measures.

Definition 1. The *conceivable outcome distributions* are the elements of $\Theta = \operatorname{supp} \mu_0$. The agent is *correctly specified* if $p^* \in \Theta$, i.e. the objective distribution is conceivable.

Throughout the paper, we will maintain the following assumption:

Assumption 1 (Regularity)**.**

(i) For all $p \in \Theta$, $a \in A$, and $y \in Y$, $p_a(y) > 0$ if and only if $p_a^*(y) > 0$.

(ii) The prior μ_0 has *subexponential decay*: there is $\Psi : \mathbb{R}_+ \to \mathbb{R}_{++}$ such that for every $p \in \Theta$ and $\varepsilon > 0$ we have $\mu_0(B_\varepsilon(p)) \geqslant \Psi(\varepsilon)$ with $\lim_{n \to \infty} \Psi(K/n) \exp(n) = \infty$ for all $K > 0$.

Assumption 1(i) requires that the outcomes that the agent thinks are possible are the same as those that objectively have positive probability. This assumption guarantees that

[9] We denote objective distributions with a superscript *.

[10] For every finite dimensional vector v, we let $||v|| = \max_i v_i$ denote the supremum norm.

Bayes rule is always well defined.[11] Assumption 1(ii) extends Diaconis and Freedman (1990)'s notion of ϕ-positivity to the misspecified case, and adds the requirement that the bounding Ψ vanishes at a subexponential rate around 0. It is always satisfied by priors with a density that is bounded away from 0 on their support, and by priors with finite support.[12]

Our specification allows the agent's subjective uncertainty to be correlated across actions. For example, if the agent is certain that every action generates the same outcome distribution, then they believe the outcome distributions are perfectly correlated across actions.

Updating Subjective Beliefs We assume throughout that the agent updates their beliefs using Bayes rule. Denote by $\mu_t(\cdot \mid (a^t, y^t))$ the subjective belief the agent obtains using Bayes rule after action sequence $a^t = (a_s)_{s=1}^t$ and outcome sequence $y^t = (y_s)_{s=1}^t$,

$$\mu_t(C \mid (a^t, y^t)) = \frac{\int_{p \in C} \prod_{\tau=1}^t p_{a_\tau}(y_\tau) d\mu_0(p)}{\int_{p \in P} \prod_{\tau=1}^t p_{a_\tau}(y_\tau) d\mu_0(p)}. \tag{Bayes Rule}$$

Since the agent's prior has support Θ, their posterior belief does as well. We sometimes suppress the dependence of the posterior belief on the realized sequence and just write μ_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. 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 let $a_{t+1} = \pi(a^t, y^t)$ denote the action taken in period t. The objective action-contingent probability distribution p^* and a policy π induce a probability measure $\mathbb{P}_\pi$ on $(a_\tau, y_\tau)_{\tau=1}^\infty$.[13] Standard results guarantee that there is an optimal policy π that is Markovian and depends on the history only through the agent's beliefs; we restrict attention to such policies.

Given a belief $\nu \in \Delta(\Theta)$ we denote by ν_a the belief over outcome distributions associated with action a, i.e. $\nu_a(C) = \int \mathbf{1}_{p_a \in C} d\nu(p)$ for all Borel sets $C \subseteq \Delta(Y)$. 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

[11] Assumption 1(i) is satisfied in most applications but it is stronger than necessary. The "if" part is enough for all our results except the non-myopic version of Theorem 1. In the Supplemental Material, B.2 we show how this result can be extended to the case where the "only if" part is not satisfied.

[12] Dirichlet priors also satisfy Assumption 1(ii), even though they do vanish at the edge of their support. Fudenberg, He, and Imhof (2017) shows by example that even correctly specified Bayesian updating can behave oddly when the prior vanishes exponentially quickly.

[13] We spell out the details of this measure at the start of the Appendix.

p_a. $A^m(\nu)$ denotes the set of myopically optimal actions given belief ν, i.e.,

$$A^m(\nu) = \operatorname*{argmax}_{a \in A} \int_{\Delta(Y)} \mathbb{E}_{p_a}[u(a,y)]\, d\nu_a(p_a).$$

2.2 Forms of Misspecification

Our model encompasses many sorts of misspecified learning, including the following:

Subjectively Exogenous Problems We say that there are subjectively exogenous outcomes when the agent believes that the realized outcome is not affected by the chosen action.

Definition 2. Outcomes are *subjectively exogenous* if for every $a, a' \in A$ and every $p \in \Theta$, we have $p_a = p_{a'}$.

Note that the agent can believe in exogenous outcomes independent of whether or not the action really does influence the distribution; if the action does influence the outcome and the agent ignores this we say the agent exhibits causation neglect. An agent who thinks the outcome distribution is exogenous updates their beliefs as if they faced an i.i.d. environment. We will establish that the beliefs in this setting concentrate on the conceivable outcome distributions closest to the empirical average. We use this result to show that if a is a uniformly strict Berk-Nash equilibrium, it is positively attractive.

Subjective Bandit Problems The other extreme case encompassed by our setup is where the agent thinks that they face a bandit problem, i.e. they believe that the distributions over outcomes induced by different actions are independent. This corresponds to the case where the agent's prior μ_0 is a product measure.

Definition 3. An agent faces a *subjective bandit problem* if $\mu_0 = \times_{a \in A} \mu_{0,a} \in (\Delta(\Delta(Y)))^A$.

We show that uniformly strict Berk-Nash equilibria are positively attractive in this setting as well, provided that the agent is sufficiently patient.

One Dimensional Problems In one-dimensional problems, the agent's uncertainty is summarized by a parameter $\gamma \in \mathbb{R}$. The parameter determines the distribution over outcomes through a function ϕ which maps parameters to action-dependent outcome distributions. Formally, the support of the agent's prior is contained in the image of this function ϕ.

Definition 4. The problem is *one-dimensional* if there exists $\Gamma \subseteq \mathbb{R}$ and a function $\phi : \Gamma \to P$ such that $\Theta \subseteq \{\phi(\gamma) \colon \gamma \in \Gamma\}$. A one-dimensional problem is *supermodular* if A can be ordered such that $(\gamma, a) \mapsto \mathbb{E}_{\phi(\gamma)_a}[u(a, y)]$ is supermodular.

EPY provides a sufficient condition for actions to converge in one-dimensional problems that are supermodular. Heidhues, Kőszegi, and Strack (2018) shows that a unique Berk-Nash equilibrium is globally attracting in supermodular problems where the outcomes are real numbers and ϕ is an additive shift. Our Example 9 shows that their result does not hold in our more general setting: a unique (and uniformly strict) Berk-Nash equilibrium may not be positively attractive. Under a stronger version of supermodularity, our positive attractiveness results do extend to extremal uniformly strict Berk-Nash equilibria.

Finite Support Another common assumption is that the support of the prior is finite. With a finite-support prior, if behavior converges to an action a, a is a best reply to all outcome distributions that minimize the Kullback-Leibler divergence from p_a^*, so it is a *uniform* Berk-Nash equilibrium. However, Example 6 shows that non-uniform Berk-Nash equilibria can be limit points when the support of the prior is infinite if Assumption 1(ii) is not satisfied.

Signals 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 the 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 μ over P, where $p_{s,a}(y)$ denotes the probability under $p \in P$ of outcome y after observing signal s playing action a. Extending the model to signals lets us incorporate the stochastic payoff perturbations assumed in Esponda and Pouzo (2016). It also lets us model cases where the agent mistakenly thinks that some information they observe is uninformative.

3 Limit Points and Berk-Nash Equilibria

We are interested in when the agent's actions converge, and their possible limit points. 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 the agent doesn't have an action choice, and conversely actions can oscillate with fixed beliefs if the agent is indifferent.[14]

We say that the action process *converges to action* a if there exists a time period $T \in \mathbb{N}$ such that $a_t = a$ for all time periods $t > T$. Action a is a *limit action* if the action process converges to a with positive probability under some optimal policy π.[15] Note that there may be several optimal policies for a given prior; 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 for which a is myopically optimal, and which assigns 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

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

Note that $-H(q, q')$ is the expected log likelihood of an outcome under subjective distribution q' when the true distribution is q, so q' with smaller $H(q, q')$ is a better explanation for the outcome frequency q. The *Kullback-Leibler* (KL) divergence between p_a^* and p_a is given by $H(p_a^*, p_a) - H(p_a^*, p_a^*)$, so any p_a that minimizes $H(p_a^*, p)$ also minimizes the KL divergence between p_a^* and p_a.

Recall that p_a denotes the outcome distribution p assigns to action a. For each a, let

$$\hat{\Theta}(a) = \underset{p \in \Theta}{\operatorname{argmin}} H(p_a^*, p_a) \subseteq \Theta \tag{1}$$

denote the set of conceivable action-contingent outcome distributions that minimize the KL divergence relative to the true distribution p_a^* given that the agent plays a. Note that the elements of $\hat{\Theta}(a)$ specify an outcome distribution for each action $a' \in A$, even though $\hat{\Theta}(a)$ only depends on the distributions corresponding to a. We call $\hat{\Theta}(a)$ the set of *KL minimizers* for action a.[16]

Berk (1966) established that the agent's beliefs concentrate on $\hat{\Theta}(a)$ if they always play a. This motivates Esponda and Pouzo (2016)'s notion of a Berk-Nash equilibrium. We

[14] The fact that beliefs can oscillate under a fixed action is the driving force behind the uniformity requirement in several of our results, such as Theorem 1.

[15] Formally, there exists a measurable set $C \subseteq A^\infty \times Y^\infty$ with $\mathbb{P}_\pi[C] > 0$ such that a_t converges to a in C.

[16] Note that if $p^* \in \Theta$ then each minimizing p explains the observed outcome distribution perfectly, $p_a = p_a^*$. In particular this is true if μ_0 has full support.

introduce variations of this concept to capture different senses in which an action is or is not a long-run outcome of the agent's learning process.

Definition 5. Two action-contingent outcome distributions p and p' are *observationally equivalent under action* a if $p_a = p'_a$. We denote by $\mathcal{E}_a(p) \subseteq \Theta$ the set of action-contingent outcome distributions in Θ that are observationally equivalent to p under a.

Definition 6.

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

(ii) Action a is a *strict BN-E* 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 BN-E* if for all KL minimizers $p \in \hat{\Theta}(a)$ there exists a belief $\nu \in \Delta(\mathcal{E}_a(p))$ such that $a \in A^m(\nu)$.

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

Uniformity requires that for each class of observationally equivalent KL minimizers for action a, there is a belief concentrated on that class for which a is the myopically optimal choice.[17] The difference between BN-E and uniform BN-E disappears in the correctly specified case, where both concepts coincide with self-confirming equilibrium. In settings where the KL minimizer is unique, the uniformity requirement has no bite. However, in frameworks with additional structure, such as symmetry or parametric restrictions, multiple KL minimizers can arise naturally. For example, suppose that agent's payoff depends on the color y of a ball drawn from an urn, and the agent's action is to bet on the color of the drawn ball. The agent correctly believes their action has no impact on the distribution of outcomes. The urn has 6 balls: 4 of them white, 1 red, 1 blue. Here there is a finite number of possible outcome distributions corresponding to the possible urn composition. If the agent wrongly believes that at most half of the balls share the same color, i.e., $p(y) \leqslant 1/2$ for $y \in \{\text{white, red, blue}\}$, the two KL minimizers are (3 white, 2 blue, 1 red) and (3 white, 1 blue, 2 red).

The following result motivates our definition of uniform BN-E. It holds regardless of the

[17] If p is a KL minimizer, i.e. $p \in \hat{\Theta}(a)$, then all observationally equivalent actions are as well, i.e. $\mathcal{E}_a(p) \subseteq \hat{\Theta}(a)$. When $\mathcal{E}_a(p)$ contains more than one element for some KL minimizer p, uniformity does not require that the equilibrium action is a best reply to every KL minimizer in $\hat{\Theta}(a)$. The only other equilibrium refinement we know of that, like uniform BN-E, tests for optimality against all beliefs in a non-singleton set is Fudenberg and He (2020), which studies non-equilibrium learning in a steady-state model where the agents are correctly specified Bayesians. They do not study the dynamics away from the steady state.

agent's discount factor, and for all optimal strategies. The same is true for all subsequent results except those where the dependence on the discount factor is made explicit.

Theorem 1. *Every limit action is a uniform BN-E.*

One implication of Theorem 1 is that limit actions must be BN-E. 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.

More strongly, Theorem 1 shows that a limit action must be a *uniform* BN-E. When a is not a uniform BN-E, there is an equivalence class of KL minimizers such that a is not a myopic best reply when beliefs concentrate on that class. The proof of Theorem 1 works by contradiction: Consider an action a which is not a uniform BN-E. If play converges to a with positive probability there must exist a history after which it is optimal in every future period to play a. We thus study the agent's belief process under the assumption that a is played in every period. As we prove in Proposition 1 in the Appendix, the agent's beliefs concentrate around the set of Kullback-Leibler minimizers relative to the realized outcome frequency exponentially fast. This result allows us to determine the agent's long run actions from the long-run frequency of outcomes. If a is not a uniform BN-E, there is a KL minimizer p' under which action a is not optimal. Moreover, the number of times each outcome is realized is a random walk, and by the Central Limit Theorem the outcome frequency converges to objective outcome frequency p^*_a at rate $1/\sqrt{t}$. This implies that the probability with which the outcome frequency will be in a ball of radius $1/\sqrt{t}$ centered around $p^*_a\left(1-1/\sqrt{t}\right)+p'_a\left(1/\sqrt{t}\right)$ in a given period t converges to a constant. These balls are chosen in the direction of the outcome frequency p'_a such that the action a is not optimal for large enough t when the empirical frequency is in these balls. We then apply the Kochen-Stone Lemma which implies that the probability that the agent's outcome frequency will be in such a ball infinitely often is non-negative and the Hewitt-Savage zero-one law implies that it must equal one. Thus with probability one, the outcome frequency will eventually be such that the agent takes an action different from a. Thus, a can not be a limit action if it is not a uniform BN-E.

The same technique can be applied to obtain a starker result in subjective bandit problems. There Corollary 1 shows that if an action performs poorly under some KL minimizer, the agent will stop playing it in finite time with probability 1, even if the action is objectively optimal and the agent is very patient. Example 6 in the Supplemental Material shows that Theorem 1 can fail without Assumption 1(ii). Here the agent's prior has countable support

and assigns vanishingly low probability to distributions that are close to one of the KL minimizers. However, Assumption 1(ii) does not ensure that a uniform BN-E exists, as shown in the following example. As a consequence, actions need not converge.

Example 1 (Non-existence of Uniform BN-E)**.** *A monopolist is uncertain about the demand for their product. Every period it posts a price in* $\{3, 4, 5, 6, 7\}$*, and then a randomly selected consumer observes the price and decides whether to buy,* $y = 1$*, or not buy,* $y = 0$*, the good. The monopolist's maximizes revenue* $u(a, y) = ay$*, and the true distribution of customer values is uniform on* $[3, 7]$*. The monopolist overestimates the variance of consumer values, and believes that they are either uniformly distributed on* $[0, 8]$ *or on* $[2, 10]$*. As we show in the Supplemental Material, the unique BN-E is non-uniform and strict, with price* 5*. Both distributions are KL-minimizing for this price, but price* 5 *is myopically optimal only if the valuations are uniformly distributed on the high range* $[2, 10]$*. Theorem 1 implies that the monopolist's actions do not converge, even though there is a unique and strict BN-E. This is because when* $a = 5$*, the monopolist eventually sees a sequence of outcomes where few consumers buy, becomes very confident in the low range of valuations* $[0, 8]$*, and switches to a lower price.*

Theorem 1 implies the non-convergence theorem of Nyarko (1991) as a corollary since also in that setting there is no uniform BN-E. Moreover, in the case of myopic agents, Corollary 2 in the Appendix combines the result with Theorem 2 of Esponda, Pouzo, and Yamamoto (2019) to show the empirical action frequencies cannot converge to some non-uniform BN-E.

4 Sufficient Conditions for Long-Run Persistence

Theorem 1 shows that play can only converge to a given action a if that action is a uniform BN-E. This section gives sufficient conditions for a to be a long-run outcome in two different senses, namely stability and attractiveness.

4.1 Stability

We say that action a is stable if play converges to a with high probability starting from every belief in a neighborhood of a KL minimizer for a. For $\nu \in \Delta(\Theta)$, let $B_\varepsilon(\nu) = \{\nu' \in \Delta(\Theta) | d(\nu', \nu) \leqslant \varepsilon\}$ be the set of beliefs over conceivable distributions that are within ε of ν. Define the set $\hat{\Theta}^\varepsilon(a)$ as all outcome distributions whose marginal distribution with respect

to action a is at most ε away from a KL minimizer,

$$\hat{\Theta}^{\varepsilon}(a) = \{p \in \Theta\colon \text{ there exists } p' \in \hat{\Theta}(a) \text{ with } ||p'_a - p_a|| \leqslant \varepsilon\}\,. \tag{2}$$

Definition 7.

(i) An action a is *stable* if for every $\kappa \in (0,1)$, there is an $\varepsilon > 0$ and a belief $\nu \in \Delta(\Theta)$ such that for all initial beliefs in $B_{\varepsilon}(\nu)$, the action prescribed by some optimal policy converges to a with probability larger than $1-\kappa$.

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

Theorem 1 shows that stable actions must be uniform BN-E. The next theorem shows that an action is a uniformly strict BN-E if and only if it is uniformly stable.

Theorem 2. *An action is uniformly stable if and only if it is a uniformly strict BN-E.*

Theorem 2 differs from past work by providing the first if and only if characterization of the stability of actions under misspecified learning with non-binary priors, and by allowing the agent to be non-myopic and thus perceive an information value from experimentation.[18] Its proof has two parts, corresponding to the two directions of the if and only if statement. To show that every uniformly strict BN-E is uniformly stable, we first show that if beliefs assign sufficiently high probability to a neighborhood of the KL minimizers, the only optimal action is the uniformly strict BN-E a. That such a neighborhood exists for a myopic policy follows from the definition of uniformly strict BN-E. 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 than its value against the most favorable minimizer, and since a is a uniformly strict BN-E this value is strictly lower than that of a. Then we combine an observation from FII with a generalization of the arguments in Fudenberg and Levine (1992) and the Dubins' upcrossing inequality to guarantee that if the probability initially assigned to the neighborhood is sufficiently high, it is unlikely to drop below the threshold that makes action a suboptimal.

[18] Bohren and Hauser (2020) and Fudenberg, Romanyuk, and Strack (2017) characterize stability when the agent has a binary prior. FII's Theorem 1 gives a sufficient condition for stability when the agent's prior has finite support. The statement of the theorem is for their general model, which takes the evolution of the belief process as a primitive, and does not describe the agent's actions, discount factor, or optimization. The paper's three applications all assume myopic choice.

The proof of the converse direction is much simpler: If a is not a uniformly strict BN-E, there is a distribution p in $\hat{\Theta}(a)$ that makes some other action b the best response, and if we set ν to be a point mass on p the agent always plays b.

Theorem 2 is in contrast to the non-convergence in the monopoly pricing example of Heidhues, Kőszegi, and Strack (2021), where there is a continuum of actions, and actions that are sufficiently near the strict best response are best responses to nearby beliefs. As we explain in Section 6, it is not clear what the right definition of uniform stability is for that setting.

Example 1 shows that Theorem 2 does not extend to strict BN-E that are not uniformly strict. The next example shows that in Theorem 2 we cannot replace uniformly stable with stable.

Example 2 (A stable BN-E that is not uniformly strict)**.** *Suppose there are 2 actions, a and b, that induce the same distribution on $Y = \{0, 1\}$ and such that $u(a, \cdot) = u(b, \cdot)$. The agent has an arbitrary belief supported on $\{p : p_a = p_b\}$, i.e., they know the actions induce the same distribution. Here, since the agent is always indifferent, even though action a is not a uniformly strict BN-E, it is stable under the optimal policy that always prescribes a.*

In general there is a gap between uniformly strict BN-E and stability, but in sufficiently rich problems, this gap is absent.

Definition 8. A problem is *rich* if for every action a, minimizer $p \in \hat{\Theta}(a)$ and $\varepsilon > 0$ there exists a $p' \in \Theta \backslash \hat{\Theta}(a)$ with $||p - p'|| \leqslant \varepsilon$ such that

$$\mathbb{E}_{p_a}\left[u(a, y)\right] - \max_{b \in A \backslash \{a\}} \mathbb{E}_{p_b}\left[u(b, y)\right] > \mathbb{E}_{p'_a}\left[u(a, y)\right] - \max_{b \in A \backslash \{a\}} \mathbb{E}_{p'_b}\left[u(b, y)\right].$$

In words, a problem is rich if for every KL minimizer for every action a, the support of the agent's prior includes a nearby distribution under which a performs relatively less well.[19] This rules out the previous example and also rules out finite-support priors.

Theorem 3. *If a problem is rich, the following are equivalent:*

(i) $a \in A$ *is a uniformly strict BN-E.*

(ii) $a \in A$ *is stable.*

Richness guarantees that if a is not a uniformly strict equilibrium, there is a KL minimizer for action a that can be approximated with a sequence of outcome distributions $(p^n)_{n \in \mathbb{N}}$ under

[19] Note that "relatively less well" allows the action to be a best response to all distributions near p.

which action a is strictly suboptimal. To prove this theorem, for every ν we build a sequence of beliefs $(\nu^n)_{n\in\mathbb{N}}$ that have have p^n has the unique KL minimizer for action a, and combine this with Theorem 1 to show that the probability that the actions converge to a starting from ν_n is 0. To summarize our stability results,

$$\textbf{Uniformly Strict BN-E} = \textbf{Uniformly Stable} \subseteq \textbf{Stable} \subseteq \textbf{Uniform BN-E},$$

where the first inclusion is an equality if the problem is rich, and the second inclusion can be strict as shown by Example 11 in the Supplemental Material.

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* initial beliefs. Another natural notion of a being a long-run outcome is that for *every* initial belief with support Θ there is strictly positive probability that the agent's action converges to a.

Definition 9. The action $a \in A$ is *positively attractive* if for every optimal policy π and every initial belief ν with $\operatorname{supp}\nu = \Theta$,

$$\mathbb{P}_\pi\left[\lim_{t\to\infty} a_t = a\right] > 0\,.$$

Below we give sufficient conditions for uniformly strict BN-E to be positively attractive. Benaïm and Hirsch (1999) obtains a similar conclusion for the linearly stable Nash equilibria of stochastic fictitious play.[20] These arguments rely on Proposition 1 in the Appendix, which shows that beliefs about the outcome distribution concentrate around the distributions that best fit the empirical frequency of outcomes. Importantly, our result applies pathwise and does not require that either actions or empirical frequencies converge.

Our results on positive attractiveness cover three different cases: subjectively exogenous outcomes, subjective bandit problems, and strongly supermodular problems. In the first two cases we are able to identify a particular empirical distribution that is sufficient for analyzing convergence. With subjectively exogenous outcomes, the agent only tracks a single empirical distribution. In subjective bandit problems, the agent does consider multiple empirical distributions, but it is sufficient to study the distribution corresponding to the action in

[20]The Bayesian foundation of fictitious play (Fudenberg and Kreps (1993)) assumes that the players believe that the environment is stationary. Away from a steady state the players are 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.

question. In supermodular problems, we instead show that certain outcome realizations can lead the agent to lock on to the highest or lowest action.

4.2.1 Subjectively Exogenous Problems

In subjectively exogenous problems, the agent believes that the distribution over outcomes is the same for all actions. This is a fairly stark assumption; more typically the agent might believe that their action influences some dimensions of the outcome but not others. We present the case where the agent believes the action has no effect at all because the extension to "partially exogenous" outcomes does not bring any additional insight.

Theorem 4. *Suppose outcomes are subjectively exogenous. If a is a uniformly strict BN-E then it is positively attractive.*

To prove Theorem we first use Proposition 1 to show that beliefs concentrate around the distributions that minimize the KL divergence from the empirical frequency on every path of outcome realizations. We then 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 BN-E, if beliefs concentrate around the minimizers, a becomes the unique best reply. While using a, the relative probability the agent assigns to distributions in $\hat{\Theta}(a)$ 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 $\hat{\Theta}(a)$ to show that, with positive probability, the agent will stick to action a forever.

Proposition 4 in EPY shows that for every uniformly strict BN-E a, there exists at least one prior with support equal to Θ under which the policy converges to a with positive probability. FII provides sufficient conditions for the system to converge with probability 1 to a specific BN-E from any initial belief. Our Theorem 4 concludes that every uniformly strict BN-E has positive probability of being the limit behavior starting from *every* initial prior without imposing conditions that imply global convergence to a specific outcome.

Example 3 (Stackelberg game perceived as Cournot)**.** *The agent is a seller who every period faces a competing seller randomly drawn from a large population. The agent first chooses whether to produce low output, $a = 1$, or high output, $a = 2$. The competitor sets their quantity y at 1 or 2 after observing the agent's action: If the agent chooses low output the competitor produces high output with probability $2/3$, while if the agent chooses high output the competitor produces a high quantity with probability $1/3$.*[21] *The agent believes that the*

[21] The randomness could arise from a distribution over production costs in the population of competitors.

competitor chooses output without observing the agent's action, and that they choose an high output with some unknown probability p: $\Theta = \{p \in \Delta(Y)^A : p_2(2) = p_1(2)\}$. *The true distribution is* $p_2^*(2) = 1/3 = p_1^*(1)$.

The demand function of the consumers is linear, and the agent has no production cost; the utility function of the agent is $u(a, y) = a(4.5 - a - y)$. *High output is objectively optimal for the agent, and is also a uniformly strict BN-E. However, low output is also a uniformly strict BN-E, supported by the wrong belief that the observed high level of production of the competitor would be the same even if the agent increased output. By Theorem 4 both actions have a positive probability of arising as limit outcomes starting from every initial prior.*

Without the assumption of subjectively exogenous outcomes, uniformly strict BN-E need not be positively attractive.

Example 4 (A uniformly strict BN-E that is not positively attractive)**.** *A central bank decides whether to keep a flexible exchange rate,* $a = f$, *or peg the currency to the dollar,* $a = c$. *The outcome has two binary components,* $y = (y^e, y^s)$, *where* y^e *says whether the economy is in a boom, and* y^s *whether there is a speculative attack on the currency. The bank likes booms and dislikes speculative attacks:* $u(f, y) = y^e, u(c, y) = \frac{3}{2}y^e - y^s$. *The bank correctly believes that whether there is a speculative attack is independent of the state of the economy. Furthermore, the bank knows that if they maintain a flexible exchange rate, the probability of a currency attack is* 0, *and believes that the probability of a currency attack under a fixed exchange rate is either* 10% *(the true value) or* 90%. *The bank correctly believes that pegging the currency to the dollar increases the probability of a boom by* $33.\bar{3}\%$ *over a baseline probability, which the bank believes is either* $33.\bar{3}\%$ *or* $66.\bar{6}\%$, *and the belief is independent across the two dimensions. In truth the baseline is* 50%, *so the bank is misspecified.*[22]

Here pegging the currency to the dollar is a uniformly strict BN-E, but it is not positively attractive: For any discount factor, if the prior assigns sufficiently high probability to the states where a currency attack happens with probability 90% *if the currency is not pegged to the dollar, the bank starts out choosing a flexible exchange rate, and sticks with that action forever. To see why, note that when the currency is floating the bank does not update its beliefs about the likelihood of a currency attack under a pegged exchange rate.*

[22] That is, the bank believes that the probabilities of a boom with or without peg are either $(100\%, 66.\bar{6}\%)$ or $(66.\bar{6}\%, 33.\bar{3}\%)$, respectively, while in truth they are $(83.\bar{3}\%, 50\%)$.

4.2.2 Subjective Bandit Problems

Recall that in a subjective bandit problem (Definition 3), the agent believes that the outcome distribution is independent across actions. An argument similar to that for subjectively exogenous problems shows that uniformly strict BN-E are positively attractive in subjective bandit problems if the agent is sufficiently patient. However, uniformly strict BN-E is a very demanding concept in subjective bandit problems, as the Kullback-Leibler divergence between the true and subjective outcome distributions induced by an action 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, in these problems we can replace the uniformity requirement with the requirement that the equilibrium is *weakly identified* introduced in Esponda and Pouzo (2016).

Definition 10. A BN-E action a is *weakly identified* 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. When the agent thinks the outcome distribution is exogenous, the equilibrium can only be weakly identified if 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.

Theorem 5. *For every subjective bandit problem there is a $\bar{\beta} < 1$ such that if the discount factor $\beta \geqslant \bar{\beta}$, then every weakly identified strict BN-E is positively attractive.*

The proof uses the fact that patient agents experiment with actions that they believe might give them a higher payoff. The conclusion of the theorem is false for myopic agents even in the correctly specified case, where the BN-E correspond to the self-confirming equilibria, and with probability 1 the agent may always play whichever action is myopically optimal given their initial beliefs.

In subjective bandit problems, we can sharpen the conclusion of Theorem 1 for actions that perform poorly under one of the KL minimizers. We say that action a is *quasi-dominated* if there are $\hat{p} \in \hat{\Theta}(a)$ and $b \in A$ such that $\mathbb{E}_{\hat{p}_a}[u(a, y)] < \mathbb{E}_{p_b}[u(b, y)]$ for all $p \in \Theta$. That is, there is a KL minimizer $\hat{p}$ for action a such that the utility of a under $\hat{p}$ is lower than that of action b under any of the p in the support of the prior. Quasi-dominated actions are not uniform BN-E, so play cannot converge to them with positive probability.

In a subjective bandit problem even more is true; quasi-dominated actions can be played only a finite number of times.

Corollary 1. *In a subjective bandit problem, any quasi-dominated action is almost surely played only a finite number of times.*

In particular, in two-armed subjective bandit problems where one action is quasi-dominated, play converges to the other one. Note that this result does not depend on the discount factor, and is true even if the quasi-dominated action is objectively optimal and the agent assigns positive probability to it being optimal. In contrast, the probability that a correctly specified agent locks on to an incorrect action goes to 0 as the discount factor goes to 1.

4.2.3 Strongly Supermodular Problems

Definition 11. We say that the problem is *strongly supermodular* if we can strictly order the space of actions $(A, >)$, outcomes $(Y, >)$, and the set of conceivable distributions $(\Theta, >)$ so that:

(i) u is strictly supermodular in a and y;

(ii) if $p, p' \in \Theta$ and $p > p'$, then for all $a \in A$ and $y \in Y \backslash \bar{y}$, we have $p_a(\{y' : y' > y\}) > p'_a(\{y' : y' > y\})$, where $\bar{y}$ denotes the highest outcome.

Theorem 6. *In a strongly supermodular problem, if $p^*_{\underline{a}}$ (resp. $p^*_{\bar{a}}$) has full support, and the highest action $\bar{a}$ (resp. the lowest action $\underline{a}$) is a uniform and strict BN-E, then $\bar{a}$ (resp. $\underline{a}$) is positively attractive.*

Strong supermodularity implies that the agent will use action $\bar{a}$ if they observe the highest y's sufficiently often. Moreover, the antisymmetric ordering of the elements of Θ guarantees that every uniform and strict BN-E is uniformly strict, and so Theorem 2 guarantees that there is positive probability that once the agent plays $\bar{a}$ they will stick to it forever.

Example 5 (Under-investment trap)**.** *Each period the agent decides how much effort $a \in \{0, 1, 2\}$ to exert on a task. The effort can be either successful, $y = 1$, or unsuccessful, $y = 0$. Higher effort makes success more likely: $p^*_2(1) = 9/10 > p^*_1(1) = 1/2 > p^*_0(1) = 1/12$. Moreover, higher effort also increases the benefit of a success: $u(a, y) = ay - a/2$. Thus the objectively optimal action is to exert high effort, $a = 2$.*

The agent mistakenly believes that the probability of success depends on their effort and their intrinsic skill ψ, and Θ is consists of all p such that $p_2(1) = 2/3 + \psi > p_1(1) = 1/2 + \psi > p_0(1) = 1/3 + \psi$ for some $\psi \in [-1/4, 1/4]$.

Here there are two BN-E: $a = 0$ and $a = 2$. In the bad equilibrium $a = 0$, the KL-minimizing outcome distribution corresponds to the lowest possible skill level $\psi = -1/4$, which leads the agent to exert the low effort. Since both BN-E are uniformly strict and the problem is strongly supermodular, Theorem 6 implies that both the Nash equilibrium and the bad equilibrium with low effort are positively attractive.

5 Signals

Suppose each period before taking an action the agent observes a signal s from a compact set S, equipped with its Borel sigma algebra. Thus the analog of an action in the previous sections is now a *strategy*, i.e. a measurable map $\sigma : S \to A$ from signals to actions. Signals may be payoff relevant, so now utility is a map $u : A \times Y \times S \to \mathbb{R}$, and signals may also be useful for predicting the outcome distributions, so now $p_{a,s} \in \Delta(Y)$ depends both on this period's action and on the signal observed at the start of the period. A policy $\pi(a^t, y^t, s^{t+1})$ specifies the action in each period t as a function of past actions, outcomes and signals.

To complete the model we also need to specify the objective distribution of signals. We focus on the case where the distribution of s is fixed (iid) with distribution ζ that is known to the agent, as in Esponda and Pouzo (2016).[23]

Subjective Beliefs 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 $P = \Delta(Y)^{A \times S}$ be the space of all signal and action dependent outcome distributions. The agent's uncertainty is captured by a prior belief $\mu_0 \in \Delta\big(P\big)$, again with $\Theta = \operatorname{supp} \mu_0$.

Assumption 1′.

(i) For all $p \in \Theta$, $a \in A$, $y \in Y$, and $s \in S$, $p_{a,s}(y) > 0$ if and only if $p^*_{a,s}(y) > 0$.

(ii) The prior μ_0 has *subexponential decay*: there is $\Psi : \mathbb{R}_+ \to \mathbb{R}_{++}$ such that for every $p \in \Theta$ and $\varepsilon > 0$ we have $\mu_0(B_\varepsilon(p)) \geqslant \Psi(\varepsilon)$ with $\lim \Psi(K/n) \exp(n) = \infty$ for all $K > 0$.

Let $\mu_t(\cdot \mid (s^t, a^t, y^t)) \in \Delta(P)$ denote 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.

We say that two outcome distributions $p, p' \in \Theta$ are *observationally equivalent under the strategy* σ if $p_{\sigma(s),s}(y) = p'_{\sigma(s),s}(y)$ for all $s \in S$ and $y \in \operatorname{supp} p^*_{\sigma(s),s}$, and we let $\mathcal{E}_\sigma(p)$ denote

[23] A continuum of signals allows payoff shocks that generate continuous best-response distributions.

the outcome distributions that are observationally equivalent to p under σ. To simplify the analysis, we make the following assumption, which is satisfied for example if the signals are payoff shocks, or if there are only finitely many signals.

Definition 12. The environment is *finite dimensional* if there is a partition $\Xi = \{\xi_1, ...\xi_N\}$ of S into a finite number of measurable sets such that the agent correctly believes the same outcome distribution applies for all s in ξ_i: $p_{a,s} = p_{a,s'}$ for all $p \in \Theta \cup \{p^*\}$, $a \in A$, $i \in \{1, ..., N\}$, and $s, s' \in \xi_i$.

Under this assumption, we abuse the notation by letting p_{a,ξ_i} denote the outcome distribution prescribed by p after action a and an arbitrary signal in ξ_i. With this, the relevant set of "closest beliefs to the truth" is now

$$\hat{\Theta}(\sigma) = \underset{p \in \Theta}{\operatorname{argmin}} \sum_{\xi_i \in \Xi} \zeta(\xi_i) H\left(p^*_{\sigma(s),\xi_i}, p_{\sigma(s),\xi_i}\right).$$

We use this modified definition of the minimizers to extend the definition of the equilibrium concepts to this more general setting. The proofs for all of the results of this section are in the Supplemental Material.

Definition 6′.

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

(ii) Strategy σ is a uniform BN-E *if for all $p \in \hat{\Theta}(\sigma)$ there exists a belief $\nu \in \Delta\left(\mathcal{E}_\sigma(p)\right)$ such that σ is myopically optimal given ν.*

(iii) Strategy σ is a uniformly strict BN-E *if σ is the unique myopic best reply to any belief in $\nu \in \Delta(\hat{\Theta}(\sigma))$.*[24]

Theorem 1′. *Suppose the agent's beliefs are finite dimensional. If σ is a limit strategy, then σ is a uniform BN-E.*

The proof of this result is very similar to the proof of Theorem 1. The main difference is that the relevant random walk is the empirical distribution over joint realizations of signals and outcomes.

Similarly, we can extend our result on the stability of uniformly strict BN-E. Specifically:

[24]Here uniqueness is up to a set of signals that have zero probability under ζ.

Theorem 2′. *Suppose σ is a uniformly strict BN-E. Then there is a belief $\nu \in \Delta(\Theta)$ such that for every $\kappa \in (0,1)$ there exists an $\varepsilon' > 0$ such that starting from any prior belief in $B_{\varepsilon'}(\nu)$:*

$$\mathbb{P}_\pi \left[\lim_{t\to\infty} \frac{1}{t+1} \sum_{r=0}^{t} \mathbf{1}_{\pi(a^r,y^r,s^{r+1})=\sigma(s_{r+1})} \geqslant 1-\kappa \right] > 1-\kappa\,.$$

Example 10 in the Supplemental Material illustrates the long-run biases that can be induced when the agent mistakenly thinks that signals are uninformative. There, a seller receives a signal about the current period's market, and decides whether to undertake an investment that may boost sales. The seller does not realize that when more consumers show up, a lower fraction of them buy; we show that this can lead to persistent underinvestment when market attendance is high.

When the agent thinks the signals are uninformative, their prior has support on distributions of y given a that are independent of s. Here the only reason they might influence the agent's choices is that they may directly enter their payoff function. The next result shows that all uniformly strict BN-E are positively attractive when signals are subjectively uninformative and the true data generating process has full support.

Theorem 4′. *If signals are finite, subjectively uninformative, outcomes are subjectively exogenous, and that the true outcome distribution p^* has full support, then any uniformly strict BN equilibrium σ is positively attractive.*

The proof of this result is similar to that of Theorem 4, because when signals are subjectively uninformative we can apply Proposition 1 to the *uncontingent* empirical distribution.

6 Concluding Remarks

Learning in Large Population Games The biases we consider are relevant in non-equilibrium models of learning about the prevailing distribution of strategies. Consider a finite I player game, and suppose there is a continuum of agents in each player role $i \in I$ who are matched every period to play the game, and observe the actions played in their matches but nothing else. In a steady state,[25] the problem faced by an agent in population i is equivalent to the one we 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

[25] These models do have steady states when there is a steady outflow of agents balanced by an inflow of new ones; see e.g. Proposition 3 in Fudenberg and He (2018).

period's distribution of opponents' strategies. Causation neglect corresponds to the bias of an agent who thinks they are playing a simultaneous-move game, when in reality their 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 may have signal neglect, and incorrectly believe that the game has independent private values.

Our results help characterize the possible limit actions in these situations. Of course, extensive-form games may not have strict equilibria, so some of our results will not apply, but it may be possible to extend some of our 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.[26]

Infinitely Many 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 BN-E is an action a that is the unique best response to every belief in a neighborhood of the 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 BN-E are not equivalent. Indeed, as shown by an example in Heidhues, Kőszegi, and Strack (2021), some BN-E that are uniformly strict BN in the sense of Definition 6 may not be positively attractive. However, we conjecture that the positive attractiveness result continues to hold under the alternative definition.

Summary and Discussion 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 these 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 BN-E are stable, and that under a mild condition only uniform BN-E can be stable. Moreover we show that play can only converge to uniform BN-E. Our work thus suggests uniformity should be imposed as a refinement of BN-E. We then provide the first sufficient conditions for an action to be positively attractive

[26] Alternatively we could consider a model with one agent per player role and payoff perturbations, as in Fudenberg and Kreps (1993) and Esponda and Pouzo (2016).

under misspecified learning. Here we highlight the role played by the correlation that the agent perceives between the outcome distributions associated with different actions.

A Appendix

Section A.1 formally describes the space where our stochastic processes are defined, Section A.2 states some preliminary technical lemmas, Section A.3 proves that beliefs concentrate around the KL minimizers at and exponential rate, and Section A.4 contains the results of the main text for the models that do not have signals.

A.1 Sample Space

We work with the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The sample space $\Omega = (Y^\infty)^A$ consists of infinite sequences of action dependent outcome realizations $(x_{a,1}, x_{a,2}, \ldots)_{a\in A}$, where $x_{a,k}$ 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^*. The outcome observed by the agent in period t after action a_t is $y_t = x_{a_t,k}$, where $k = |\{\tau \leqslant t\colon a_\tau = a_t\}|$ is the number of times the agent has taken action a_t up to and including period t.[27] The probability measure $\mathbb{P}_\pi$ over $(a_\tau, y_\tau)_{\tau=1}^\infty$ induced by the policy π is defined as follows: For every $t \in \mathbb{N}$ and cylinder $(a_\tau, y_\tau)_{\tau=1}^t$,

$$\mathbb{P}_\pi\left[(a_\tau, y_\tau)_{\tau=1}^t\right] = \begin{cases} 0, & \text{if there exists } t' \in \{1, ..., t\} : a_{t'} \neq \pi((a_\tau, y_\tau)_{\tau=1}^{t'-1}) \\ \prod_{\tau=1}^t p_{a_\tau}(y_\tau) & \text{otherwise.} \end{cases}$$

A.2 Preliminary Lemmas and Definitions

Denote the set of conceivable outcome distributions for action a that best match p_a^* by

$$\hat{\Theta}_a(a) = \underset{p_a : p\in\Theta}{\operatorname{argmin}} H\left(p_a^*, p_a\right) \subset \Delta(Y).$$

Lemma 1. *For every $a \in A$ and $\varepsilon > 0$, $\hat{\Theta}(a)$ defined in equation (1), $\hat{\Theta}_a(a)$, $\hat{\Theta}^\varepsilon(a)$ defined in equation (2), and $\Delta(\hat{\Theta}(a))$ are compact.*

[27] An alternative specification has sample space $(x_{a,1}, x_{a,2}, \ldots)_{a\in A}$, with $x_{a,k}$ denoting the outcome realization if the agent takes action a in period k. An argument similar to that of Lemma 5 of Fudenberg and He (2017) shows that this would not change our results.

The proof of Lemma 1 is routine and relegated to the Supplemental Material.

For every $p \in P$ and every policy π let $\mathbb{E}_{p,\pi}$ denote the expectation operator over action and outcome sequences that is induced by policy π under outcome distribution p. We work with the agent's normalized value throughout, which is

$$V(\pi, \nu) = (1-\beta) \int_P \mathbb{E}_{p,\pi} \left[\sum_{t=1}^{\infty} \left[\beta^{t-1} u(a_t, y_t) \right] \right] d\nu(p).$$

The set of policy functions is $\Pi = A^{\bigcup_{t=0}^{\infty} A^t \times Y^t}$.

Lemma 2. *Π is compact in the product topology, and for all $\nu \in \Delta(\Theta)$, $V(\cdot, \nu)$ is continuous with respect to the product topology.*

Lemma 2 is a consequence of the more general Lemma 11 which covers cases where each period the agent observes a signal before choosing their action. This lemma is proved in the Supplemental Material.

Next we 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 $\hat{\Theta}(a)$. Denote the set of beliefs over conceivable distributions that assign at least probability $1-\varepsilon$ to $\hat{\Theta}^{\varepsilon}(a)$ by

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

The following lemma shows that if the agent's beliefs are sufficiently concentrated on the set of KL minimizers associated with a uniformly strict BN-E a, the agent will play a, even if the agent is not myopic.

Lemma 3. *If $a \in A$ is a uniformly strict BN-E, then for every optimal policy π, there exists an $\hat{\varepsilon} > 0$ such that for all $\varepsilon < \hat{\varepsilon}$, $\nu \in M_{\varepsilon,a} \implies \pi(\nu) = a$.*

Proof. Let π^a denote the policy that prescribes to always play a. By Lemma 2, 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, and because $\beta \in [0,1)$, the value function is continuous at infinity, so $V(\pi^a, \nu) - V(\cdot, \nu)$ is a continuous function of the policy. Moreover since $\mathbb{E}_{p,\pi}\left[\sum_{t=1}^{\infty} \left[\beta^{t-1} u(a_t, y_t)\right]\right]$ is continuous in p, $V(\pi^a, \cdot) - V(\tilde{\pi}, \cdot)$ is continuous in ν.

Define $G(\varepsilon)$ as the minimal gain from playing a forever instead of using some best policy $\tilde{\pi}$ that does not play a at a belief ν in $M_{\varepsilon,a}$: $G(\varepsilon) = \min_{\tilde{\pi}:\tilde{\pi}(\nu)\neq a} \min_{\nu \in M_{\varepsilon,a}} \left(V(\pi^a, \nu) - V(\tilde{\pi}, \nu)\right)$.

Given that $\varepsilon \to M_{\varepsilon,a}$ is an upper hemicontinuous and compact valued correspondence, G is continuous in ε.

Note that $M_{0,a} = \hat{\Theta}(a)$ which is the set of KL minimizers given action a. For every belief supported on this set a is the unique maximizer as it is a uniformly strict BN-E. If the prior belief is supported on $\hat{\Theta}(a)$ so is the posterior after every history, which implies that the action a is strictly optimal after every history. Consequently, $V(\pi^a, \nu) - V(\tilde{\pi}, \nu) > 0$ for every $\nu \in \Delta(\hat{\Theta}(a))$ and every strategy $\tilde{\pi}$ that does not prescribe action a in every period with probability 1. Since $G(0) = \min_{\tilde{\pi}:\tilde{\pi}(\nu)\neq a} \min_{\nu\in\Delta(\hat{\Theta}(a))} (V(\pi^a, \nu) - V(\tilde{\pi}, \nu))$, this implies that $G(0) > 0$. By the continuity of G there is an $\hat{\varepsilon}$ such that if $\varepsilon \leqslant \hat{\varepsilon}$, $G(\varepsilon) > 0$. This implies that for any $\varepsilon \leqslant \hat{\varepsilon}$, any optimal policy prescribes the action a for all $\nu \in M_{\varepsilon,a}$. ■

The next Lemma extends an argument of Fudenberg and Levine (1992) to take into account misspecification. It establishes that if the expectation of the l-th power of the likelihood ratio between two subjective outcome distributions is greater 1 then the l-th power of the likelihood ratio of the subjective probability assigned to small environments of these outcome distributions is a sub-martingale.

Lemma 4. *Let $p, p', p^* \in \Delta(Y)$, and $l \in (0,1)$ be such that*

$$\sum_{y\in Y} p^*(y) \left(\frac{p(y)}{p'(y)}\right)^l < 1. \tag{3}$$

Then there is $\varepsilon' > 0$ such that for all $\nu \in \Delta(\Delta(Y))$, if we let $\nu(C \mid y) = \frac{\int_{q\in C} q(y)d\nu(q)}{\int_{q\in\Delta(Y)} q(y)d\nu(q)}$, then

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

Proof. The lemma is trivially true if $\nu(B_\varepsilon(p')) = 0$ for some ε. Therefore, without loss of generality, we can assume that $\nu(B_\varepsilon(p')) > 0$ for all ε. Let $\hat{\varepsilon}$ be such that $||q - p'|| \leqslant \hat{\varepsilon}$ implies that $q(y) = 0$ only if $p'(y) = 0$. Let $C_\varepsilon = \Delta(B_\varepsilon(p)) \times \Delta(B_\varepsilon(p'))$ and define $G : [0, \frac{\hat{\varepsilon}}{2}] \to \mathbb{R}$ by

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

By the Maximum Theorem, the compactness of $\Delta(B_\varepsilon(p'))$ and $\Delta(B_\varepsilon(p))$ and the fact that

$G(0) < 1$ by equation (3), there is $\varepsilon' > 0$ such that for all $\nu' \in \Delta\left(B_{\varepsilon'}\left(p'\right)\right)$, $\bar{\nu} \in \Delta\left(B_{\varepsilon'}\left(p\right)\right)$

$$\sum_{y \in Y} p^*(y) \left(\frac{\int_{B_{\varepsilon'}(p)} \bar{q}(y) d\bar{\nu}\left(\bar{q}\right)}{\int_{B_{\varepsilon'}(p')} q(y) d\nu'\left(q\right)} \right)^l \leqslant 1. \tag{4}$$

Then

$$\begin{aligned}
\sum_{y \in Y} p^*(y) \left(\frac{\nu(B_{\varepsilon'}\left(p\right) \mid y)}{\nu(B_{\varepsilon'}\left(p'\right) \mid y)} \right)^l &= \sum_{y \in Y} p^*(y) \left(\frac{\int_{B_{\varepsilon'}(p)} \nu(B_{\varepsilon'}\left(p\right)) \bar{q}(y) d\frac{\nu(\bar{q})}{\nu(B_{\varepsilon'}(p))}}{\int_{B_{\varepsilon'}(p')} \nu\left(B_{\varepsilon'}\left(p'\right)\right) q(y) d\frac{\nu(q)}{\nu(B_{\varepsilon'}(p'))}} \right)^l \\
&= \sum_{y \in Y} p^*(y) \left(\frac{\int_{B_{\varepsilon'}(p)} \bar{q}(y) d\frac{\nu(\bar{q})}{\nu(B_{\varepsilon'}(p))}}{\int_{B_{\varepsilon'}(p')} q(y) d\frac{\nu(q)}{\nu(B_{\varepsilon'}(p'))}} \right)^l \left(\frac{\nu(B_{\varepsilon'}\left(p\right))}{\nu\left(B_{\varepsilon'}\left(p'\right)\right)} \right)^l \\
&\leqslant \left(\frac{\nu(B_{\varepsilon'}\left(p\right))}{\nu\left(B_{\varepsilon'}\left(p'\right)\right)} \right)^l
\end{aligned}$$

where the inequality follows from equation (4). ■

The next lemma shows that if for every initial belief supported on Θ, always playing b almost surely leads to a belief at which action b is not prescribed by any optimal policy, then b is not a limit action.

Lemma 5. *Suppose that for any prior belief ν_0 supported on Θ and any optimal policy $\tilde{\pi}$ $\mathbb{P}_{\pi^b}[b = \tilde{\pi}(\nu_\tau)$ for all $\tau \geqslant 0] = 0$, then b is not a limit action.*

Proof. Suppose by way of contradiction that there is an optimal policy $\tilde{\pi}$ and a history (a^t, y^t) with $\mathbb{P}_{\tilde{\pi}}[(a^t, y^t)] > 0$ such that with positive probability $\tilde{\pi}$ prescribes b after (a^t, y^t) in every future period. Define $\nu_0 = \mu(\cdot|(a^t, y^t))$, and notice that $\operatorname{supp} \nu_0 = \operatorname{supp} \mu_0 = \Theta$. Define ν_t to be the belief if the agent uses the policy π^b, i.e. plays b in every period. As the evolution of beliefs under π^b is the same as under $\tilde{\pi}$ for every history where the agent continues to play b, we have that $\mathbb{P}_{\tilde{\pi}}[b = \tilde{\pi}(\mu_\tau)$ for all $\tau \geqslant t] > 0$ if and only if $\mathbb{P}_{\pi^b}[b = \tilde{\pi}(\nu_\tau)$ for all $\tau \geqslant 0] > 0$. However, the later equals zero by the assumption of the lemma, which establishes that b can not be a limit action. ■

The next lemma extends Lemma 3 of FII to show that there exists a uniform l such that all KL minimizers dominate all the distributions that are ε away from the minimizers in the sense that the expectation of the l-th power of the likelihood ratio is lower than 1.

Lemma 6. *Fix an action a and $\varepsilon > 0$. There exists $\bar{l} > 0$ such that for all $l \leqslant \bar{l}$, for every KL minimizer $q \in \hat{\Theta}(a)$, and every outcome distribution $p' \notin \hat{\Theta}^{\varepsilon}(a)$*

$$f_l(q, p') := \sum_{y \in Y} p_a^*(y) \left(\frac{p'_a(y)}{q_a(y)} \right)^l < 1 .$$

Proof. As noted by FII in their Lemma 3, (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 $f_l(q, p') < 1$ for all $l \leqslant l(q, p')$ and (ii) for all $q, q' \in \Theta$, if $\hat{l} > l$ and $f_l(q, q') \geqslant 1$, then $f_{\hat{l}}(q, q') \geqslant 1$. We will now prove that there exists a uniform l that works for every $q \in \hat{\Theta}(a)$ and $p' \notin \hat{\Theta}^{\varepsilon}(a)$.

Suppose by way of contradiction that there was no $\bar{l} > 0$ such that for all $l \leqslant \bar{l}$, $f_l(q, p') < 1$ for all $q \in \hat{\Theta}(a)$ and $p' \notin \hat{\Theta}^{\varepsilon}(a)$. Then define a sequence $(q_n, p'_n)_{n \in \mathbb{N}} \in (\hat{\Theta}(a), \Theta \backslash \hat{\Theta}^{\varepsilon}(a))^{\mathbb{N}}$ such that $f_{\frac{1}{n}}(q_n, p'_n) \geqslant 1$. Sequential compactness of $\hat{\Theta}(a) \times \text{cl}\{p \in \Delta(\Theta) \colon p_a \notin \hat{\Theta}^{\varepsilon}(a))\}$ guarantees that this sequence has an accumulation point (q, p') with $q \in \hat{\Theta}(a)$ and $p' \notin \hat{\Theta}(a)$.[28] However, for $n > \frac{1}{l(q,p')}$, $f_{\frac{1}{n}}(q_n, p'_n) \geqslant 1$ implies $f_{l(q,p')}(q_n, p'_n) \geqslant 1$, and the continuity of $f_{l(q,p')}$ at (q, p') leads to a contradiction with $f_{l(q,p')}(q, p') < 1$. ∎

A.3 Exponential Concentration of Beliefs

We show next that repeated use of action a implies that the beliefs about the outcome distribution induced by a concentrate at an exponential rate around the distributions that "best fit" the empirical frequency of observed outcomes. Importantly, this result does not require that either actions or empirical frequencies converge. It will be important in what follows that these results apply pathwise, as they do in the correctly specified case studied by Diaconis and Freedman (1990), although unlike their result ours only applies for empirical distributions that are near the true distribution p^*. For brevity, we limit our analysis to this set of distribution, since this is enough for our results. In a separate note, Fudenberg, Lanzani, and Strack (2021), we provide a result that resembles more closely the original result in Diaconis and Freedman (1990).

For every $a \in A$, $\eta \in (0, 1)$ and $q \in \Delta(Y)$, let $q_\eta = (1 - \eta)p_a^* + \eta q$, $\eta_t = 2t^{-\frac{1}{2}}$, and $D = \min \{(p'_a(y)/p_a(y)) : p, p' \in \Theta, a \in A, y \in Y, p_a^*(y) > 0\}$.

Proposition 1. *Let $(a_i, y_i)_{i=1}^{\tau}$ be a history with positive probability, and suppose that only action a is played in periods $(\tau + 1, ..., \tau + t)$. For every $\hat{q} \in \hat{\Theta}_a(a)$, there exist $I, \hat{K}, K' \in \mathbb{R}_{++}$*

[28] We denote the closure of a set by cl.

such that if the empirical outcome frequency $f_t = \frac{1}{t}\sum_{i=\tau+1}^{\tau+t} \mathbf{1}_{y_i=y}$ *satisfies* $||\hat{q}_{\eta_t} - f_t|| < ||\hat{q} - p_a^*||t^{-\frac{1}{2}}/K'$, *then*

$$\frac{\mu_{\tau+t}\left(\{p \in \Theta\colon \forall y \in \operatorname{supp} p_a^*, |p_a(y) - \hat{q}(y)| < \varepsilon\}\right)}{1 - \mu_{\tau+t}\left(\{p \in \Theta\colon \forall y \in \operatorname{supp} p_a^*, |p_a(y) - \hat{q}(y)| < \varepsilon\}\right)} \geqslant D^{\tau}\Psi\left(\hat{K}\varepsilon^2\frac{2}{It^{\frac{1}{2}}}\right)\exp\left(2\hat{K}t^{\frac{1}{2}}\varepsilon^2\right).$$

To establish Proposition 1 we first prove a sequence of auxiliary results. Given two outcome distributions $q, q' \in \Delta(Y)$, $\eta \in (0,1)$, and $\varepsilon > 0$, let

$$U_\varepsilon(q, q', \eta) = \{q'' \in \Delta(Y) : ||\eta q + (1-\eta)q' - q''|| \leqslant \varepsilon\}$$

denote the ball of radius ε around $\eta q + (1-\eta)q'$. The next result establishes a form of local Lipschitz continuity of the function $\min_{q' \in C} H(\cdot, q') - H(\cdot, q)$ for suitably chosen $q \in \Delta(Y)$ and compact $C \subseteq \Delta(Y)$.

Lemma 7. *Fix* $q \in \Delta(Y)$ *with* $\operatorname{supp} q \subseteq \operatorname{supp} p_a^*$ *and a compact set* $C \subseteq \Delta(Y)$ *such that all the elements of* C *are absolutely continuous with respect to* p_a^*. *Then there exists a* $K > 0$ *such that for every* $f' \in U_\varepsilon(q, p_a^*, \eta)$ *with* $\operatorname{supp} f' \subseteq \operatorname{supp} p_a^*$

$$|\min_{q' \in C} H\left((1-\eta)p_a^* + \eta q, q'\right) - H\left((1-\eta)p_a^* + \eta q, q\right) - \min_{q' \in C} H\left(f', q'\right) + H\left(f', q\right)| \leqslant K\varepsilon\,.$$

The proof of Lemma 7 is in the Supplemental Material.

Let χ be a Borel probability measure over probability distributions on Y, let

$$Q_{\varepsilon,\chi}(\bar{q}) = \left\{q' \in \Delta(Y) : \exists q'' \in \Delta(Y) : H\left(\bar{q}, q''\right) \leqslant \min_{q \in \operatorname{supp}\chi} H\left(\bar{q}, q\right), ||q' - q''||_\infty < \varepsilon\right\}$$

be the distributions that are within ε of a distribution q'' with a lower Kullback-Leibler divergence with the given $\bar{q}$ than the minimum over $\operatorname{supp}\chi$, and let

$$g\left(p', \varepsilon\right) = \min_{p \in \Delta(Y)\setminus Q_{\varepsilon,\chi}} H\left(p', p\right) - \min_{p \in \operatorname{supp}\chi} H\left(p', p\right) > 0$$

be the minimal increase of the relative entropy from p' when it is minimized over $\Delta(Y)\setminus Q_{\varepsilon,\chi}$ instead of $\operatorname{supp}\chi$.

Lemma 8. *Let* χ_0 *be a Borel probability measure over* $\Delta(Y)$ *and for every* $t \in \mathbb{N}$ *and every sequence of outcomes* $y^t \in Y^t$ *let* $\chi_t(\cdot|y^t)$ *denote the posterior belief after observing the outcome sequence* y^t *starting from the prior* χ_0. *Then for all* $\varepsilon \in \mathbb{R}_{++}$ *and* $f_t(y) =$

$\frac{1}{t}\sum_{\tau=1}^{t}\mathbf{1}_{y_\tau=y}$ *we have that*

$$\frac{\chi_t\left(Q_{\varepsilon,\chi_0}\left(f_t\right)\mid y^t\right)}{1-\chi_t\left(Q_{\varepsilon,\chi_0}\left(f_t\right)\mid y^t\right)}\geqslant\chi_0\left(Q_{\frac{g(f_t,\varepsilon)}{2R(f_t,\varepsilon)},\chi_0}\left(f_t\right)\right)e^{.5tg(f_t,\varepsilon)}$$

where

$$R(f_t,\varepsilon)=\sup_{q,q'\in Q_{\varepsilon,\chi_0}(f_t)}\frac{|H(f_t,q)-H(f_t,q')|}{||q-q'||}.$$

Proof. Fix $\varepsilon\in\mathbb{R}_{++}$ and for any $\bar{\varepsilon}\in\mathbb{R}_{++}$, let $Q(\bar{\varepsilon})=Q_{\bar{\varepsilon},\chi_0}\left(p'\right)$. By definition of $R(f_t,\varepsilon)$, $g(f_t,\varepsilon)/2R(f_t,\varepsilon)\leqslant\varepsilon$, and so $\min_{p\in\Delta(Y)\setminus Q(\varepsilon)}H\left(p',p\right)-\max_{p\in Q\left(\frac{g(f_t,\varepsilon)}{2R(f_t,\varepsilon)}\right)}H\left(p',p\right)\geqslant.5g\left(p',\varepsilon\right)$. From the definition of χ_t we have that for all y^t where the empirical distribution is f_t,

$$\begin{aligned}\frac{\chi_t\left(Q(\varepsilon)\mid y^t\right)}{1-\chi_t\left(Q(\varepsilon)\mid y^t\right)} &= \frac{\int_{Q(\varepsilon)}\prod_{y\in Y}q(y)^{tf_t(y)}d\chi_0(q)}{\int_{\operatorname{supp}\chi_0\setminus Q(\varepsilon)}\prod_{y\in Y}q(y)^{tf_t(y)}d\chi_0(q)}\geqslant\frac{\int_{Q\left(\frac{g(f_t,\varepsilon)}{2R(f_t,\varepsilon)}\right)}\exp(-tH\left(f_t,q\right))d\chi_0(q)}{\exp(-t\min_{p\notin Q(\varepsilon)}H\left(p',p\right))}\\ &= \int_{Q\left(\frac{g(f_t,\varepsilon)}{2R(f_t,\varepsilon)}\right)}\exp(t\min_{p\notin Q(\varepsilon)}H\left(p',p\right)-tH\left(f_t,q\right))d\chi_0(q)\\ &\geqslant \chi_0\left(Q\left(\frac{g(f_t,\varepsilon)}{2R(f_t,\varepsilon)}\right)\right)e^{.5tg(p',\varepsilon)},\end{aligned}$$

where the first inequality follows from $g(f_t,\varepsilon)/2R(f_t,\varepsilon)\leqslant\varepsilon$. ∎

Lemma 9. *For $\varepsilon>0$ and $\eta\in(0,1)$, if $p\in\hat{\Theta}(a),q=p_a,\operatorname{supp}\chi=\{q'\in\Delta(Y):q'=p'_a,p'\in\Theta\}$ then $g\left((1-\eta)p_a^*+\eta q,\varepsilon\right)\geqslant2\eta\varepsilon^2$.*

Proof. H is linear in its first argument, so for $\eta\in(0,1)$, $\operatorname{argmin}_{p'_a:p'\in\Theta}H((1-\eta)p_a^*+\eta q,p'_a)=\{q\}$. Then

$$\begin{aligned}&g\left((1-\eta)p_a^*+\eta q,\varepsilon\right)\\ &\geqslant\min_{q'\in\Delta(Y)\setminus Q_{\varepsilon,\chi_0}}\sum_{y\in Y}\left[(1-\eta)p_a^*\left(y\right)+\eta q\left(y\right)\right]\log q'\left(y\right)-\sum_{y\in Y}\left[(1-\eta)p_a^*\left(y\right)+\eta q\left(y\right)\right]\log q\left(y\right)\\ &\geqslant(1-\eta)\min_{q'\in\Delta(Y)\setminus Q_{\varepsilon,\chi_0}}\sum_{y\in Y}p_a^*\left(y\right)\log\left(\frac{q'\left(y\right)}{q\left(y\right)}\right)+\eta\inf_{q'\in\Delta(Y)\setminus B_\varepsilon(q)}\sum_{y\in Y}q\left(y\right)\log\left(\frac{q'\left(y\right)}{q\left(y\right)}\right)\\ &\geqslant0+\eta\inf_{q'\in\Delta(Y)\setminus B_\varepsilon(q)}\sum_{y\in Y}q\left(y\right)\log\left(\frac{q'\left(y\right)}{q\left(y\right)}\right)\geqslant2\eta\varepsilon^2,\end{aligned}$$

where the first inequality follows from the definition of g, the second from concavity of the minimum, the third from the fact that q is a KL minimizer, and the fourth is Pinsker inequality. ∎

Remark 1. Observe that after every finite time t, the posterior μ_t satisfies the Assumption 1. That (i) is satisfied follows from the fact that $\operatorname{supp}\mu_t \subseteq \operatorname{supp}\mu_0$. For (ii), let $\Psi : \mathbb{R}_+ \to \mathbb{R}_{++}$ be the function whose existence is guaranteed by the regularity assumption (ii). Bayesian updating implies that for every $p \in \Theta$, $\varepsilon > 0$, $\mu_t(B_\varepsilon(p)) \geqslant \mu_0(B_\varepsilon(p))D^t \geqslant \Psi(\varepsilon)D^t$ a.s. Therefore, by defining $\Psi_t = D^t\Psi$ we have $\lim_{n\to\infty} \Psi_t(K/n)e^n = \lim_{n\to\infty} \Psi(K/n)e^n D^t = \infty$ for all $K > 0$, so (ii) is satisfied. ▲

Proof of Proposition 1. Set $I = R(\hat{q}_{\eta_t}, \varepsilon)$. If $\hat{q}_{\eta_t} = \frac{\sum_{i=\tau+1}^{\tau+t} \mathbf{1}_{y_i=y}}{t}$, we have

$$\begin{aligned}
&\frac{\mu_{\tau+t}\left(\{p \in \Theta\colon \forall y \in \operatorname{supp} p_a^*, |p_a(y) - \hat{q}(y)| < \varepsilon\}\right)}{1 - \mu_{\tau+t}\left(\{p \in \Theta\colon \forall y \in \operatorname{supp} p_a^*, |p_a(y) - \hat{q}(y)| < \varepsilon\}\right)} \\
\geqslant\ & \mu_\tau\left(\{p \in \Theta\colon \forall y \in \operatorname{supp} p_a^*, |p_a(y) - \hat{q}(y)| < \frac{g(\hat{q}_{\eta_t}, \varepsilon)}{2I}\}\right) e^{.5tg\left(\hat{q}_{\eta_t}, \varepsilon\right)} \\
\geqslant\ & \mu_\tau\left(\{p \in \Theta\colon \forall y \in \operatorname{supp} p_a^*, |p_a(y) - \hat{q}(y)| < \varepsilon^2 \frac{2}{It^{\frac{1}{2}}}\}\right) \exp\left(t\eta_t\varepsilon^2\right) \geqslant D^\tau\Psi\left(\varepsilon^2\frac{2}{It^{\frac{1}{2}}}\right)\exp\left(2t^{\frac{1}{2}}\varepsilon^2\right),
\end{aligned}$$

where the first inequality follows from Lemma 8, the second from Lemma 9, and the third from Assumption 1(ii) and Remark 1.

By Lemma 7 there exists a $\hat{K}, K' > 0$ such that if $||\hat{q}_{\eta_t} - f_t|| < ||\hat{q} - p_a^*||t^{-\frac{1}{2}}/K'$ then

$$\frac{\mu_{\tau+t}\left(\{p \in \Theta\colon \forall y \in \operatorname{supp} p_a^*, |p_a(y) - \hat{q}(y)| < \varepsilon\}\right)}{1 - \mu_{\tau+t}\left(\{p \in \Theta\colon \forall y \in \operatorname{supp} p_a^*, |p_a(y) - \hat{q}(y)| < \varepsilon\}\right)} \geqslant D^\tau\Psi\left(\hat{K}\varepsilon^2\frac{2}{It^{\frac{1}{2}}}\right)\exp\left(2\hat{K}t^{\frac{1}{2}}\varepsilon^2\right).$$

■

A.4 Proof of Results Stated in the Text

Proof of Theorem 1. We prove the statement by contraposition. Suppose that a is a limit action under the optimal policy π, and let $(a_i, y_i)_{i=1}^\tau$ be a history with positive probability. We show that if the agent plays a at every period after $(a_i, y_i)_{i=1}^\tau$ almost surely the belief μ_t reaches a region where no optimal policy prescribes a. By Lemma 5 this is enough to obtain the desired conclusion. Since a is not a uniform BN-E, then there is $p' \in \hat{\Theta}(a)$ such that if $\operatorname{supp}\nu \subseteq \mathcal{E}_a(p')$, then $a \notin A^m(\nu)$. We set $q = p'_a$ throughout this proof.

Claim 1. *There exists $\varepsilon > 0$ such that if $\nu \in \Delta(\Theta)$ is such that*

$$\frac{\nu\left(\{p \in \Theta\colon \forall y \in \operatorname{supp} p_a^*, |p_a(y) - q(y)| < \varepsilon\}\right)}{1 - \nu\left(\{p \in \Theta\colon \forall y \in \operatorname{supp} p_a^*, |p_a(y) - q(y)| < \varepsilon\}\right)} > \frac{1-\varepsilon}{\varepsilon},$$

then $\pi(\nu) \neq a$.

Proof. Suppose by contradiction that for every $n \in \mathbb{N}$ there exists a $\nu_n \in \Delta(\Theta)$ such that

$$\frac{\nu_n\left(\{p \in \Theta\colon \forall y \in \operatorname{supp} p_a^*, |p_a(y) - q(y)| < 1/n\}\right)}{1 - \nu_n\left(\{p \in \Theta\colon \forall y \in \operatorname{supp} p_a^*, |p_a(y) - q(y)| < 1/n\}\right)} \geqslant \frac{1 - 1/n}{1/n}$$

and $a = \pi(\nu_n)$. Because $\Delta(\Theta)$ is sequentially compact, $(\nu_n)_{n\in\mathbb{N}}$ has a converging subsequence $(\nu_{n_i})_{i\in\mathbb{N}} \to \nu^*$.

To show that this leads to a contradiction, define $G(\nu) = \max_{\tilde{\pi}} V(\tilde{\pi}, \nu) - \max_{\tilde{\pi}:\tilde{\pi}(\nu)=a} V(\tilde{\pi}, \nu)$. We claim that if $\operatorname{supp}\nu \subseteq \{p \in \Theta\colon \forall y \in \operatorname{supp} p_a^*, p_a(y) = q(y)\}$, then $G(\nu) > 0$. This is because the definition of q implies $\operatorname{supp}\nu \subseteq \mathcal{E}_a(p')$, so $a \notin A^m(\nu)$, and $\operatorname{supp}\nu \subseteq \mathcal{E}_a(p')$, together with Assumption 1(i), implies that the experimentation value of a is 0.

Next note that as shown in Lemma 2, the space of policy functions endowed with the product topology is compact and $V(\cdot, \nu) - V(\cdot, \nu)$ is a continuous function of the policy. Since for every policy $\tilde{\pi}$, $V(\tilde{\pi}, \cdot)$ is continuous in ν, from the Maximum Theorem G is continuous. But then $\nu^*(\{p \in \Theta\colon \forall y \in \operatorname{supp} p_a^*, p_a(y) = q(y)\}) = 1$ and $G(\nu^*) = \lim_n G(\nu_n) = 0$, a contradiction. ∎

In what follows, we fix an $\varepsilon \in \mathbb{R}_{++}$ that satisfies the conditions of Claim 1. Also, fix an outcome $y^0 \in \operatorname{supp} p_a^*$, and let $\tilde{f}_t$ be the empirical frequency of the other $|\operatorname{supp} p_a^*| - 1$ outcomes in the support of p_a^*. Denote by $\tilde{p}_a^*$ the true probabilities of the same $|\operatorname{supp} p_a^*| - 1$ outcomes.

Claim 2. $\tilde{f}_t \cdot t - \tilde{p}_a^* t$ *is a* $|\operatorname{supp} p_a^*| - 1$ *dimensional random walk under the distribution* $\tilde{p}_a^*$*, and the covariance matrix of its increments is nonsingular.*

Proof. Let $y \in \operatorname{supp} p_a^* \backslash \{y^0\}$. The increment of the y dimension at time $t+1$ is equal to

$$\tilde{f}_{t+1}(y) \cdot (t+1) - p_a^*(y) \cdot (t+1) - \tilde{f}_t(y) \cdot t - p_a^*(y) \cdot t = \mathbf{1}_{y_{t+1}=y} - p_a^*(y)$$

and has expected value 0. Therefore, $\tilde{f}_t \cdot t - \tilde{p}_a^* t$ is a $|\operatorname{supp} p_a^*| - 1$ dimensional random walk. Moreover, the covariance matrix for the increments of $\tilde{f}_t \cdot t - \tilde{p}_a^* t$ is given by $\Sigma_{y,y'} = -\tilde{p}_a^*(y)\tilde{p}_a^*(y')$ if $y \neq y'$ and $\tilde{p}_a^*(y)(1 - \tilde{p}_a^*(y))$ if $y = y'$. To see this, observe that the covariance

between 1_y and $1_{y'}$ is given by:

$$\begin{aligned}
& \tilde{p}_a^*(y)\left(1-\mathbb{E}_{\tilde{p}_a^*}(1_y)\right)\left(0-\mathbb{E}_{\tilde{p}_a^*}(1_{y'})\right)+\tilde{p}_a^*(y')\left(0-\mathbb{E}_{\tilde{p}_a^*}(1_y)\right)\left(1-\mathbb{E}_{\tilde{p}_a^*}(1_{y'})\right) \\
& +\left(1-\tilde{p}_a^*(y')-\tilde{p}_a^*(y)\right)\left(0-\mathbb{E}_{\tilde{p}_a^*}(1_y)\right)\left(0-\mathbb{E}_{\tilde{p}_a^*}(1_{y'})\right) \\
= \; & \tilde{p}_a^*(y)\left(1-\tilde{p}_a^*(y)\right)\left(-\tilde{p}_a^*(y')\right)+\tilde{p}_a^*(y')\left(-\tilde{p}_a^*(y)\right)\left(1-\tilde{p}_a^*(y')\right) \\
& +\left(1-\tilde{p}_a^*(y')-\tilde{p}_a^*(y)\right)\left(-\tilde{p}_a^*(y')\right)\left(-\tilde{p}_a^*(y)\right) \\
= \; & -\tilde{p}_a^*(y)\,\tilde{p}_a^*(y')\left[2-\tilde{p}_a^*(y)-\tilde{p}_a^*(y')-1+\tilde{p}_a^*(y')+\tilde{p}_a^*(y)\right]=-\tilde{p}_a^*(y)\,\tilde{p}_a^*(y').
\end{aligned}$$

By part M35 of Theorem 2.3 of Berman and Plemmons (1994), page 137, if for every row of the covariance matrix the entry on the diagonal is larger than the sum of the off-diagonal entries, then the matrix is diagonal dominant, and so non singular.[29] And for all $y' \in Y$, we have that

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

concluding the proof of the claim. ■

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 $F_t=B_{\frac{||q-p_a^*||/K'}{\sqrt{t}}}\left(\tilde{p}_a^*+\frac{1}{\sqrt{t}}(q-p_a^*)\right)$. We have that

$$\mathbb{P}\left[\tilde{f}_t\in F_t\right]=\mathbb{P}\left[\sqrt{t}(\tilde{f}_t-\tilde{p}_a^*)\in B_{||q-p_a^*||/K'}(q-p_a^*)\right]$$

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

$$\lim_{t\to\infty}\mathbb{P}\left[\tilde{f}_t\in F_t\right]=\mathbb{P}\left[\tilde{Z}\in B_{||q-p_a^*||/K'}(q-p_a^*)\right]$$

where $\tilde{Z}$ is a random variable that is normally distributed with mean $\vec{0}$ and covariance matrix $\Sigma_{y,y'}$. Thus if we let E_t denote the event $f_t\in F_t$, it follows that $\sum_{t=1}^{\infty}\mathbb{P}[E_t]=\infty$. Moreover,

$$\liminf_{t\to\infty}\frac{\sum_{s=1}^{t}\sum_{r=1}^{t}\mathbb{P}[E_s \text{ and } E_t]}{\left(\sum_{s=1}^{t}\mathbb{P}[E_s]\right)^2}=\liminf_{t\to\infty}\frac{\frac{1}{t^2}\sum_{s=1}^{t}\sum_{r=1}^{t}\mathbb{P}[E_s \text{ and } E_r]}{\left(\frac{1}{t}\sum_{t=1}^{\infty}\mathbb{P}[E_t]\right)^2}\leqslant\liminf_{t\to\infty}\frac{\frac{1}{t^2}\sum_{s=1}^{t}\sum_{r=1}^{t}\mathbb{P}[E_r]}{\left(\frac{1}{t}\sum_{s=1}^{t}\mathbb{P}[E_s]\right)^2}$$

$$=\liminf_{t\to\infty}\frac{\frac{1}{t}\sum_{r=1}^{t}\mathbb{P}[E_r]}{\left(\frac{1}{t}\sum_{s=1}^{t}\mathbb{P}[E_s]\right)^2}=\frac{1}{\lim_{t\to\infty}\mathbb{P}[E_t]}=\frac{1}{\mathbb{P}\left[\tilde{Z}\in B_{||q-p_a^*||/K'}(q-p_a^*)\right]}.$$

It then follows from the Kochen-Stone lemma (see Kochen and Stone (1964) or Exercise

[29] This statement is the special case in which D is the identity.

2.3.20 in Durrett (2008)) that

$$\mathbb{P}\left[\bigcap_{t=1}^{\infty}\bigcup_{s=t}^{\infty} E_s\right] \geqslant \mathbb{P}\left[\tilde{Z} \in B_{||q-p_a^*||/K'}\left(q-p_a^*\right)\right] > 0\,.$$

The event $\bigcap_{t=1}^{\infty}\bigcup_{s=t}^{\infty} E_s$ is invariant under finite permutations of the increments $\left(\mathbf{1}_{y_t=y^1},...,\mathbf{1}_{y_t=y^{|\operatorname{supp} p_a^*|-1}} - p_a^*\right)$ 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 $\bigcap_{t=1}^{\infty}\bigcup_{s=t}^{\infty} E_s$ is zero or one, and since it is strictly positive it must equal one.

This implies that $f_t \in F_t$ infinitely often with probability 1. So, by Proposition 1 the agent will eventually take an action different from a. ■

Proof of Theorem 2. *If.* Consider a uniformly strict BN-E a, an optimal policy π and $\kappa \in (0,1)$. By Lemma 3, there exists an ε such that if $\nu(\hat{\Theta}^{\varepsilon}(a)) \geqslant 1-\varepsilon$, then $\pi(\nu) = a$.

Recall that for every $l \in (0,1)$, the function $f_l : P \times P \to \bar{\mathbb{R}}$ is defined by

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

By Lemma 6, since $\hat{\Theta}^{\varepsilon}(a)$ is compact by Lemma 1, and since f_l is lower semicontinuous in its first argument, there exists $\varepsilon' \in (0,\varepsilon)$ such that $\bar{p} \in \hat{\Theta}^{\varepsilon'}(a)$ implies that $f_l(\bar{p},p') < 1$ for all p' with $p' \notin \hat{\Theta}^{\varepsilon}(a)$. Let $K = \left(\frac{\varepsilon}{1-\varepsilon}\right)^l$. Then

$$\left(\frac{1-\nu\left(\hat{\Theta}^{\varepsilon}(a)\right)}{\nu\left(\hat{\Theta}^{\varepsilon'}(a)\right)}\right)^l < K \implies \frac{1-\nu\left(\hat{\Theta}^{\varepsilon}(a)\right)}{\nu\left(\hat{\Theta}^{\varepsilon}(a)\right)} < \frac{\varepsilon}{1-\varepsilon}$$
$$\implies \nu\left(\hat{\Theta}^{\varepsilon}(a)\right) > 1-\varepsilon \implies \pi(\nu) = a.$$

Let $\bar{\varepsilon}$ be such that $\nu\left(\hat{\Theta}^{\bar{\varepsilon}}(a)\right) > 1-\bar{\varepsilon}$ implies that

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

Then if the agent starts with a belief ν_0 with $\nu_0(\hat{\Theta}^{\varepsilon}(a)) > \bar{\varepsilon}$, $A(\nu_0) = \{a\}$. Moreover, by Lemma 4, Dubins' upcrossing inequality, the compactness of $\hat{\Theta}^{\varepsilon}(a)$ guaranteed by Lemma 1, and the union bound, there is a probability $(1-\kappa)$ that the positive supermartingale

$\left(\frac{1-\nu_t'\left(\hat{\Theta}^\varepsilon(a)\right)}{\nu_t'\left(\hat{\Theta}^\varepsilon(a)\right)}\right)^l$ never rises above K, so the action played is always a, and $\bar{\varepsilon}$ satisfies the requirement of the statement.

Only if. If a is not a uniformly strict BN-E, there exists $p \in \hat{\Theta}(a)$ and $b \neq a$ such that $\{b\} \in A^m(\delta_p)$. But then if we let $\nu = \delta_p$ we have that $\nu\left(\hat{\Theta}(a)\right) = 1$. Moreover, there exists a policy π that prescribes b at belief ν, so that the agent will never update their belief and will play b forever. ■

Proof of Theorem 3. $(i) \Rightarrow (ii)$ Immediately follows by Theorem 2.

$(ii) \Rightarrow (i)$ We prove the statement by contraposition. Suppose that a is not a uniformly strict BN-E, and let $\nu \in \Delta(\Theta)$, $\varepsilon > 0$. We construct an initial belief ν_ε that is ε close to ν but such that the actions do not converge to a.

Since a is not a uniformly strict BN-E, there exists $\hat{p} \in \hat{\Theta}(a)$ with $\{a\} \neq A^m(\delta_{\hat{p}})$. By Lemma 1, we can pick a finite collection of open balls $(C_{\varepsilon,i})_{i=1}^n$ of radius ε in $\Delta(P)$ that covers $\hat{\Theta}(a)$ and such that for each $C_{\varepsilon,i} \cap \hat{\Theta}(a) \neq \varnothing$. For every $C_{\varepsilon,i}$, choose $q_{\varepsilon,i} \in C_{\varepsilon,i} \backslash \hat{\Theta}(a)$ whose existence follows from the assumption of the theorem.

Define $\Phi_\varepsilon : \Theta \to 2^\Theta$ as

$$\Phi_\varepsilon(p) = \begin{cases} \{q_{\varepsilon,i} : p \in C_{\varepsilon,i}\} & \text{if } p \in C_{\varepsilon,i} \text{ for some } i \\ \{p\} & \text{otherwise} \end{cases}.$$

The correspondence Φ_ε is Borel measurable, nonempty, and closed valued, so it has a measurable selection ϕ_ε by the Kuratowski Selection Theorem (see, e.g., Theorem 18.13 in Aliprantis and Border (2013)). Define $\bar{\nu}_\varepsilon(C) = \nu(\phi_\varepsilon^{-1}(C))$. Because the problem is rich, there is $p' \in \Theta \cap B_\varepsilon(\hat{p})$ such that $H(p_a', p_a^*) < \min_{p \in \operatorname{supp} \bar{\nu}_\varepsilon} H(p_a, p_a^*)$ and $a \notin A^m(\delta_{p'})$. Set $\nu_\varepsilon = \varepsilon\delta_{p'} + (1-\varepsilon)\bar{\nu}_\varepsilon$. Then $\nu_\varepsilon \to \nu$, but $\operatorname*{argmin}_{p' \in \operatorname{supp} \nu_\varepsilon} H(p_a^*, p_a') = \{\hat{p}\}$, so by Theorem 1, the probability of converging to a starting from belief ν_ε is 0. ■

Proof of Theorem 4. By the hypothesis of the theorem $p_a(y) = p_{a'}(y)$, and $p_a^*(y) > 0$ if and only if $p_a(y) > 0$ for all $p \in \Theta$ by Assumption 1(i). Thus $p_a^*(y) > 0$ if and only if $p_{a'}^*(y) > 0$ for all $a, a' \in A$, i.e. $p_a^*, p_{a'}^*$ are mutually absolutely continuous. Since the agent believes that actions do not change the outcome distribution, every $p \in \Theta$ can be identified with an element of $\Delta(Y)$, and every belief $\nu \in \Delta(\Theta)$ can be identified with an element of $\Delta(\Delta(Y))$.

Consider a uniformly strict BN-E a. By Lemma 1, $\Delta(\hat{\Theta}(a))$ is compact. For every $\bar{\varepsilon} > 0$ and $q \in \Delta(Y)$ let $Q_{\bar{\varepsilon}}(q) = Q_{\bar{\varepsilon},\mu_{0,a}}(q)$. By Theorem 2, there exists $\varepsilon' > 0$ such that if $\varepsilon' > \varepsilon$

and $\nu\left(\text{cl}\left(Q_\varepsilon\left(p_a^*\right)\right)\right) > 1-\varepsilon$ implies $A^m\left(\nu\right) = \{a\}$ the probability of playing a forever starting from belief ν is larger than $1/2$. By the Maximum Theorem, the correspondence Q_ε is upper-hemicontinuous, so there is a sequence of outcomes y^t with corresponding empirical frequency $\hat{p}_t(y) = \frac{1}{t}\sum_{i=1}^{t} \mathbf{1}_{y_i=y}$ sufficiently close to p_a^* to have

$$\hat{q} \in Q_{\varepsilon'/2}\left(\hat{p}_t\right), q \in Q_{\varepsilon'/2}\left(p_a^*\right) \implies ||\hat{q} - q|| < \varepsilon/2.$$

This implies $Q_{\varepsilon'/2}\left(\hat{p}_t\right) \subseteq Q_{\varepsilon'}\left(p_a^*\right)$ from the triangle inequality. Thus by Lemma 8 there is a time T such that for all $t' > T$, if the empirical frequency is $\hat{p}_{t'} = \hat{p}_t$, the agent assigns a relative probability higher than K to an ε' ball around p_a^*:

$$\frac{\mu_{t'}(Q_{\varepsilon'}\left(p_a^*\right))}{1-\mu_{t'}(Q_{\varepsilon'}\left(p_a^*\right))} \geqslant \frac{\mu_{t'}(Q_{\varepsilon'/2}\left(\hat{p}_t\right))}{1-\mu_{t'}(Q_{\varepsilon'}\left(p_a^*\right))} > \frac{K}{2}.$$

Replicating y^t sufficiently many times yields a sequence $y^{t'}$ with empirical frequency $\hat{p}_{t'} = \hat{p}_t$ and $t' > T$. Since p_a^* is absolutely continuous with respect to $p_{a'}^*$ for all $a' \in A$, this sequence of outcomes has positive probability, and after it occurs the agent plays a. By Lemma 4 and the law of iterated expectations, conditional on a being played $\left(\frac{1-\mu_{t'}(Q_\varepsilon(p_a^*))}{\mu_{t'}(Q_{\varepsilon'/2}(\hat{p}_t))}\right)^l$ is a positive supermartingale. Then by Dubins' upcrossing inequality, there is positive probability that this positive supermartingale never rises above $1/K^l$, so a is played forever. ∎

Proof of Theorem 5. Let b be a weakly identified strict BN-E. Then there is $\nu \in \Delta(\hat{\Theta}(b))$ with $\{b\} = A^m(\nu)$. Since b is a strict BN-E, and the agent believes the outcome distributions are independent across actions, we can let $\nu = \delta_p$ where $p_b = \text{argmax}_{p'_b : p' \in \Theta} \mathbb{E}_{p'_b}\left[u(b,y)\right]$, and $p_a = \text{argmin}_{p'_a : p' \in \Theta} \mathbb{E}_{p'_a}\left[u(a,y)\right]$ for $a \in A \backslash \{b\}$. Let $\{y\left(b\right)_i\}_{i=1}^{\infty}$ be a sequence of outcomes such that the empirical frequency $\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{y(b)_i=y}$ is converging to p_b. By Lemma 8, for every $\varepsilon \in (0,1)$, there exists K_ε such that for all $t > K_\varepsilon$, $\mu_{0,b}\left(B_\varepsilon(p_b) \mid y\left(b\right)^t\right) > 1-\varepsilon$.

Because $\{b\} = A^m\left(\nu\right)$, there is $\bar{\beta} \in (0,1)$ such that for all $\beta > \bar{\beta}$, there is $\left(\varepsilon_a\right)_{a \in A} \in \mathbb{R}_+^A$ such that if the belief $\bar{\nu}$ is such that $\bar{\nu}_b \in \{\mu_{0,b}\left(\cdot \mid y\left(b\right)_t\right) : 0 \leqslant t \leqslant K_{\varepsilon_b}\} \bigcup \{\nu'_b : \nu'_b\left(B_\varepsilon(p_b)\right) > 1-\varepsilon\}$, and for all $a' \neq b$, $\nu_{a'}\left(B_{\varepsilon_{a'}}\left(p_{a'}\right)\right) > 1-\varepsilon_{a'}$, then b has the highest Gittins index. For each $\beta > \bar{\beta}$, let $\varepsilon_\beta < \varepsilon$ be such that if $\bar{\nu}_b\left(B_{\varepsilon_\beta}(p(b))\right) > (1-\varepsilon_\beta)$ then the probability of converging to play action a is larger than $\frac{1}{2}$ under any optimal policy given the discount factor β, whose existence is guaranteed by Lemma 14 and the fact that b is weakly identified.

For every $a \neq b$, let $\bar{n}_a \geqslant n_a$ and $\{y\left(a\right)_i\}_{i=1}^{n_a}$ be a sequence of outcomes such that the empirical frequency $\hat{p}_{n_a}\left(a\right)$ converges to p_a. By Lemma 8, for every $a \neq b$ there is a finite n_a such that after n_a observations $\nu_a\left(B_{\varepsilon_a}(p_a) \mid \hat{p}_{n_a}\right) > 1-\varepsilon_a$. Finally, let $n_b = K_{\varepsilon_\beta}$. Then

the array $\left(\{y(a)_i\}_{i=1}^{n_a}\right)_{a\in A}$ has positive probability, so 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 Corollary 1. Let π be an optimal policy. If a is quasi-dominated, with $\hat{p} \in \hat{\Theta}(a)$ as in the definition, there exists $\varepsilon \in (0,1)$ such that if $\nu_a(\{q : ||q - \hat{p}_a|| \leqslant \varepsilon\}) > 1 - \varepsilon$ implies $\pi(\nu) \neq a$. Suppose by way of contradiction that a is played infinitely many times. Then by the last part of the proof of Theorem 1, since the problem is a subjective bandit there is t such that $\mu_a(\{q : ||q - \hat{p}_a|| \leqslant \varepsilon\}|(a^t, y^t)) > 1 - \varepsilon$, so the agent switches to another action b. Since while playing an action different from a the agent does not update μ_a, $\mu_a(\{q : ||q - \hat{p}_a|| \leqslant \varepsilon\}|(a^\tau, y^\tau)) > 1 - \varepsilon$ for all $\tau > t$, so they will not switch to a anymore, a contradiction. ■

Proof of Theorem 6. We prove the statement for $\bar{a}$, the proof for $\underline{a}$ is analogous. Denote the optimal policy used by the agent as π. Since the environment is strongly supermodular, every class of observationally equivalent outcome distributions under action $\bar{a}$ is a singleton, so $\bar{a}$ is a uniformly strict BN-E. Theorem 2 and the strong supermodularity of the environment then imply there is $\bar{p} \in \Theta$ and $K \in (0,1)$ such that if $\nu(\{p : p > \bar{p}\}) > K$, then the probability that a is used forever is larger than $\frac{1}{2}$. Denote the highest outcome as $\bar{y}$. Since the environment is strongly supermodular, for every action $b \in A$,

$$\frac{\mu_{t+1}(\{p : p > \bar{p}\} \mid (a^t, y^t), (b, \bar{y}))}{1 - \mu_{t+1}(\{p : p > \bar{p}\} \mid (a^t, y^t), (b, \bar{y}))} > \frac{\mu_t(\{p : p > \bar{p}\} \mid (a^t, y^t))}{1 - \mu_t(\{p : p > \bar{p}\} \mid (a^t, y^t))}.$$

Therefore, there exists a finite number $n(b)$ such that if $a_t = b$ and $y_t = \bar{y}$ for all $t \leqslant n(b)$, then $\mu_t(\{p : p > \bar{p}\} \mid (a^t, y^t)) \geqslant K$.

Consider the event E that for all $b \in A$ and $t \leqslant n(b)$, $x_{t,b} = \bar{y}$. This event has strictly positive probability $\mathbb{P}_\pi[E]$. Moreover, if E realizes, after some $\hat{T} \leqslant \sum_{b\neq\bar{a}} (n(b) - 1) + 1$, the policy of the agent prescribes action $\bar{a}$. Therefore, after $\hat{T} + n(\bar{a})$, for all $\tau \leqslant \hat{T} + n(\bar{a})$, and for all $y \in Y, \mathbb{P}[x_{\tau,\bar{a}} = y|E] = \mathbb{P}[x_{\tau,\bar{a}} = y]$. Therefore, by Theorem 2 the probability of converging to $\bar{a}$ is at least $\frac{\mathbb{P}_\pi[E]}{2}$. ■

A.5 Action Frequencies and Mixed Equilibria

By Theorem 1, if action a is not a uniform BN-E, the agent will use a different action b infinitely often. We can the use a result from of Esponda, Pouzo, and Yamamoto (2019) to show that if action b's outcome distribution does not induce a as a myopic best reply, the agent will spend a nontrivial fraction of time using actions different from a. For every $a \in A$,

let $\underline{\Theta}(a) = \left\{p \in \hat{\Theta}(a) : a \notin A^m(\delta_p)\right\}$. Let $C_a \subseteq \Delta(\Theta)$ be the largest convex set such that (i) it contains all ν with $\operatorname{supp}\nu = \Theta(a)\backslash\underline{\Theta}(a)$, and (ii) $a \in A^m(\nu)$ for all $\nu \in C_a$. That is, C_a contains all the beliefs supported on the "good" KL minimizers for action a that induce a as a best reply, as well as the beliefs around them that still support a as a myopic best reply. Also, let $A_a = \{b : \exists \nu \in C_a, b \in A^m(\nu)\}$.

Corollary 2. *Let $\beta = 0$, and suppose $a \in A$ is a non-uniform BN-E. If there is $\bar{p} \in \underline{\Theta}(a)$ such that $H(p_b^*, \bar{p}_b) < H(p_b^*, \hat{p}_b)$ for all $b \in A_a$ and $\hat{p} \in \Theta\backslash\{\bar{p}\}$, then $\liminf \frac{1_{a_t=a}}{t} \neq 1$ a.s.*

Proof of Corollary 2. Let $\varepsilon > 0$ be such that if $||\bar{p} - p|| \leqslant \varepsilon$, then $a \notin A^m(\delta_p)$. By assumption, there exists $\varepsilon' > 0$ such that $\hat{\Theta}(\alpha) \subseteq \{p \in \Theta : ||\bar{p} - p|| \leqslant \varepsilon\}$ for all $\alpha \in \Delta(A)$ such that $||\alpha - a|| < \varepsilon'$, $\operatorname{supp}\alpha = A_a \bigcup \{a\}$. Suppose by way of contradiction that $\liminf \frac{1_{a_t=a}}{t} = 1$. Let $W_\tau(\alpha) \subseteq \Delta(A)^{[\tau,\infty)}$ be the set of all differentiable functions $\gamma : [\tau, \infty) \to \Delta(A)$ such that

$$\frac{\partial \gamma_t}{\partial t} \in \Delta\left(A^m\left(\Delta\left(\hat{\Theta}(\gamma_t)\right)\right)\right) - \gamma_t$$

and $\gamma_0 = \alpha$. Define the random variable $\hat{\alpha}_t$ to be the empirical frequency of actions up to time t, i.e., $\hat{\alpha}_t(b) = \frac{1_{a_t=b}}{t}$ for all $b \in A$. For every $\tau \in [t, t+1]$ let $\hat{\alpha}_\tau(b) = \hat{\alpha}_t(b)(\tau - t) + \hat{\alpha}_{t+1}(b)(t + 1 - t)$. From the convergence result (Theorem 2) of Esponda, Pouzo, and Yamamoto (2019), for all $T > 0$ $\lim_{t\to\infty} \inf_{\gamma_t \in W_t(\hat{\alpha}_t)} \sup_{0\leqslant s\leqslant T} ||\hat{\alpha}_{t+s}(a) - \gamma_{t+s}(a)|| = 0$ a.s. By Theorem 1, for all $t' \in \mathbb{N}$ almost surely there is a $\hat{t} \geqslant t'$ such that $\mu_{\hat{t}} \notin C_a$. But then, since the frequency of action a decreases in a ball of size ε outside C_a, for all $\gamma \in W_{\hat{t}}(\hat{\alpha}_{\hat{t}})$, we have $||\gamma_{\hat{t}+\varepsilon'\hat{t}}(a) - 1|| > \varepsilon'$ and $\lim_{s\to\infty} \hat{\alpha}_{\hat{t}+s} = a$, a contradiction. ■

There are two reasons that multiple actions can be played with positive probability in a BN-E: Either every action played can be justified with the same belief over the KL minimizers, or different beliefs are needed to justify some of them. The first case requires the agent to be indifferent between the different actions, so here the BN-E cannot be uniformly strict. However, signals that take the form of payoff perturbations can allow us to obtain such equilibria as the limit of uniformly strict Berk-Nash equilibria, and the associated purification can be uniformly stable and positively attractive.

Formally, for every $\alpha \in \Delta(A)$ and $p \in \Theta$, let

$$H_\alpha(p^*, p) = \sum_{b\in A} \alpha(b)\, p_b^*(y) \log p_b(y) \text{ and } \hat{\Theta}(\alpha) = \operatorname*{argmin}_{p\in\Theta} H_\alpha(p^*, p).$$

Definition 13. The mixed action $\alpha \in \Delta(A)$ is a *strongly uniform mixed BN-E* if all actions $a \in \operatorname{supp} \alpha$ are myopically optimal for all $\theta \in \hat{\Theta}(\alpha)$.

Given a problem (A, Y, p^*, u, Θ) without signals, a problem with signals $(A, Y, S, \zeta, \tilde{p}^*, \tilde{u}, \tilde{\Theta})$ is its (ε, v) *perturbation*, $\varepsilon \in \mathbb{R}_+, v : A \times Y \times S \to \mathbb{R}$, if (i) $\tilde{u}(a, y, s) = u(a, y) + \varepsilon v(a, y, s)$, (ii) $\tilde{p}^*_{a,s}(y) = p^*_a(y)$ and (iii) $\tilde{\Theta} = \{\tilde{p} : \exists p \in \Theta, \tilde{p}_{a,s}(y) = p_a(y), \forall (a, y, s) \in A \times Y \times S\}$.

Corollary 3. *If α is a strongly uniform mixed BN-E in (A, Y, p^*, u, Θ), there is a sequence of strategies $(\sigma_n)_{n\in\mathbb{N}}$ such that each $\sigma_{1/n}$ is a uniformly stable BN-E of a $(1/n)$-perturbation of (A, Y, p^*, u, Θ) and $\lim_{n\to\infty} \zeta(s : \sigma_n(s) = a) = \alpha(a)$ for all $a \in A$. If (A, Y, p^*, u, Θ) is subjectively exogenous and p^* has full support, there are positively attractive $\sigma_{1/n}$.*

The proof is in Section B.2 of the Supplemental Material.

References

Aliprantis, C. and K. Border (2013). *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Berlin: Springer-Verlag.

Arrow, K. and J. Green (1973). "Notes on Expectations Equilibria in Bayesian Settings". Working Paper No. 33, Stanford University. URL: https://scholar.harvard.edu/green/publications/notes-expectations-equilibria-bayesian-settings-institute-mathematical-studies-s.

Bell, A. et al. (2019). "Do tax cuts produce more Einsteins? The impacts of financial incentives versus exposure to innovation on the supply of inventors". *Journal of the European Economic Association*.

Benaïm, M. and M. W. Hirsch (1999). "Mixed Equilibria and Dynamical Systems Arising from Fictitious Play in Perturbed Games". *Games and Economic Behavior* 29, pp. 36–72.

Berk, R. H. (1966). "Limiting Behavior of Posterior Distributions when the Model is Incorrect". *The Annals of Mathematical Statistics* 37, pp. 51–58.

Berman, A. and R. J. Plemmons (1994). *Nonnegative Matrices in the Mathematical Sciences*. New York, New York: SIAM.

Bohren, J. A. (2016). "Informational Herding with Model Misspecification". *Journal of Economic Theory* 163, pp. 222–247.

Bohren, J. A. and D. Hauser (2020). "Learning with Model Misspecification: Characterization and Robustness". URL: https://www.dropbox.com/s/f1c73d8t3mnt9p2/BohrenHauser_LearningModelMisspecification_20200821.pdf?dl=0.

Bray, M. (1982). "Learning, estimation, and the stability of rational expectations". *Journal of economic theory* 26, pp. 318–339.

Bray, M. M. and N. E. Savin (1986). "Rational expectations equilibria, learning, and model specification". *Econometrica*, pp. 1129–1160.

Cho, I.-K. and K. Kasa (2015). "Learning and model validation". *The Review of Economic Studies* 82, pp. 45–82.

— (2017). "Gresham's Law of Model Averaging". *American Economic Review* 107, pp. 3589–3616.

Diaconis, P. and D. Freedman (1990). "On the Uniform Consistency of Bayes Estimates for Multinomial Probabilities". *The Annals of Statistics* 18, pp. 1317–1327.

Dudley, R. M. (2018). *Real Analysis and Probability*. Cambridge, UK: Chapman and Hall/CRC.

Durrett, R. (2008). *Probability Models for DNA Sequence Evolution*. New York, New York: Springer.

Eliaz, K. and R. Spiegler (2018). "A Model of Competing Narratives". arXiv: 1811.04232.

Esponda, I. and D. Pouzo (2016). "Berk–Nash equilibrium: A Framework for Modeling Agents with Misspecified Models". *Econometrica* 84, pp. 1093–1130.

— (2019). "Equilibrium in Misspecified Markov Decision Processes". arXiv: 1502.06901.

Esponda, I., D. Pouzo, and Y. Yamamoto (2019). "Asymptotic Behavior of Bayesian Learners with Misspecified Models". arXiv: 1904.08551.

Frick, M., R. Iijima, and Y. Ishii (2019). "Misinterpreting Others and the Fragility of Social Learning". URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3316376.

— (2020). "Stability and Robustness in Misspecified Learning Models". URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3600633.

Fudenberg, D. and K. He (2017). "Player-compatible Equilibrium". arXiv: 1712.08954.

— (2018). "Learning and type compatibility in signaling games". *Econometrica* 86, pp. 1215–1255.

— (2020). "Payoff information and learning in signaling games". *Games and Economic Behavior* 120, pp. 96–120.

Fudenberg, D., K. He, and L. A. Imhof (2017). "Bayesian posteriors for arbitrarily rare events". *Proceedings of the National Academy of Sciences* 114, pp. 4925–4929.

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

Fudenberg, D. and G. Lanzani (2020). "Which Misperceptions Persist?" *Available at SSRN*. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3709932.

Fudenberg, D., G. Lanzani, and P. Strack (2021). *Pathwise Concentration Bounds for Misspecified Bayesian Beliefs*. Tech. rep. Working Paper.

Fudenberg, D. and D. K. Levine (1992). "Maintaining a Reputation When Strategies are Imperfectly Observed". *Review of Economic Studies* 59, pp. 561–581.

Fudenberg, D., G. Romanyuk, and P. Strack (2017). "Active Learning with a Misspecified Prior". *Theoretical Economics* 12, pp. 1155–1189.

Gagnon-Bartsch, T. M. (2016). "Taste Projection in Models of Social Learning". PhD thesis. URL: https://scholar.harvard.edu/files/tasteprojectionearlydraft.pdf.

Gagnon-Bartsch, T., M. Rabin, and J. Schwartzstein (2018). *Channeled Attention and Stable Errors*. Harvard Business School.

Gibbons, R., M. LiCalzi, and M. Warglien (2019). *What situation is this? Coarse cognition and behavior over a space of games*. Tech. rep. Department of Management, Università Ca'Foscari Venezia. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3034826.

He, K. (2019). "Mislearning from Censored Data: The Gambler's Fallacy in Optimal-Stopping Problems". arXiv: 1803.08170.

He, K. and J. Libgober (2020). "Evolutionarily Stable (Mis)specifications: Theory and Application". *In preparation*.

Heidhues, P., B. Kőszegi, and P. Strack (2018). "Unrealistic Expectations and Misguided Learning". *Econometrica* 86, pp. 1159–1214.

— (2019). "Overconfidence and Prejudice". arXiv: 1909.08497.

— (2021). "Convergence in models of misspecified learning". *Theoretical Economics* 16, pp. 73–99.

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

Kagel, J. H. and D. Levin (1986). "The winner's curse and public information in common value auctions". *The American economic review*, pp. 894–920.

Kochen, S. and C. Stone (1964). "A Note on the Borel-Cantelli Lemma". *Illinois Journal of Mathematics* 8, pp. 248–251.

Levy, G., R. Razin, and A. Young (2020). "Misspecified Politics and the Recurrence of Populism". Working Paper. URL: https://drive.google.com/file/d/1yvY-K5Zt9M4NIzeES88DPq_QpgISOK5z/view.

Mailath, G. J. and L. Samuelson (2020). "Learning under Diverse World Views: Model-Based Inference". *American Economic Review* 110, pp. 1464–1501.

Molavi, P. (2019). "Macroeconomics with Learning and Misspecification: A General Theory and Applications". URL: https://www.cemfi.es/ftp/pdf/papers/Seminar/CREE.pdf.

Morrison, W. and D. Taubinsky (2019). "Rules of Thumb and Attention Elasticities: Evidence from Under-and Overreaction to Taxes". NBER Working Paper. URL: https://www.nber.org/papers/w26180.

Neveu, J. (1975). *Discrete-parameter martingales*. Amsterdam: North-Holland.

Nyarko, Y. (1991). "Learning in Mis-specified Models and the Possibility of Cycles". *Journal of Economic Theory* 55, pp. 416–427.

Parthasarathy, K. R. (2005). *Probability Measures on Metric Spaces*. New York, New York: American Mathematical Soc.

Rabin, M. and J. L. Schrag (1999). "First impressions matter: A model of confirmatory bias". *The Quarterly Journal of Economics* 114, pp. 37–82.

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

Rees-Jones, A. and D. Taubinsky (2020). "Measuring "schmeduling"". *The Review of Economic Studies* 87, pp. 2399–2438.

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