Berk-Nash Equilibrium: A Framework for Agents with Misspecified Models*

Ignacio Esponda  Demian Pouzo
(WUSTL) (UC Berkeley)
May 18, 2015

Abstract

We develop an equilibrium framework that relaxes the standard assumption that people have a correctly-specified view of their environment. Each player is characterized by a (possibly misspecified) subjective model, which describes the set of possible beliefs (over distributions over others’ actions and payoff relevant states). We introduce the notion of a Berk-Nash equilibrium: Each player follows a strategy that is optimal given her belief, and her belief is restricted to be the best fit among the set of beliefs she considers possible. The notion of best fit is formalized in terms of minimizing the relative entropy and depends on the equilibrium strategy profile. Standard solution concepts such as Nash equilibrium and self-confirming equilibrium constitute special cases where players learn with correctly-specified models. The framework provides a systematic approach to modeling players with misspecified views and also unifies a specific bounded rationality literature where mistakes are driven by misspecifications. Importantly, we provide a learning foundation for Berk-Nash equilibrium by extending and combining results from the statistics literature on misspecified learning and the economics literature on learning in games.

*We thank Vladimir Asriyan, Pierpaolo Battigalli, Larry Blume, Aaron Bodoh-Creed, Sylvain Chassang, Emílio Espino, Erik Eyster, Drew Fudenberg, Yuriy Gorodnichenko, Philippe Jehiel, Stephan Lauermann, Natalia Lazzati, Kristóf Madarász, Matthew Rabin, Ariel Rubinstein, Joel Sobel, Jörg Stoye, and several seminar participants for helpful comments. Esponda: Olin Business School, Washington University in St. Louis, 1 Brookings Drive, Campus Box 1133, St. Louis, MO 63130, iesponda@wustl.edu; Pouzo: Department of Economics, UC Berkeley, 530-1 Evans Hall #3880, Berkeley, CA 94720, dpouzo@econ.berkeley.edu.
1 Introduction

Economic models provide a simplified framework to understand complex environments. Most economists recognize that the simplifying assumptions underlying our models are often wrong, but we nevertheless make these assumptions in our search for insights. Despite recognizing that our models are likely to be misspecified, the standard approach in economics is to assume that the economic agents themselves have a correctly specified view of their environment. In this paper, we introduce an equilibrium framework that relaxes this standard assumption and forces the modeler to take a stand on the subjective view of the world held by the economic agents.

We define a game to be composed of an objective game and a subjective model. An objective game represents the true environment faced by the players. Payoff relevant states and privately observed signals are realized according to some objective probability distribution. Each player observes her own private signal and then players simultaneously choose actions. The action profile and the realized payoff-relevant state determine a consequence for each player, and these consequences in turn determine each player’s payoffs. This objective description of the game is fairly standard in economics.

While it is also standard to implicitly assume that players know the objective game, we deviate from this practice by assuming that each player has a subjective model that represents her own view of the game. Formally, a subjective model is a set of probability distributions over own consequences conditional on a player’s own action. A key feature is that we allow the subjective model of one or more players to be misspecified, which roughly means that the set of subjective distributions do not include the true, objective distribution. For example, firms might incorrectly believe that sales depend only on their own price and not also on the price of other firms. Or a consumer might perceive a nonlinear price schedule to be linear and, therefore, respond to average, not marginal, prices. Or traders might not realize that the value of trade is partly determined by the terms of trade.

Players play the objective game repeatedly. They believe that the environment they face is stationary—it might not be, if other players are also present and learning simultaneously. In particular, we ignore repeated game considerations where players take into account how their actions affect others’ future play. Players start with a prior over a set of subjective distributions over consequences. In each period, they
play the objective game and use the observed consequences to update their beliefs according to Bayes’ rule. Players maximize discounted expected utility, for some fixed discount factor. The problem of each player can then be cast recursively as a dynamic optimization problem where the state variable is a player’s own belief. The main objective is to characterize limiting behavior when players behave optimally but learn with a possibly misspecified subjective model.

The main result is that, if players’ behavior converges, then it converges to what we call an equilibrium of the game. An equilibrium is defined to be a strategy profile such that, for each player, there exists a belief with support in that player’s subjective model that satisfies two conditions. First, the strategy must be optimal (in a static sense) given the belief. Second, the belief puts probability one on the set of subjective distributions over consequences that are “closest” to the true distribution, where the true distribution is determined by the objective game and the actual strategy profile. The notion of “closest” is given by a weighted version of the Kullback-Leibler divergence, also known as relative entropy, that we define formally in the main text.

A converse of the main result, showing that we can converge to any equilibrium of the game for some initial (non-doctrinaire) prior, does not hold. But we do obtain a positive convergence result by relaxing the assumption that players exactly optimize. We show that, for any equilibrium, there exists a policy rule that is myopic and asymptotically optimal (in the sense that optimization mistakes vanish with time) under which convergence to equilibrium occurs with probability one.

Our notion of equilibrium includes Nash equilibrium as a special case. Suppose that the game is correctly specified, which means that the support of each player’s prior contains the true, objective distribution. Suppose, in addition, that the game is strongly identified, which means that there is always a unique distribution (whether or not correct) that matches the observed data. Then, our notion of equilibrium is equivalent to Nash equilibrium. The novelty of our framework is to force the modeler to make explicit any assumptions about players’ subjective models and to entertain the possibility that these subjective models might be misspecified.

There is a longstanding interest among economists in studying the behavior of economic agents who hold misspecified views of the world. As we illustrate in the paper, there are examples from such diverse fields as industrial organization, mechanism design, psychology and economics, macroeconomics, and information economics, although many times there is no explicit reference to a problem of misspecified learn-
ing. Most of the literature, however, focuses on particular settings, and there has been little progress in developing a unified framework. Moreover, one message that emerges from some of the literature is often discouraging, emphasizing that misspecified models lead to non-convergent behavior, multiplicity of equilibrium, or even non-existence of equilibrium. Our unifying treatment clarifies that these are all natural features of equilibrium analysis—whether or not players are misspecified—and that modeling the behavior of misspecified players does not constitute a large departure from the standard framework.

Early examples of misspecified learning are provided by Arrow and Green (1973), Kirman (1975), Sobel (1984), Nyarko (1991), and Sargent (1999), among others. Arrow and Green (1973) also provide a general treatment and make a distinction between an objective and subjective game. Their framework, though, is more restrictive than ours in terms of the types of misspecifications that players are allowed to have. Moreover, they do not establish existence of equilibrium and they do not provide a learning foundation for equilibrium.\(^1\)

More recently, new equilibrium concepts have been proposed to capture the behavior of players who are boundedly rational and who might be viewed as learning from past interactions: sampling equilibrium (Osborne and Rubinstein (1998), Spiegler (2006)), the inability to recognize patterns (Piccione and Rubinstein (2003), Eyster and Piccione (2013)), valuation equilibrium (Jehiel and Samet, 2007), analogy-based expectation equilibrium (Jehiel (2005), Jehiel and Koessler (2008)), cursed equilibrium (Eyster and Rabin, 2005), behavioral equilibrium (Esponda, 2008), sparse Nash equilibrium (Gabaix, 2012), cursed expectations equilibrium (Eyster et al., 2013), and personal equilibrium (Spiegler, 2014). In particular, analogy-based expectation, (fully) cursed, and behavioral equilibrium can all be integrated into our framework, thus clarifying the underlying misspecification in each of these cases.\(^2\)

\(^1\)Misspecified models have also been studied in contexts that are outside the scope of our paper either because the decision problem is dynamic (instead, we focus on the repetition of a static problem) or because a market mechanism mediates the interactions between agents. Examples include the early literature on rational expectations with misspecified players (e.g., Blume and Easley (1982), Bray (1982), and Radner (1982)), the macroeconomics literature on bounded rationality (e.g., Sargent (1993), Evans and Honkapohja (2001)), a behavioral finance literature that studies under and overreaction to information (e.g., Barberis et al., 1998), and a literature that formalizes psychological biases and studies related applications (e.g., Rabin (2002), Rabin and Vayanos (2010), Spiegler (2013)).

\(^2\)Of course, many aspects of bounded rationality do not seem naturally fitted to misspecified learning, such as the literature that studies biases in information processing due to computational complexity (e.g., Rubinstein (1986), Salant (2011)), bounded memory (e.g., Wilson, 2003), or self-
Our paper is also related to the bandit literature, which shows that agents in a decision problem might optimally end up with incorrect beliefs if experimentation is costly (e.g., Rothschild (1974), McLennan (1984), Easley and Kiefer (1988)). The self-confirming equilibrium literature (Battigalli (1987), Fudenberg and Levine (1993a), Dekel et al. (2004)) extends this insight to games by requiring players to have beliefs that are consistent with observed past play, though not necessarily correct when feedback is coarse. In our framework, we also allow beliefs to be incorrect due to lack of experimentation: our equilibrium notion is equivalent to self-confirming equilibrium when we drop the assumption that the game is strongly identified but maintain the assumption that the subjective model is correctly specified.\(^3\) In all of these cases with coarse feedback, players’ beliefs are endogenous and depend on everyone’s actions. But, when players have misspecified models, we show that beliefs continue to be endogenous even if players persistently experiment with all actions. Thus, an equilibrium framework is needed to characterize behavior even in single-agent settings with perpetual experimentation.\(^4\)

From a technical point of view, our results extend and combine results from two literatures. First, the idea that equilibrium is a result of a learning process is taken from the literature on learning in games. This literature studies explicit learning models in order to justify Nash and self-confirming equilibrium (e.g., Fudenberg and Kreps (1988), Fudenberg and Kreps (1993), Fudenberg and Kreps (1995), Fudenberg and Levine (1993b), Kalai and Lehrer (1993)).\(^5\) In particular, we follow Fudenberg and Kreps (1993) in making the assumption that payoffs are perturbed, à la Harsanyi (1973), to guarantee that behavior is continuous in beliefs and, therefore, to justify how players might learn to play mixed strategy equilibria. We also rely on an idea by Fudenberg and Kreps (1993) to prove the converse of the main result. We extend this literature to account for the possibility that players learn with models of the world that are misspecified even in steady state.

Second, we rely on and contribute to the statistics literature that studies the deception (e.g., Bénabou and Tirole (2002), Compte and Postlewaite (2004)).

\(^3\)In the macroeconomics literature, the term “self-confirming equilibrium” is sometimes used in a broader sense to include cases where agents have misspecified models (e.g., Sargent, 1999).

\(^4\)The literature on self-confirming equilibrium considers two interesting extensions, neither of which are captured in our paper: refinements that restrict beliefs by allowing players to introspect about other players’ motivations (e.g., Rubinstein and Wolinsky, 1994), and non-Bayesian models of updating that capture ambiguity aversion (Battigalli et al., 2012).

\(^5\)See Fudenberg and Levine (1998, 2009) for a survey of this literature.
consistency of Bayesian updating in order to characterize limiting beliefs. In decision problems with correctly-specified models, the standard approach is to use a martingale convergence theorem to prove that beliefs converge (e.g., Easley and Kiefer, 1988). This result guarantees convergence of beliefs from a subjective point of view, which is, unfortunately, not useful for our results because beliefs might still not converge in an objective sense when the agent has a misspecified model. Thus, we take a different route and follow the statistics literature on misspecified learning. This literature characterizes limiting beliefs in terms of the Kullback-Leibler divergence (e.g., Berk (1966), Bunke and Milhaud (1998)). We extend this statistics literature to the case where agents are not only passively learning about their environment but are also actively learning by taking actions.

Finally, in this paper, we take players’ misspecifications as given and characterize the resulting behavior. This is a natural first step towards endogenizing the subjective model. It is important to emphasize, however, that Bayesian players in our setting have no reason to “discover” that they are misspecified.6

We present the framework and several examples in Section 2, discuss the relationship to other solution concepts in Section 3, and provide a learning foundation in Section 4.

2 The framework

2.1 The environment

A (simultaneous-move) game \( \mathcal{G} = \langle \mathcal{O}, \mathcal{Q} \rangle \) is composed of a (simultaneous-move) objective game \( \mathcal{O} \) and a subjective model \( \mathcal{Q} \). We now describe each of these two components in detail. For simplicity, the main proofs in the text assume that all sets are finite.

**Objective game.** A (simultaneous-move) **objective game** is a tuple

\[
\mathcal{O} = \langle I, \Omega, S, p, X, Y, f, \pi \rangle,
\]

where: \( I \) is the set of players (a special case is a single-agent decision problem); \( \Omega \)

---

6Some explanations for why agents may have misspecified models include the use of heuristics (Tversky and Kahneman, 1973), complexity (Aragones et al., 2005), the desire to avoid over-fitting the data (Al-Najjar (2009), Al-Najjar and Pai (2013)), and costly attention (Schwartzstein, 2009).
is the set of payoff-relevant states; $S = \times_{i \in I} S^i$ is the set of profiles of signals, where $S^i$ is the set of signals of player $i$; $p$ is a full-support probability distribution over $\Omega \times S$; $X = \times_{i \in I} X^i$ is a set of profiles of actions, where $X^i$ is the set of actions of player $i$; $Y = \times_{i \in I} Y^i$ is a set of profiles of (observable) consequences, where $Y^i$ is the set of consequences of player $i$; $f = (f^i)_{i \in I}$ is a profile of feedback or consequence functions, where $f^i : \Omega \times X \rightarrow Y^i$ map outcomes in $\Omega \times X$ into consequences of player $i$; and $\pi = (\pi^i)_{i \in I}$, where $\pi^i : X^i \times Y^i \rightarrow \mathbb{R}$ is the payoff function of player $i$.\footnote{While it is redundant to have $\pi^i$ also depend on $x^i$, it helps to reduce notation, particularly in applications.} For simplicity, we prove the results for the case where all of the above sets are finite. In the Online Appendix, we provide technical assumptions under which the results extend to nonfinite $\Omega$ and $Y$.

The timing of the objective game is as follows: First, a state and a profile of signals are drawn according to the probability distribution $p$. Second, each player privately observes her own signal. Third, each player simultaneously chooses an action. Finally, each player observes her consequence and obtains a payoff.\footnote{For simplicity, the setup implicitly assumes that players observe at least their own payoffs and actions.}

A strategy of player $i$ is a mapping $\sigma^i : S^i \rightarrow \Delta(X^i)$. The probability that player $i$ chooses action $x^i$ after observing signal $s^i$ is denoted by $\sigma^i(x^i | s^i)$. A strategy profile is a vector of strategies $\sigma = (\sigma^i)_{i \in I}$; let $\Sigma$ denote the space of all strategy profiles.

**Subjective model.** The subjective model represents the set of distributions over consequences that players consider possible a priori. For a fixed objective game, a **subjective model** is a tuple

$$Q = (\Theta, (Q^i)_{\theta \in \Theta}) ,$$

where: $\Theta = \times_{i \in I} \Theta^i$ is a parameter space, where $\Theta^i$ is the parameter space of player $i$; and $Q^i_{\theta} = (Q^i_{\theta^i})_{i \in I}$, where $Q^i_{\theta^i} = (Q^i_{\theta^i}(\cdot | s^i, x^i))_{(s^i, x^i)}$ is a profile of conditional distributions over player $i$'s consequences parameterized by $\theta^i \in \Theta^i$.\footnote{For simplicity, this definition of a subjective model assumes that players know the distribution over their own signals.} We maintain the following assumptions throughout the paper.\footnote{As usual, the superscript $-i$ denotes a profile where the $i$'th component is excluded.}

**Assumption 1.** For all $i \in I$: (i) $\Theta^i$ is a compact subset of an Euclidean space,
(ii) $Q^i_{\theta^i}(y^i \mid s^i, x^i)$ is continuous as a function of $\theta^i$ for all $(y^i, s^i, x^i) \in \mathcal{Y}^i \times \mathcal{S}^i \times \mathcal{X}^i$.

(iii) For every $\theta^i \in \Theta^i$, there exists a sequence $(\theta^i_n)_{n}$ in $\Theta^i$ such that $\lim_{n \to \infty} \theta^i_n = \theta^i$ and, for all $n$, $Q^i_{\theta^i_n}(y^i \mid s^i, x^i) > 0$ for all $y^i \in f^i(\Omega, x^i, x^{-i})$ and all $(s^i, x^i) \in \mathcal{S}^i \times \mathcal{X}^i$.

The last of these assumptions plays two roles. First, it rules out a stark form of misspecification by guaranteeing that there exists a parameter that can rationalize every feasible observation. Second, it requires that if indeed a feasible event is deemed impossible by some parameter, then that parameter is not isolated in the sense that there are nearby parameters that consider every feasible event to be possible. In Section 4.4, we show that equilibrium may fail to exist and steady-state behavior need not be characterized by equilibrium without this assumption.

While the objective game represents the true environment, the subjective model represents the players’ perception of their environment. This separation between objective and subjective models, which is often implicit in standard treatments of games, is crucial in this paper.

By defining the subjective model directly on the space of consequences, the framework can accommodate more general types of misspecifications. For example, a player does not even need to understand that the consequences of her actions are partly determined by the actions of other players. We now introduce a special type of subjective model that places more structure on the problem and facilitates comparison of our approach to previous solution concepts.

**Strategic subjective model.** Each player $i$ is characterized by a parameter space $\Theta^i$ where each $\theta^i \in \Theta^i$ corresponds to a profile $P^i_{\theta^i}(: \mid s^i) \in \Delta(\Omega \times \mathcal{X}^{-i})$ of probability distributions over the state space and the actions of other players, conditional on her private signal $s^i \in \mathcal{S}^i$. The corresponding subjective model is

$$Q^i_{\theta^i}(y^i \mid s^i, x^i) = P^i_{\theta^i}((\omega, x^{-i}) : f^i(\omega, x^i, x^{-i}) = y^i \mid s^i).$$

The interpretation is that each player understands the game being played; any possibility of mistaken beliefs is captured by restrictions on the probability distribution over $\Omega \times \mathcal{X}^{-i}$. There is a further special case of a strategic model in which players understand that $x^{-i}$ are determined by strategy profiles $\sigma^{-i}$. In this special case, each player has a parameterized model of $(p_\theta, \sigma^{-i}_\theta)$, where $p_\theta \in \Delta(\Omega \times \mathcal{S})$ is a distribution over states and signals and $\sigma^{-i}_\theta$ is a profile of other players’ strategies. Each $(p_\theta, \sigma^{-i}_\theta)$
determines $P_{\theta^i}(\cdot \mid s^i)$ and, subsequently, $Q_{\theta^i}(\cdot \mid s^i, x^i)$.\(^{11}\)

2.2 Examples

**Example 3.1. Monopolist with unknown demand.** A monopolist faces demand function $y = f(x, \omega)$, where $x \in X$ is the price chosen by the monopolist and $\omega$ is a random shock, distributed according to $p \in \Delta(\Omega)$. The monopolist observes sales $y$, but not the random shock. The monopolist’s payoff function is $\pi(x, y) = xy$ (i.e., there are no costs). The monopolist does not know the true demand function $f$. Her subjective model is given by

$$y = a - bx + \omega,$$

where $\omega$ follows a standard normal distribution and $\theta = (a, b) \in \Theta$ is a parameter vector. Then $Q_{\theta}(\cdot \mid x)$ is a normal distribution with mean $a - bx$ and unit variance.

**Example 3.2. Non-linear pricing.** An agent chooses effort $x \in X$ and then income $z = x + \omega$ is realized, where $\omega$ is a random shock with zero mean. The agent pays taxes $\tau = \tau(z)$, where $\tau(\cdot)$ is a nonlinear tax schedule. The agent observes $y = (z, \tau)$ and obtains payoff $\pi(x, z, \tau) = z - \tau - c(x)$, where $c(x)$ is the cost of effort $x$. The agent correctly understands how effort translates into income. She believes, however, that she faces a linear tax schedule: $\tau = \theta z + \varepsilon$, where $\theta \in \Theta$ is the constant tax rate and $\varepsilon$ follows a standard normal distribution and measures unknown aspects of the schedule (such as eligible deductions). This model captures an agent who responds to the average, not the marginal, tax rate.

**Example 3.3. Misspecified market structure.** Two firms compete by simultaneously choosing prices. First, a vector of costs $s = (s_1, s_2)$ and demand shocks $\omega = (\varepsilon_1, \varepsilon_2)$ are drawn according to $p \in \Delta(\Omega \times S)$. Next, each firm $i$ privately observes her own cost $s^i$ and chooses a price $x^i$. The quantity sold by firm $i$ is given by the demand system

$$\ln y^i = f^i(\omega, x) = \alpha^* + \beta^* \ln x^j - \gamma^* \ln x^i + \varepsilon^i,$$

where $\gamma^* > 1$ is the demand elasticity (in absolute value). There are no costs, so the payoff of firm $i$ is $\pi^i(x^i, y^i) = x^i y^i$. Each firm $i = 1, 2$ (incorrectly) believes that they

\(^{11}\)Note that, in the general definition of a strategic subjective model, players are allowed to have correlated beliefs about their opponents’ strategies, as in Fudenberg and Levine (1993a).
are a monopolist in this market and that the demand function they face is

\[
\ln y^i = \alpha^i - \gamma^i \ln x^i + \varepsilon^i, \tag{2}
\]

where \(\varepsilon^i\) follows a standard normal distribution and \(\theta^i = (\alpha^i, \gamma^i)\) parameterizes the subjective model. In particular, the subjective distribution \(Q^i_{\theta^i}(\cdot \mid x^i, s^i)\) over \(Y^i\) follows a lognormal distribution.

**Example 3.4. Regression to the mean.** An instructor observes the initial performance \(s\) of a student and decides whether to praise or criticize him, \(x \in \{C, P\}\). The student then performs again and the instructor observes his final performance, \(s'\). The truth is that performances \(Y = (S, S')\) are independent, standard normal random variables. The instructor’s payoff is \(\pi(x, s, s') = s' - c(x, s)\), where \(c(x, s) = \kappa |s| > 0\) if either \(s > 0, x = C\) or \(s < 0, x = P\), and, in all other cases, \(c(x, s) = 0\).\(^{12}\) The interpretation is that the instructor bears a (reputation) cost from lying that is increasing in the size of the lie, where lying is defined as either criticizing an above-average performance or praising a below-average performance. Because the instructor has no influence on performance, it is optimal to praise if \(s > 0\) and to criticize if \(s < 0\). The instructor, however, does not admit the possibility of regression to the mean and believes that

\[
s' = s + \theta x + \varepsilon,
\]

where \(\varepsilon\) has a standard normal distribution, and \(\theta = (\theta_C, \theta_P) \in \Theta\) parameterizes her perceived influence on the performance of the student.\(^{13}\) We will show that an instructor who ignores regression to the mean will have incorrect beliefs about the influence of her feedback on the student’s performance.

**Example 3.5. Classical monetary policy.** There are two players, the government (G) and the public (P). The government chooses monetary policy \(x^G\) and the public chooses inflation forecasts \(x^P\). Inflation, \(e\), and unemployment, \(U\), are

\(^{12}\)Formally, a state is \(\omega = (s, s')\) and the feedback function is \(y = f(\omega, x) = \omega\).

\(^{13}\)A model that allows for regression to the mean is \(s' = \alpha s + \theta x + \varepsilon\); in this case, the agent will correctly learn that \(\alpha = 0\) and \(\theta x = 0\) for all \(x\).
determined as follows:\(^{14}\)

\[
\begin{align*}
e &= x^G + \varepsilon_e \\
U &= u^* - \lambda(e - x^P) + \varepsilon_U,
\end{align*}
\] (3)

where \(\lambda \in (0, 1)\) and where \(\omega = (\varepsilon_e, \varepsilon_U)\) are random shocks with joint distribution \(p\) and \(Var(\varepsilon_e) > 0\). Thus, inflation is determined by the government’s action and a random term. And unemployment is determined by surprise inflation according to a standard Phillips curve, where \(u^* > 0\) is the natural rate of unemployment.

Realized inflation and unemployment, but not the error terms, are observed by both the public and the government. The government’s payoff is \(\pi(x^G, e, U) = -(U^2 + e^2)\).

For simplicity, we focus on the government’s problem and assume that the public has correct beliefs and chooses \(x^P = x^G\). Under the classical subjective model, the government believes (correctly) that its policy \(x^G\) affects inflation, but it does not realize that unemployment is affected by surprise inflation, and not just by inflation:

\[
\begin{align*}
e &= x^G + \varepsilon_e \\
U &= \theta_1 - \theta_2 e + \varepsilon_U.
\end{align*}
\]

The subjective model is parameterized by \(\theta = (\theta_1, \theta_2)\).

**Example 3.6. Trade with adverse selection.** A buyer and a seller simultaneously submit a (bid) price \(x \in X\) and an ask price \(a \in A\), respectively. If \(a \leq x\), then trade takes place at price \(x\), and the buyer obtains payoff \(v - x\), where \(v\) is the buyer’s value of the object. If \(a > x\), then no trade takes place and each player receives 0. At the time she makes an offer, the buyer does not know her value or the ask price of the seller. The seller’s ask price and the buyer’s value are drawn from \(p \in \Delta(A \times V)\).\(^{15}\)

We consider two different feedback functions to illustrate the importance of making explicit what players observe about the outcome of the game. Under perfect feedback, the buyer observes the ask price and her own value at the end of each period. Under partial feedback, the buyer observes the ask price, but she only observes her own value

\(^{14}\)Formally, a state is \(\omega = (\varepsilon_e, \varepsilon_U)\) and the feedback function \(y = (e, U) = f(\omega, x^G, x^P)\) is given by the system of equations (3).

\(^{15}\)The typical story is that there is a population of sellers each of whom follows the weakly dominant strategy of asking for her valuation; thus, the ask price is a function of the seller’s valuation and, if buyer and seller valuations are correlated, then the ask price and buyer valuation are also correlated.
if she trades in that period. Finally, suppose that $A$ and $V$ are actually correlated but that the buyer naively believes that they are independent. This is formally captured by letting $\Theta = \Delta(A) \times \Delta(V)$.

References for examples. **Example 3.1**: Nyarko (1991) studies a special case and shows that a steady state does not exist, although he does not allow for mixed strategies. **Example 3.2**: Sobel (1984) considers a similar misspecification to capture a consumer who responds to average, not marginal, pricing. **Example 3.3**: Arrow and Green (1973) and Kirman (1975) study similar examples in which firms have incorrect perceptions about market structure. **Example 3.4**: The story of the instructor who does not understand regression to the mean is taken from Tversky and Kahneman (1973); we are not aware of previous attempts to formalize the underlying misspecification. **Example 3.5**: The monetary policy example is based on Sargent (1999, Chapter 7). **Example 3.6**: This misspecification is first discussed in the lemons context by Kagel and Levin (1986). Recently, it has been generalized by Eyster and Rabin (2005), Jehiel and Koessler (2008), and Esponda (2008). The framework in this paper unifies all of the previous examples. In addition, previous work considered special cases with relatively simple misspecifications, in which equilibrium beliefs are intuitively given by simple averages. Our framework, in contrast, provides a characterization of beliefs essentially for all types of misspecified models, and this characterization is not necessarily intuitive in more complex examples.

### 2.3 Definition of equilibrium

**Objective distribution.** Fix an objective game. Given a strategy profile $\sigma$, we obtain an objective distribution over player $i$'s consequences, $Q_i^\sigma$, where, for each $(s^i, x^i) \in S^i \times X^i$, $Q_i^\sigma(\cdot \mid s^i, x^i) \in \Delta(Y^i)$ is defined as follows:

$$Q_i^\sigma(y^i \mid s^i, x^i) = \sum_{\{(\omega, x^{-i}): f_i(\omega, x^i, x^{-i}) = y^i\}} \sum_{s^{-i}} \prod_{j \neq i} \sigma^j(x^j \mid s^j)p(\omega, s^{-i} \mid s^i).$$

The objective distribution represents the true distribution over consequences, conditional on a player’s own action and signal, given the objective game and a strategy profile followed by the players.

**Distance to true model.** In equilibrium, we will require players’ beliefs to put
probability one on the set of subjective distributions over consequences that are “closest” to the objective distribution. In order to describe the right notion of “closest”, we need some additional definitions. The following function, which we call the weighted Kullback-Leibler divergence (wKLD) function of player $i$, is a weighted version of the standard Kullback-Leibler divergence in statistics (Kullback and Leibler, 1951). It represents a non-symmetric measure between the objective distribution over $i$’s consequences given a strategy profile $\sigma \in \Sigma$ and the distribution as parameterized by $\theta^i \in \Theta^i$,

$$K^i(\sigma, \theta^i) = \sum_{(s^i, x^i) \in S^i \times X^i} E_{Q^i_{\sigma}}(s^i, x^i) \left[ \ln \frac{Q^i_{\sigma}(Y^i | s^i, x^i)}{Q^i_{\theta^i}(Y^i | s^i, x^i)} \right] \sigma^i(x^i | s^i) p^i_{s^i}(s^i),$$

where we use the convention that $-\ln 0 = \infty$ and $0 \ln 0 = 0$. The set of closest parameters of player $i$ given $\sigma$ is the set

$$\Theta^i(\sigma) \equiv \arg \min_{\theta^i \in \Theta^i} K^i(\sigma, \theta^i).$$

The interpretation is that $\Theta^i(\sigma) \subset \Theta^i$ is the set of parameters of the world that player $i$ believes to be possible after observing feedback consistent with strategy profile $\sigma$.

**Remark 1.** (a) The use of the Kullback-Leibler divergence to measure distance is not an arbitrary assumption. We show in Section 4 that this is the right notion of distance when players are Bayesian. (b) Because the wKLD function is weighted by a player’s own strategy, it will in general place no restrictions on beliefs about outcomes that only arise following out-of-equilibrium actions.

The next result collects some useful properties of the wKLD. The proof is non-standard and relies on Assumption 1 because the wKLD function $K^i$ is not jointly continuous.

**Lemma 1.** (i) For all $\sigma \in \Sigma$, $\theta^i \in \Theta^i$, and $i \in I$, $K^i(\sigma, \theta^i) \geq 0$, with equality holding if and only if $Q^i_{\theta^i}(\cdot | s^i, x^i) = Q^i_{\sigma}(\cdot | s^i, x^i)$ for all $(s^i, x^i)$ such that $\sigma^i(x^i | s^i) > 0$. (ii) For every $i \in I$, $\Theta^i(\cdot)$ is non-empty, upper hemicontinuous, and compact valued.

**Proof.** See the Appendix. \(\Box\)

\(^{16}\)The notation $E_Q$ denotes expectation with respect to the probability measure $Q$. 

12
Optimality. In equilibrium, we will require each player to choose a strategy that is optimal given her beliefs. A strategy $\sigma^i$ for player $i$ is \textbf{optimal} given $\mu^i \in \Delta(\Theta^i)$ if $\sigma^i(x^i \mid s^i) > 0$ implies that

$$x^i \in \arg \max_{x^i \in X^i} E_{Q^i_{\mu^i}(\cdot \mid s^i, x^i)} [\pi^i(x^i, y^i)]$$

(6)

where, for all $i$ and $(y^i, s^i, x^i)$,

$$Q^i_{\mu^i}(y^i \mid s^i, x^i) = \int_{\Theta^i} Q^i_{\mu^i}(y^i \mid s^i, x^i) \mu^i(d\theta^i)$$

is the distribution over consequences of player $i$ induced by $\mu^i$.

Definition of equilibrium. We propose the following solution concept.

\textbf{Definition 1.} A strategy profile $\sigma$ is a Berk-Nash equilibrium of game $G$ if, for all players $i \in I$, there exists $\mu^i \in \Delta(\Theta^i)$ such that

(i) $\sigma^i$ is optimal given $\mu^i$, and

(ii) $\mu^i \in \Delta(\Theta^i(\sigma))$.

Definition 1 places two types of restrictions on equilibrium behavior: (i) optimization given beliefs, and (ii) endogenous restrictions on beliefs.\textsuperscript{17} For comparison, notice that the definition of a Nash equilibrium is identical to Definition 1 except that condition (ii) is replaced with the condition that $\bar{Q}^i_{\mu^i} = Q^i_{\sigma}$; in other words, players must have correct beliefs in a Nash equilibrium.

Existence of equilibrium. The standard proof of existence of Nash equilibrium cannot be used to show existence of a Berk-Nash equilibrium because the corresponding version of a best response correspondence is not necessarily convex valued.\textsuperscript{18} To prove existence, we first perturb the payoffs of the game and establish that equilibrium exists in the perturbed game. We then consider a sequence of equilibria of perturbed

\textsuperscript{17}Our setting is restricted to unitary beliefs (cf., Fudenberg and Levine, 1993a), which requires the same belief to rationalize each action in the support of a strategy. This is the standard assumption when there is one player in each role, as opposed to a setting where there is a population of players in each player role (e.g., Fudenberg and Levine, 1993b).

\textsuperscript{18}For example, fix $\sigma$ and suppose that $\Theta^i(\sigma) = \{\theta^i_1, \theta^i_2\}$. Then it is possible that $\sigma^i_1$ is optimal if player $i$ puts probability 1 on $\theta^i_1$ and that $\sigma^i_2$ is optimal if she puts probability 1 on $\theta^i_2$, but that a convex combination of $\sigma^i_1$ and $\sigma^i_2$ is not optimal for any belief satisfying $\mu^i \in \Delta(\Theta^i(\sigma))$. 

13
games where the perturbations go to zero and establish that the limit is a Berk-Nash equilibrium of the (unperturbed) game. The part of the proof that is not standard is establishing existence of equilibrium in the perturbed game. The perturbed best response correspondence is still not necessarily convex valued. Our approach is to characterize equilibrium as a fixed point of a belief correspondence and show that this correspondence satisfies the requirements of a generalized version of Kakutani’s fixed point theorem.

**Theorem 1.** Every game has at least one Berk-Nash equilibrium.

*Proof.* See the Appendix.

2.4 Examples: Finding a Berk-Nash equilibrium

We illustrate the definition by finding the set of Berk-Nash equilibrium for some of the examples in Section 2.2.

**Example 3.1, continued from pg. 8. Monopolist with unknown demand.** Suppose that the monopolist can choose a price of 2 or 10, i.e., $X = \{2, 10\}$ and that the true demand function is given by equation (1), with $\theta^0 = (a^0, b^0) = (40, 5)$. Figure 1 shows the true demand parameter and the set of parameters $\Theta = [12, 32] \times [1, 3]$ that the monopolist considers possible. In particular, $\theta^0 \notin \Theta$ and, therefore, we say that the monopolist has a *misspecified* model. The dashed line in Figure 1 depicts optimal behavior: price 10 is uniquely optimal to the left, price 2 is uniquely optimal to the right, and the monopolist is indifferent for parameters on the dashed line.

Because this is a single-agent problem, the objective distribution does not depend on the strategy $\sigma = (\sigma(2), \sigma(10))$; we denote it by $Q_{\theta^0}(\cdot \mid x)$ and note that it is a normal distribution with mean $\mu_{\theta^0}(x) = a^0 - b^0 x$ and unit variance. Similarly, for $\theta = (a, b)$, $Q_{\theta}(\cdot \mid x)$ is a normal distribution with mean $\mu_{\theta}(x)$ and unit variance. It

---

19The idea of perturbations and the strategy of the existence proof date back to Harsanyi (1973); Selten (1975) and Kreps and Wilson (1982) also used these ideas to prove existence of perfect and sequential equilibrium, respectively.
follows from equation (5) that

\[ K(\sigma, \theta) = \sum_{x \in \{2, 10\}} \sigma(x) E_{Q(\cdot|x)} \frac{1}{2} \left( (Y - \mu_\theta(x))^2 - (Y - \mu_{\theta_0}(x))^2 \right) \]

\[ = \sum_{x \in \{2, 10\}} \sigma(x) \frac{1}{2} (\mu_{\theta_0}(x) - \mu_\theta(x))^2. \]  

(7)

The derivatives are

\[ \frac{\partial K(\sigma, \theta)}{\partial a} = \sum_{x \in \{2, 10\}} \sigma(x) (\mu_{\theta_0}(x) - \mu_\theta(x)) \]  

(8)

\[ \frac{\partial K(\sigma, \theta)}{\partial b} = \sum_{x \in \{2, 10\}} \sigma(x) (\mu_{\theta_0}(x) - \mu_\theta(x)) x. \]  

(9)

Consider first pure strategies. If \( \sigma = (0, 1) \) (i.e., the price is \( x = 10 \)), the first order condition becomes \( \mu_{\theta_0}(10) = \mu_\theta(10) \). Any parameter \( \theta \in \Theta \) such that the mean of \( Y \) under \( \theta \) and \( x = 10 \) is equal to the true mean under \( x = 10 \) minimizes equation (7). In Figure 1, these are the parameters that lie on the line with slope 10 \((a^0 - b^010 = a - b10)\) and belong to \( \Theta \); i.e., the line segment \( AB \). Thus, if the monopolist were to choose \( \sigma = (0, 1) \) in equilibrium, then its beliefs must have support in the segment \( AB \). But this segment lies to the right of the dashed line, where it is not optimal to set a price of 10. Thus, \( \sigma = (0, 1) \) is not an equilibrium. A similar argument establishes that \( \sigma = (1, 0) \) is not an equilibrium: If it were, then the monopolist would believe that the parameter is \( D \). But \( D \) lies to the left of the dashed line, where it is not optimal to set a price of 2.

Finally, consider a totally mixed strategy \( \sigma \). Because expressions (8) and (9) cannot be simultaneously equal to zero, the parameter that minimizes \( K(\sigma, \theta) \) lies on the boundary of \( \Theta \). In fact, a bit of algebra shows that, for each totally mixed \( \sigma \), there is a unique minimizer \( \theta_\sigma = (a_\sigma, b_\sigma) \) characterized as follows. If \( \sigma(2) \leq 3/4 \), the minimizer is in the segment \( BC: b_\sigma = 3 \text{ and } a_\sigma = 16\sigma(2) + 20 \) is such that equation (8) is zero. If \( \sigma(2) \in [3/4, 15/16] \), then \( \theta_\sigma = C \) is the northeast vertex of \( \Theta \). Finally, if \( \sigma(2) > 15/16 \), the minimizer is in the segment \( CD: a_\sigma = 32 \text{ and } b_\sigma = (420 - 416\sigma(2))/(100 - 96\sigma(2)) \) is such that equation (9) is zero. Because the monopolist is mixing, optimality requires that the equilibrium belief \( \theta_\sigma \) lies on the dashed line. Thus, \( \sigma^\ast = (23/24, 1/24) \) is the unique Berk-Nash equilibrium, and the
Figure 1: Monopolist with misspecified demand function.
supporting belief \( \theta_{\sigma^*} = (32, 8/3) \) is given by the intersection of the dashed line and line segment \( CD \). Note that, in this example, it is not the case that the equilibrium belief about the mean of \( Y \) is correct. Thus, an approach that had focused on fitting the mean, rather than minimizing \( K \), would have led to the wrong conclusion.\(^{20}\)

**Example 3.2, continued from pg. 8.** Non-linear pricing. For any pure strategy \( x \) and parameter \( \theta \),

\[
K(x, \theta) = E_{\tau, z} \left( \ln \frac{Q(\tau \mid z)p(z \mid x)}{Q_{\theta}(\tau \mid z)p(z \mid x)} \mid X = x \right) \\
= -\frac{1}{2}E_z ((\tau(z) - \theta z)^2 \mid X = x) + C,
\]

where \( C \) is a constant that does not depend on \( \theta \). It is straightforward to check that \( \theta(x) = E ((x + \omega)\tau(x + \omega)) / E(x + \omega)^2 \) is the unique parameter that minimizes \( K(x, \theta) \). In particular, it would be a mistake to conclude that this naive agent believes that the constant marginal tax rate is given by the expected average tax rate, i.e., \( \theta(x) \neq E(\tau(x + \omega)/(x + \omega)) \)—this conjecture would be correct, however, if the agent had the following “random-coefficient” misspecified subjective model: \( \tau = (\theta + \varepsilon)z \).

A strategy \( x^* \) is a Berk-Nash equilibrium if and only if \( x = x^* \) maximizes \((1 - \theta(x^*)) x - c(x) \). In contrast, \( x^{opt} \) is the optimal strategy if and only if \( x = x^{opt} \) maximizes \( x - E\tau(x + \omega) - c(x) \). For example, suppose that \( \tau(z) = z^2 \) is progressive, \( c(\cdot) \) is increasing and strictly convex, and \( \omega \) follows a normal distribution with zero mean and variance \( \sigma^2 \). Then there is a unique Berk-Nash equilibrium and unique optimal strategy. Moreover, there is a threshold variance \( \bar{\sigma}^2 \) of the error term such that the Berk-Nash equilibrium strategy is lower than the optimal one for all higher variances and higher than optimal for all lower variances. Thus, despite a progressive tax schedule, naiveté can either decrease or increase the incentives to put effort. \( \square \)

**Example 3.2, continued from pg. 8.** Misspecified market structure. For convenience, we focus on the case where prices and costs are chosen from a compact interval and players follow pure strategies, where \( \sigma^i(s^i) \) is the price chosen by firm \( i \)

\(^{20}\)The example also illustrates the importance of allowing for mixed strategies for existence of Berk-Nash equilibrium, even in single-agent settings; see Esponda and Pouzo (2011) for a related example in a voting context.
given cost $s^i$. Optimality implies that each firm $i$ follows strategy

$$
\sigma^i(s^i) = \left[\gamma^i(\sigma)/(\gamma^i(\sigma) - 1)\right] s^i.
$$

(10)

where $\gamma^i(\sigma)$ is firm $i$’s equilibrium belief under strategy profile $\sigma$. Since the error term is normally distributed, the minimizer of the wKLD function is given by the ordinary least squares estimand of equation (2). Thus, for all $\sigma = (\sigma^1, \sigma^2)$,

$$
\gamma^i(\sigma) = \frac{-Cov(\ln \sigma^i(S^i), \alpha^* + \beta^* \ln \sigma^j(S^j) - \gamma^* \ln \sigma^i(S^i) + \varepsilon^i)}{Var(\ln \sigma^i(S^i))} = \gamma^* - \beta^* \frac{Cov(\ln \sigma^i(S^i), \ln \sigma^j(S^j))}{Var(\ln \sigma^i(S^i))}.
$$

(11)

By replacing (10) into (11), we obtain that, in equilibrium, $\gamma^i(\sigma)$ is independent of $\sigma$ and given by

$$
\gamma^i = \gamma^* - \beta^* \frac{Cov(\ln S^i, \ln S^j)}{Var(\ln S^i)}.
$$

(12)

Thus, there is a unique equilibrium and it is supported by belief $\gamma^i$ for firm $i$.\footnote{This is true as long as we make an assumption on the primitives that makes $\gamma^i > 1$.}

Moreover, (12) shows that firms estimate demand elasticity with a bias that depends on the sign of $\beta^* Cov(\ln S^i, \ln S^j)$. For example, suppose that $\beta^* > 0$, so that the products are substitutes, and that $Cov(\ln S^i, \ln S^j) > 0$. Then firms believe that demand is less elastic compared to the true elasticity. The intuition is that, when a firm chooses a higher price, it is because its costs are higher. But then the competitor’s cost is also likely to be higher, so the other firm is also likely to choose a higher price. Because products are substitutes, the increase in the price of the other firm mitigates the fall in demand due to the increase in own price. This under-estimation of elasticity leads firms to set higher prices compared to the Nash equilibrium, which is unique and given by the dominant strategy $\sigma^i(s^i) = [\gamma^*/(\gamma^* - 1)] s^i$. Note that disregarding the competition leads to biased behavior in the Berk-Nash equilibrium, even though actions are strategically independent and the (correct) best response does not depend on the competitor’s choice. □

Example 3.4, continued from pg. 9. Regression to the mean. A strategy maps initial performances to actions and it is straightforward to check that optimal strategies are characterized by a cutoff. Thus, we let $\sigma \in \mathbb{R}$ represent the strategy
where the instructor praises initial performances that are above $\sigma$ and criticizes the remaining performances. The wKLD function is

$$K(\sigma, \theta) = \int_{-\infty}^{\sigma} E_{Q(\cdot | s_1, C)} \left[ \ln \frac{Q(Y | s_1, C)}{Q_0(Y | s_1, C)} \right] \varphi(s_1) ds_1 + \int_{\sigma}^{\infty} E_{Q(\cdot | s_1, P)} \left[ \ln \frac{Q(Y | s_1, P)}{Q_0(Y | s_1, P)} \right] \varphi(s_1) ds_1$$

$$= \int_{-\infty}^{\sigma} E \left[ \ln \frac{\varphi(S_2)}{\varphi(S_2 - (\theta_C + s_1))} \right] \varphi(s_1) ds_1 + \int_{\sigma}^{\infty} E \left[ \ln \frac{\varphi(S_2)}{\varphi(S_2 - (\theta_P + s_1))} \right] \varphi(s_1) ds_1,$$

where $\varphi$ is the density of the standard normal distribution and expectations are with respect to the true distribution. It is straightforward to show that, for each $\sigma$, the unique parameter vector that minimizes $K(\sigma, \cdot)$ is

$$\theta_C(\sigma) = E[S_2 - S_1 | S_1 < \sigma]$$

$$= 0 - E[S_1 | S_1 < \sigma] > 0$$

and, similarly, $\theta_P(\sigma) = 0 - E[S_1 | S_1 > \sigma] < 0$. The intuition is that the instructor is critical for performances below a threshold and, therefore, the mean performance conditional on a student being criticized is lower than the unconditional mean performance; thus, a student who is criticized delivers a better next performance in expectation. Similarly, a student who is praised delivers a worse next performance in expectation.

Therefore, if the instructor follows strategy cutoff $\sigma$, she believes that, after observing initial performance $s_1 > 0$, her expected payoff is $s_1 + \theta_C(\sigma) - \kappa s_1$ if she criticizes and $s_1 + \theta_P(\sigma)$ if she praises. By optimality, the cutoff makes her indifferent between praising and criticizing. Thus, $\sigma^* > 0$ is an equilibrium cutoff if and only if

$$\sigma^* = \frac{1}{\kappa} (\theta_C(\sigma^*) - \theta_P(\sigma^*)) > 0.$$ 

Similar steps establish that there is no equilibrium with $\sigma^* \leq 0$. Thus, the instructor is excessively critical in equilibrium because she incorrectly believes that criticizing a student improves performance and that praising a student worsens it. Moreover, as the reputation cost $\kappa \to 0$, meaning that instructors care only about performance and not lying, $\sigma^* \to \infty$: instructors only criticize.□
3 Relationship to other solution concepts

3.1 Properties of games

We introduce several properties of games that are used in the sequel to relate Berknash equilibrium to other solution concepts and to provide a learning foundation for equilibrium.

**CORRECTLY-SPECIFIED GAMES.** In Bayesian statistics, a model is correctly specified if the support of the prior includes the true data generating process. The extension to single-agent decision problems is straightforward. For example, the monopoly problem in Example 3.1 is not correctly specified because $\theta^0 \notin \Theta$. In games, however, we must account for the fact that the objective distribution over consequences (i.e., the true model) depends on the strategy profile. We say that a game is correctly specified if players do not a priori rule out any of the objective distributions that might possibly arise in equilibrium.\(^{22}\)

**Definition 2.** A game is correctly specified given $\sigma$ if the subjective model satisfies the following condition for all $i \in I$: there exists $\theta^i \in \Theta^i$ such that $Q^i_{\theta^i} (y^i | s^i, x^i) = Q^i_{\theta} (y^i | s^i, x^i)$ for all $(s^i, x^i) \in S^i \times X^i$ and $y^i \in Y^i$; otherwise, the subjective model is misspecified given $\sigma$. A game is correctly specified if it is correctly specified for all $\sigma$; otherwise, it is misspecified.

**IDENTIFICATION.** From the point of view of the subject, what matters is identification of the distribution over consequences $Q_\theta$, not the parameter $\theta$. If the model is correctly specified, then the true $Q_{\theta^0}$ is trivially identified.\(^{23}\) Of course, this is not true if the model is misspecified, because the true distribution will never be learned. But we want a definition that captures the same spirit: If two distributions are equally a best fit (given the true distribution), then we want these two distributions to be identical; otherwise, we cannot identify which distribution is a best fit. The fact that

---

\(^{22}\)A more precise, though longer, terminology would say that a game is correctly specified in steady state. The reason is that, in the dynamic model, players believe that they face a stationary environment while in fact the environment might be non-stationary before reaching the steady state.

\(^{23}\)Similarly, the true parameter $\theta^0$ is identified if we impose the condition of identifiability of the parameter, i.e., $Q_\theta = Q_{\theta^0}$ implies that $\theta = \theta^0$. But, as mentioned earlier, it suffices to consider identification of distributions, not the parameter.
our agents also take actions introduces an additional nuance to the definition of identification. We can either ask for identification of the distribution over consequences for those actions that are taken by the agent (i.e., on the path of play) or for all actions (i.e., on and off the path). We refer to the former as weak identification (because it highlights the potential to identify the distribution over consequences, if all actions are taken) and to the latter as strong identification.

**Definition 3.** A game is **weakly identified given** \( \sigma \) if the subjective model satisfies the following condition for all \( i \in I \): if \( \theta^i_1, \theta^i_2 \in \Theta^i(\sigma) \), then \( Q^{i}_{\theta^i_1}(\cdot \mid s^i, x^i) = Q^{i}_{\theta^i_2}(\cdot \mid s^i, x^i) \) for all \((s^i, x^i) \in S^i \times X^i \) such that \( \sigma^i(x^i \mid s^i) > 0 \); if the condition is satisfied for all \((s^i, x^i) \in S^i \times X^i \), then we say that the game is **strongly identified given** \( \sigma \). A game is [weakly or strongly] identified if it is [weakly or strongly] identified for all \( \sigma \).

**Feedback.** The following property says that each player believes (either correctly or not) that the distribution over her consequences does not depend on her actions. Thus, players believe that they will get feedback about their payoff-relevant consequences irrespective of the action they take. This condition rules out those types of incorrect beliefs that are mainly due to lack of experimentation (but not due to other reasons, such as having a misspecified model).24

**Definition 4.** A game has **own-action independent feedback** if the subjective model satisfies: for all \( i \in I \) and all \( \theta^i \in \Theta^i \), \( Q^{i}_{\theta^i}(y^i \mid s^i, x^i) \) does not depend on \( x^i \), i.e., \( Q^{i}_{\theta^i}(y^i \mid s^i, x^i) = Q^{i}_{\theta^i}(y^i \mid s^i, \hat{x}^i) \) for all \( x^i, \hat{x}^i, s^i, y^i \).

**Remark 2.** For the special case of strategic subjective models defined in Section 2.1, an equivalent definition of own-action independent feedback is that, for each player \( i \), the feedback function does not depend on \( x^i \).

We conclude by establishing useful connections between these properties.

---

24 Arrow and Green (1973) defined a similar condition and restricted their setup to satisfy this condition.
Proposition 1. A game that is correctly specified is weakly identified. A game that is correctly specified and has own-action independent feedback is strongly identified.

Proof. Fix any $\sigma$ and $i \in I$. Because the game is correctly specified, there exists $\theta^*_i \in \Theta^i$ such that $Q^i_{\theta^*_i} = Q^i_\sigma$. In particular, $K^i(\theta^*_i, \sigma) = 0$. By Lemma ??(i), $K^i(\theta^*_i, \sigma) \geq 0$ for all $\theta^i \in \Theta^i$; therefore, $\theta^*_i \in \Theta^i(\sigma)$. Now consider any $\hat{\theta}^i \in \Theta^i(\sigma)$. Since $K^i(\theta^*_i, \sigma) = 0$, it must also be true that $K^i(\hat{\theta}^i, \sigma) = 0$. Lemma ??(i) then implies that $Q^i_{\hat{\theta}^i}(\cdot | s^i, x^i) = Q^i_{\theta^*_i}(\cdot | s^i, x^i)$ for all $(s^i, x^i)$ such that $\sigma^i(x^i | s^i) > 0$. Thus, a game that is correctly specified is also weakly identified. Now suppose, in addition, that the game has own-action independent feedback. Then it follows that $Q^i_{\hat{\theta}^i}$ does not depend on $x^i$ and, therefore, $Q^i_{\hat{\theta}^i}(\cdot | s^i, x^i) = Q^i_{\theta^*_i}(\cdot | s^i, x^i)$ for all $(s^i, x^i)$. Thus, the game is strongly identified.

As illustrated in the next subsection by the solution to Example 3.5, a correctly specified game might be strongly identified even if it does not satisfy own-action independent feedback.

3.2 Relationship to Nash equilibrium

The next result shows that Berk-Nash equilibrium is equivalent to Nash equilibrium when the game is both correctly specified and strongly identified.

Proposition 2. (i) Suppose that the game is correctly specified given $\sigma$ and that $\sigma$ is a Nash equilibrium of its objective game. Then $\sigma$ is a Berk-Nash equilibrium of the (objective and subjective) game.

(ii) Suppose that $\sigma$ is a Berk-Nash equilibrium of a game that is correctly specified and strongly identified given $\sigma$. Then $\sigma$ is a Nash equilibrium of the corresponding objective game.

Proof. (i) Let $\sigma$ be a Nash equilibrium and fix any $i \in I$. Then $\sigma^i$ is optimal given $Q^i_\sigma$. Because the game is correctly specified given $\sigma$, there exists $\theta^*_i \in \Theta^i$ such that $Q^i_{\theta^*_i} = Q^i_\sigma$ and, therefore, $\theta^*_i \in \Theta^i(\sigma)$. Then $\sigma^i$ is also optimal given $Q^i_{\theta^*_i}$ and $\theta^*_i \in \Theta^i(\sigma)$. Thus, $\sigma$ is a Berk-Nash equilibrium. (ii) Let $\sigma$ be a Berk-Nash equilibrium and fix any $i \in I$. Then $\sigma^i$ is optimal given $\bar{Q}^i_{\mu^i}$, where $\mu^i \in \Delta(\Theta^i(\sigma))$. Because the game is correctly specified given $\sigma$, there exists $\theta^*_i \in \Theta^i$ such that $Q^i_{\theta^*_i} = Q^i_\sigma$ and, therefore,
\( \theta^*_i \in \Theta^i(\sigma) \). Moreover, because the game is strongly identified given \( \sigma \), any \( \hat{\theta}^i \in \Theta^i(\sigma) \) satisfies \( Q^i_{\hat{\theta}} = Q^i_{\theta^*_i} = Q^i_{\sigma} \). Then \( \sigma^i \) is also optimal given \( Q^i_{\sigma} \). Thus, \( \sigma \) is a Nash equilibrium.

Example 3.5, continued from pg. 9. Classical monetary policy. We show that the game is correctly specified. Fix a strategy \( x^P \) for the public. Note that \( U = u^* - \lambda(x^G - x^P + \varepsilon_e) + \varepsilon_U \), whereas the government believes that \( U = \theta_1 - \theta_2(x^G + \varepsilon_e) + \varepsilon_U \). Thus, by choosing \( \theta^* = (\theta^*_1, \theta^*_2) \) such that \( \theta^*_1 = u^* + \lambda x^P \) and \( \theta^*_2 = \lambda \), it follows that the distribution over \( Y = (U, e) \) parameterized by \( \theta^* \) coincides with the objective distribution given \( x^P \). The game, however, does not satisfy own-action independent feedback; in particular, the distribution of \( U \) for any given \( \theta \) depends on the policy \( x^G \). Nevertheless, the model is strongly identified: since \( \text{Var}(\varepsilon_e) > 0 \), it is easy to check that \( \theta^* \) is the unique minimizer of the wKL divergence. Thus, Proposition 2 implies that Berk-Nash equilibrium is equivalent to Nash equilibrium. In particular, the equilibrium policies are the same whether or not the government realizes that unemployment is driven by surprise, not actual, inflation.\(^{25} \)

For games that are correctly specified but not strongly identified, Berk-Nash equilibrium is equivalent to self-confirming equilibrium (e.g., Dekel et al., 2004); it is well known that the equilibrium set in that case includes Nash but might also include non-Nash outcomes. The examples throughout the paper illustrate that a Nash equilibrium is not necessarily a Berk-Nash equilibrium when players have a misspecified view of the game.\(^{26} \)

3.3 Relationship to fully cursed and ABEE

An analogy-based game satisfies the following four properties:

(i) The information structure is finite and partitional: The state space \( \Omega \) has a finite number of elements; we denote the true distribution over \( \Omega \) by \( p_\Omega \). In addition,\(^{25} \) Sargent (1999) established this result for a government that does OLS-based learning.\(^{26} \) Arrow and Green (1973) impose a condition that requires the game to be correctly specified but only on the equilibrium path. This is equivalent to the requirement that the wKL divergence is zero at the equilibrium belief. In this case, equilibrium can also differ from Nash equilibrium, but this is no longer true if a small amount of experimentation leads to every action being played with positive probability in equilibrium.

\(^{25} \) Sargent (1999) established this result for a government that does OLS-based learning.

\(^{26} \) Arrow and Green (1973) impose a condition that requires the game to be correctly specified but only on the equilibrium path. This is equivalent to the requirement that the wKL divergence is zero at the equilibrium belief. In this case, equilibrium can also differ from Nash equilibrium, but this is no longer true if a small amount of experimentation leads to every action being played with positive probability in equilibrium.
for each $i$, there is a partition $S^i$ of $\Omega$, and the element of $S^i$ that contains $\omega$ is denoted by $s^i(\omega)$. In particular, the signal obtained by player $i$ when the state is $\omega$ is $s^i(\omega)$.

(ii) The subjective model is a strategic subjective model.

(iii) For each $i$, $f^i(\omega, x) = (\omega, x^{-i})$ for all $(\omega, x)$, i.e., each player gets perfect feedback about $(\omega, x^{-i})$.

Properties (i)-(iii) imply that the subjective model can be written as follows. For all $\theta^i \in \Theta^i$,

$$Q^i_{\theta^i}(\omega, x^{-i} | s^i(\omega'), x^i) = P^i_{\theta^i}(\omega, x^{-i} | s^i(\omega')),$$

which is equal to zero if $\omega \notin s^i(\omega')$.

(iv) For each $i$, there exists a partition of $\Omega$, denoted by $A^i$, and the element of $A^i$ that contains $\omega$ is denoted by $\alpha^i(\omega)$. In addition, every player $i$ believes that $x^{-i}$ and $\omega$ are independent conditional on the partition, i.e., $\pi^i$ is the set of all joint probability distributions over $\Omega \times X^{-i}$ that satisfy

$$P^i_{\theta^i}(\omega, x^{-i} | s^i(\omega')) = P^i_{\Omega, \theta^i}(\omega | s^i(\omega'))P^i_{X^{-i}, \theta^i}(x^{-i} | i(\omega)).$$

Two special cases are noteworthy. If $A^i = \{\Omega\}$ for all $i$, then each player believes that the actions of other players are independent of the state of the world. If $A^i = I^i$ for all $i$, then each player believes that the actions of other players are independent of the state, conditional on their own private information.

**Definition 5.** (Jehiel and Koessler, 2008) A strategy profile $\sigma$ is an analogy-based expectation equilibrium (ABEE) if for all $i \in I$, $\omega \in \Omega$, and $x^i$ such that $\sigma^i(x^i | s^i(\omega)) > 0$,

$$x^i \in \arg \max_{x^i \in X^i} \sum_{\omega' \in \Omega} p_{\Omega}(\omega' | s^i(\omega)) \sum_{x^{-i} \in X^{-i}} \tilde{\sigma}^{-i}(x^{-i} | \omega') \pi^i(\bar{x}^i, \omega', x^{-i}),$$

where $\tilde{\sigma}^{-i}(x^{-i} | \omega') = \sum_{\omega'' \in \Omega} p_{\Omega}(\omega'' | \alpha^i(\omega')) \prod_{j \neq i} \sigma^j(x^j | s^j(\omega''))$.

**Proposition 3.** In an analogy-based game, $\sigma$ is a Berk-Nash equilibrium if and only if it is an ABEE.

---

27This assumption is made only to facilitate comparison to the original definition of an analogy-based expectation equilibrium.
Proof. See the Appendix.

Proposition 3 shows the equivalence of Berk-Nash and ABEE for games that satisfy the four properties above. As mentioned by Jehiel and Koessler (2008), ABEE is equivalent to Eyster and Rabin’s (2005) fully cursed equilibrium in the special case where \( A^i = T^i \) for all \( i \).\(^{28}\) In particular, Proposition 3 clarifies the type of misspecification underlying these solution concepts.

Example 3.6, continued from pg. 10. Trade with adverse selection. As a benchmark, the Nash equilibrium (NE) price maximizes

\[
\Pi^{NE}(x) = \Pr(A \leq x) \left( E[V \mid A \leq x] - x \right).
\]

In Online Appendix D, we show that \( x^* \) is a Berk-Nash equilibrium price if and only if \( x = x^* \) maximizes an equilibrium belief function \( \Pi(x, x^*) \) which represents the belief about expected profit from choosing any price \( x \) under a steady-state \( x^* \). The function \( \Pi \) depends on the feedback/misspecification assumptions. With perfect feedback and subjective model \( \Delta(A) \times \Delta(V) \), the buyer learns the true marginal distributions of \( A \) and \( V \) and believes the joint distribution is given by the product of the marginal distributions. Therefore, the buyer’s equilibrium belief function is

\[
\Pi^{CE}(x) = \Pr(A \leq x) \left( E[V] - x \right),
\]

and, in particular, does not depend on the equilibrium \( x^* \). The equilibrium in this case coincides with a fully cursed equilibrium.

With partial feedback and subjective model \( \Delta(A) \times \Delta(V) \), the price offered by the buyer affects the sample of valuations that she observes. Also, the buyer does not realize that this selected sample would change if she were to change her price. Suppose that the buyer’s behavior has stabilized to some price \( x^* \). Then, the buyer’s equilibrium belief function is

\[
\Pi^{BE}(x, x^*) = \Pr(A \leq x) \left( E[V \mid A \leq x^*] - x \right).
\]

\(^{28}\)For experimental evidence of this type of naiveté, see the review by Kagel and Levin (2002) and the recent work by Charness and Levin (2009), Ivanov et al. (2010), and Esponda and Vespa (2013).
The equilibrium in this case coincides with a naive behavioral equilibrium (Esponda, 2008).

Finally, suppose that there is perfect feedback and consider a more general mis-specified model. There is a partition of the set \( V \) into \( k \) “analogy classes” \( (V_j)_{j=1,...,k} \), where \( \cup_j V_j = V \) and \( V_i \cap V_j = 0 \) for all \( i \neq j \). The buyer believes that \((A,V)\) are independent conditional on \( V \in V_i \), for each \( i = 1,...,k \). The parameter space is \( \Theta_A = \times_j \Delta(A) \times \Delta(V) \), where, for a parameter \( \theta = (\theta_1,\ldots,\theta_k,\theta_V) \in \Theta_A \), \( \theta_V \) parameterizes the marginal distribution over \( V \) and, for each \( j = 1,...,k \), \( \theta_j \in \Delta(A) \) parameterizes the distribution over \( A \) conditional on \( V \in V_j \). In this game, beliefs are as in a cursed equilibrium conditional on each analogy class, and so the equilibrium belief function is\(^{29}\)

\[ \Pi^{ABEE}(x) = \sum_{j=1}^{k} \Pr(V \in V_j) \{ \Pr(A \leq x \mid V \in V_j) (E[V \mid V \in V_j] - x) \}. \] (15)

The equilibrium in this case coincides with the analogy-based expectation equilibrium. In the Online Appendix, we also characterize this case under partial feedback. □

4 Equilibrium foundation

In this section, we provide a learning foundation for equilibrium. We follow Fudenberg and Kreps (1993) in considering games with (slightly) perturbed payoffs because, as they highlight in the context of providing a learning foundation for mixed-strategy Nash equilibrium, behavior need not be continuous in beliefs in the unperturbed game. Thus, even if beliefs were to converge, behavior would not necessarily settle down in the unperturbed game. Perturbations guarantee that if beliefs converge, then behavior also converges.

4.1 Perturbed game

A perturbation structure is a tuple \( \mathcal{P} = \langle \Xi, P_\xi \rangle \), where: \( \Xi = \times_{i \in I} \Xi^i \) and \( \Xi^i \subseteq \mathbb{R}^{#X^i} \) is a set of payoff perturbations for each action of player \( i \); \( P_\xi = (P_{\xi^i})_{i \in I} \), where \( P_{\xi^i} \in \Delta(\Xi^i) \) is a distribution over payoff perturbations of player \( i \) that is absolutely

---

\(^{29}\)Note the well known fact that analogy-based expectation equilibrium with a single analogy class is equivalent to (fully) cursed equilibrium; these two solution concepts were developed independently of each other.
continuous with respect to the Lebesgue measure, satisfies \( \int_{\Xi} ||\xi_i|| P_\xi(d\xi_i) < \infty \), and is independent from the perturbations of other players. A **perturbed game** \( G^P = \langle G, \mathcal{P} \rangle \) is composed of a game \( G \) and a perturbation structure \( \mathcal{P} \).

The timing of a perturbed game \( G^P \) coincides with the timing of its corresponding (unperturbed) game \( G \), except for two modifications. First, before taking an action, each player not only observes a signal \( s^i \) but now she also privately observes a vector of own payoff perturbations \( \xi_i \in \Xi_i \), where \( \xi_i(x^i) \) denotes the perturbation corresponding to action \( x^i \). Second, her payoff given action \( x^i \) and consequence \( y^i \) is \( \pi^i(x^i, y^i) + \xi^i(x^i) \).

A strategy \( \sigma^i \) for player \( i \) is **optimal in the perturbed game** given \( \mu_i \in \Delta(\Theta_i) \) if, for all \((s^i, x^i) \in S^i \times X^i, \sigma^i(x^i | s^i) = P_\xi(\xi_i : x^i \in \Psi_i(\mu_i, s^i, \xi)) \), where

\[
\Psi_i(\mu_i, s^i, \xi) \equiv \arg \max_{x^i \in X^i} E_{\hat{Q}^i_{\mu_i}(s^i, x^i)} [\pi^i(x^i, Y^i)] + \xi^i(x^i).
\]

In other words, if \( \sigma^i \) is an optimal strategy, then \( \sigma^i(x^i | s^i) \) is the probability that \( x^i \) is optimal when the state is \( s^i \) and the perturbation is \( \xi_i \), taken over all possible realizations of \( \xi_i \).

The definition of equilibrium of a perturbed game \( G^P \) is analogous to Definition 1, with the obvious difference that optimality must be required with respect to the perturbed game.

### 4.2 Learning foundation

We fix a perturbed game \( G^P \) and consider a setting where players repeatedly play the corresponding objective game at each moment in time \( t = 0, 1, 2, ..., \), and where the time-\( t \) state and signals, \((\omega_t, s_t)\), are independently drawn from the same distribution \( p \) every period. In addition, each player \( i \) has a prior \( \mu^i_0 \) with full support over her (finite-dimensional) parameter space, \( \Theta^i \). At the end of each period \( t \), each player uses Bayes’ rule and the information obtained in that period (her own signal, action, and consequence) to update her beliefs. Players believe that they face a stationary environment and myopically maximize the current period’s expected payoff. In the Online Appendix, we extend the results to non-myopic players who experiment optimally. Finally, we assume for simplicity that players know the distribution of their

\[\text{Note:}\] We restrict attention to finite dimensional parameter spaces because, otherwise, Bayesian updating need not converge to the truth for most priors and parameter values even in correctly specified statistical settings (Freedman (1963), Diaconis and Freedman (1986)).
own payoff perturbations.

Let \( B^i : \Delta(\Theta^i) \times S^i \times X^i \times Y^i \to \Delta(\Theta^i) \) denote the Bayesian operator of player \( i \):

\[
B^i(\mu^i, s^i, x^i, y^i)(A) = \frac{\int_A Q^{i'}(y^i | s^i, x^i)\mu^i(d\theta)}{\int_{\Theta^i} Q^{i'}(y^i | s^i, x^i)\mu^i(d\theta)}.
\]

Without loss of generality, we restrict behavior at time \( t \) to depend on the belief at time \( t \) and the new information received at time \( t \), i.e., the signal and payoff perturbation.

**Definition 6.** A policy of player \( i \) is a sequence of functions \( \phi^i = (\phi^i_t)_{t \in I} \), where \( \phi^i_t : \Delta(\Theta^i) \times S^i \times X^i \to X^i \). A policy \( \phi^i \) is optimal if \( \phi^i_t \in \Psi^i \) for all \( t \). A policy profile \( \phi = (\phi^i)_{i \in I} \) is optimal if \( \phi^i \) is optimal for all \( i \in I \).

Let \( H = (S \times \Xi \times \text{graph}(\times_{i \in I} f^i(\Omega, :)))^\infty \) denote the set of observable histories, where any history \( h = (s_0, \xi_0, x_0, y_0, \ldots, s_t, \xi_t, x_t, y_t \ldots) \in H \) must satisfy the feasibility restriction that \( y_t \in \times_{i \in I} f^i(\Omega, x_i) \) for all \( t \). Let \( P^{\mu_0, \phi} \) denote the (objective) probability distribution over \( H \) that is induced by the primitives of the game, the priors \( \mu_0 = (\mu^i_0)_{i \in I} \)—which partly determine the initial actions—and the policy profiles \( \phi = (\phi^i)_{i \in I} \). Let \( (\mu^i_t)_{t \geq 1} \) denote the sequence of beliefs \( \mu^i_t : H \to \times_{i \in I} \Delta(\Theta^i) \) such that, for all \( t \geq 1 \) and all \( i \in I \), \( \mu^i_t \) is the posterior at time \( t \) defined recursively by \( \mu^i_t(h) = B^i(\mu^i_{t-1}(h), s^i_{t-1}(h), x^i_{t-1}(h), y^i_{t-1}(h)) \) for all \( h \in H \), where \( s^i_{t-1}(h) \) is player \( i \)'s signal at \( t - 1 \) given history \( h \), and similarly for \( x^i_{t-1}(h) \) and \( y^i_{t-1}(h) \).

**Definition 7.** The sequence of intended strategy profiles given policy profile \( \phi = (\phi^i)_{i \in I} \) is the sequence \( (\sigma^i_t)_{t \in I} \) of random variables \( \sigma^i_t : H \to \times_{i \in I} \Delta(\Theta^i)^{S^i} \) such that, for all \( t \) and all \( i \in I \),

\[
\sigma^i_t(h)(x^i | s^i) = P^i_t (\xi^i : \phi^i_t(h), s^i, \xi^i = x^i).
\]

An intended strategy profile \( \sigma_t \) describes how each player would behave at time \( t \) for each possible signal; it is random because it depends on the players’ beliefs at time \( t \), \( \mu^i_t \), which in turn depend on the past history.
One reasonable criteria to claim that the players’ behavior stabilizes is that their intended behavior stabilizes with positive probability (cf. Fudenberg and Kreps, 1993).

**Definition 8.** A strategy profile $\sigma$ is **stable** [or strongly stable] under policy profile $\phi$ if the sequence of intended strategies, $(\sigma_t)_t$, converges to $\sigma$ with positive probability [or with probability one], i.e.,

$$P^{\mu_0, \phi}\left(\lim_{t \to \infty} \|\sigma_t(h) - \sigma\| = 0 \right) > 0 \ [\text{or } = 1]$$

The next result establishes that, if behavior stabilizes to a strategy profile $\sigma$, then, for each player $i$, the support of the posterior beliefs converges to $\Theta^i(\sigma)$. The proof clarifies the origin of the wKLD function in the definition of Berk-Nash equilibrium. The Lemma extends results from the statistics of misspecified learning (Berk (1966), Bunke and Milhaud (1998)) by considering a setting where there is active learning (i.e., agents learn from data that is endogenously generated by their own actions). Three new issues arise with active learning: (i) Previous results need to be extended to the case of non-i.i.d. data; (ii) It is not obvious that steady-state beliefs can be characterized based on steady-state behavior, independently of the path of play (this is where Assumption 1 plays an important role; See Section 4.4 for an example); (iii) We allow KL divergence to be non-finite so that players can believe that other players follow pure strategies.\(^{31}\)

**Lemma 2.** Suppose that, for a policy profile $\phi$, the sequence of intended strategies, $(\sigma_t)_t$, converges to $\sigma$ for all histories in a set $\mathcal{H} \subseteq \mathbb{H}$ such that $P^{\mu_0, \phi}(\mathcal{H}) > 0$. Then, for all open sets $U^i \supseteq \Theta^i(\sigma)$,

$$\lim_{t \to \infty} \mu^i_t(U^i) = 1$$

$P^{\mu_0, \phi}$-a.s. in $\mathcal{H}$. 

**Proof.** It is sufficient to establish that $\lim_{t \to \infty} \int_{\Theta^i} d^i(\sigma, \theta^i) \mu^i_{t+1}(d\theta^i) = 0$ a.s. in $\mathcal{H}$,

\(^{31}\)For example, if player 1 believes that player 2 plays $A$ with probability $\theta$ and $B$ with $1 - \theta$, if we allow both $\theta = 0$ and $\theta = 1$ then the KL divergence will be infinity at $\theta = 1$ if player 2 plays $B$ with positive probability.
where \( d^i(\sigma, \theta^i) = \inf_{\hat{\theta} \in \Theta^i(\sigma)} \| \theta^i - \hat{\theta} \| \). Fix \( i \in I \) and \( h \in \mathbb{H} \). Then

\[
\int_{\Theta^i} d^i(\sigma, \theta^i) \mu^i_{t+1}(d\theta^i) = \frac{\int_{\Theta^i} d^i(\sigma, \theta^i) \prod_{\tau=1}^t Q^i_{\theta^i}(y^i_{\tau} | s^i_{\tau}, x^i_{\tau}) \mu^i_0(d\theta^i)}{\int_{\Theta^i} \prod_{\tau=1}^t Q^i_{\theta^i}(y^i_{\tau} | s^i_{\tau}, x^i_{\tau}) \mu^i_0(d\theta^i)} \equiv \frac{\int_{\Theta^i} d^i(\sigma, \theta^i) \prod_{\tau=1}^t Q^i_{\theta^i}(y^i_{\tau} | s^i_{\tau}, x^i_{\tau}) \mu^i_0(d\theta^i)}{\int_{\Theta^i} e^{tK^i(h, \theta^i)} \mu^i_0(d\theta^i)},
\]

where the first line is well-defined given the definition of a subjective model (i.e., properties (i) and (ii)), where the second line is well-defined because \( P^{\mu_0, \phi}(H) > 0 \) implies that all the terms we divide by are positive, and where we define \( K^i(h, \theta^i) = -\frac{1}{t} \sum_{\tau=1}^t \ln Q^i_{\theta^i}(y^i_{\tau} | s^i_{\tau}, x^i_{\tau}) \). For any \( \alpha > 0 \), define \( \Theta^i_\alpha(\sigma) \equiv \{ \theta^i \in \Theta^i : d^i(\sigma, \theta^i) < \alpha \} \). Then, for all \( \varepsilon > 0 \) and \( \eta > 0 \),

\[
\int_{\Theta^i} d^i(\sigma, \theta^i) \mu^i_{t+1}(d\theta^i) \leq \varepsilon + C \frac{A^i_t(h, \sigma, \varepsilon)}{B^i_t(h, \sigma, \eta)},
\]

where \( C \equiv \sup_{\theta_1, \theta_2 \in \Theta^i} \| \theta_1^i - \theta_2^i \| < \infty \) (because \( \Theta^i \) is bounded) and where

\[
A^i_t(h, \sigma, \varepsilon) = \int_{\Theta^i \setminus \Theta^i_\varepsilon(\sigma)} e^{tK^i(h, \theta^i)} \mu^i_0(d\theta^i)
\]

and

\[
B^i_t(h, \sigma, \eta) = \int_{\Theta^i_\eta(\sigma)} e^{tK^i(h, \theta^i)} \mu^i_0(d\theta^i).
\]

The proof concludes by showing that for every (sufficiently small) \( \varepsilon > 0 \), there exists \( \eta_\varepsilon > 0 \) such that \( \lim_{t \to \infty} A^i_t(h, \sigma, \varepsilon)/B^i_t(h, \sigma, \eta_\varepsilon) = 0 \). This result is achieved in several steps. First, for all \( \varepsilon > 0 \), define \( K^i_\varepsilon(\sigma) = \inf \{ K^i(\sigma, \theta^i) \mid \theta^i \in \Theta^i \setminus \Theta^i_\varepsilon(\sigma) \} \) and \( \alpha_\varepsilon = (K^i_\varepsilon(\sigma) - K^i_0(\sigma))/3 \). By continuity of \( K^i(\sigma, \cdot) \), there exists \( \bar{\varepsilon} \) and \( \bar{\alpha} \) such that \( 0 < \alpha_\varepsilon \leq \bar{\alpha} < \infty \) for all \( \varepsilon \leq \bar{\varepsilon} \). From now on, let \( \varepsilon \leq \bar{\varepsilon} \). It follows that

\[
K^i(\sigma, \theta^i) \geq K^i_\varepsilon(\sigma) > K^i_0(\sigma) + 2\alpha_\varepsilon \quad (18)
\]

\[\text{If, for some } \theta^i, Q^i_{\theta^i}(y^i_{\tau} | s^i_{\tau}, x^i_{\tau}) = 0 \text{ for some } \tau \in \{1, ..., t\}, \text{ then we define } K^i_t(h, \theta^i) = -\infty \text{ and exp } \{ tK^i_t(h, \theta^i) \} = 0. \]
for all $\theta^i$ such that $d^i(\sigma, \theta^i) \geq \varepsilon$. Also, by continuity of $K^i(\sigma, \cdot)$, there exists $\eta_{\varepsilon} > 0$ such that

$$K^i(\sigma, \theta^i) < K^i_0(\sigma) + \alpha\varepsilon/2$$  \hfill (19)

for all $\theta^i \in \Theta^i_{\eta_{\varepsilon}}$.

Second, let $\hat{\Theta}^i = \{\theta^i \in \Theta^i : Q^i_{\hat{\theta}^i}(y^i_t \mid s^i_t, x^i_t) > 0 \text{ for all } t\}$. Let $\hat{\Theta}^i_{\eta_{\varepsilon}}(\sigma) = \hat{\Theta}^i \cap \Theta^i_{\eta_{\varepsilon}}(\sigma)$. We now show that $\mu_0^i(\hat{\Theta}^i_{\eta_{\varepsilon}}(\sigma)) > 0$. By Lemma 1, $\Theta^i(\sigma)$ is nonempty. Pick any $\theta^i \in \Theta^i(\sigma)$. By Assumption 1, there exists $(\theta^i_n)_n$ in $\Theta^i$ such that $\lim_{n \to \infty} \theta^i_n = \theta^i$ and $Q^i_{\theta^i_n}(y^i_t \mid s^i_t, x^i_t) > 0$ for all $y^i \in f^i(\Omega, x^i, x^{-i})$ and all $(s^i, x^i) \in S^i \times X^i$. In particular, continuity of $Q^i$ implies that there exists $N$ and an open set $U$ such that $U \subseteq \hat{\Theta}^i_{\eta_{\varepsilon}}(\sigma)$ and $\theta^i_N \in U$. The assumption that $\mu_0$ has full support implies that $\mu_0(U) > 0$ and, therefore, $\mu_0^i(\hat{\Theta}^i_{\eta_{\varepsilon}}(\sigma)) > 0$.

Next, in Claim 2.1 in the Appendix we use a law of large numbers argument for non-iid random variables to establish that

$$\lim_{t \to \infty} K^i_t(h, \theta^i) = -K^i(\sigma, \theta^i)$$ \hfill (20)

for all $\theta^i \in \Theta^i$, a.s. in $\mathcal{H}$. Thus,

$$\liminf_{t \to \infty} B^i_t(h, \sigma, \eta_{\varepsilon})e^{t(K^i_0(\sigma)+\alpha\varepsilon/2)} \geq \liminf_{t \to \infty} \int_{\hat{\Theta}^i_{\eta_{\varepsilon}}(\sigma)} e^{t(K^i_0(\sigma)+\alpha\varepsilon/2+K^i_t(h, \theta^i))} \mu_0^i(d\theta^i)$$

$$\geq \int_{\hat{\Theta}^i_{\eta_{\varepsilon}}(\sigma)} e^{\lim_{t \to \infty} t(K^i_0(\sigma)+\alpha\varepsilon/2-K^i(\sigma, \theta^i))} \mu_0^i(d\theta^i)$$

$$= \infty,$$ \hfill (21)

a.s.-$\mathbf{P}^{\mu_0, \phi}$, where the first line follows because we do not integrate over the complement of $\hat{\Theta}^i$ and exp is a positive function, the second line follows from Fatou’s Lemma and (20), and the third line follows from (19) and the fact that $\mu_0^i(\Theta^i_{\eta_{\varepsilon}}(\sigma)) > 0$.

Next, we consider the term $A^i_t(h, \sigma, \varepsilon)$. For any $\xi > 0$, define $\Theta^i_{\sigma, \xi}$ to be the set such that $\theta^i \in \Theta^i_{\sigma, \xi}$ if and only if $Q^i_{\theta^i}(y^i_t \mid s^i_t, x^i_t) \geq \xi$ for all $(s^i, x^i, y^i)$ such that $Q^i_{\theta^i}(y^i | s^i, x^i)\sigma^i(x^i | s^i)p_{S^i}(s^i) > 0$. In Claim 2.2 in the Appendix, we show that there exists $\xi^* > 0$ and $T_{\xi^*}$ such that, for all $t \geq T_{\xi^*}$,

$$K^i_t(h, \theta^i) < -(K^i_0(\sigma) + (3/2)\alpha\varepsilon)$$ \hfill (22)
for all $\theta^i \notin \Theta_{\sigma,\xi}^i$. Also, in Claim 2.3 in the Appendix we show that, for all $\xi > 0$, there exists $\hat{T}_\xi$ such that, for all $t \geq \hat{T}_\xi$,

$$K^i_t(h, \theta^i) < -K^i(\sigma, \theta^i) + \alpha_\xi/2$$

for all $\theta^i \in \Theta_{\sigma,\xi}^i$. In particular, (18) implies that, for all $t \geq T_{\xi}$,

$$K^i_t(h, \theta^i) < -(K^i_0(\sigma) + (3/2)\alpha_\xi)$$

(23)

for all $\theta^i \in \Theta_{\sigma,\xi}^i \setminus \Theta_{\xi}^i(\sigma)$. Thus,

$$\lim_{t \to \infty} A^i_t(h, \sigma, \xi) e^{t(K^i_0(\sigma) + \alpha_\xi)} = \lim_{t \to \infty} \int_{\Theta \setminus \Theta^i(\sigma)} e^{t(K^i_0(\sigma) + \alpha_\xi + K^i_t(h, \theta^i))} \mu^i_0(d\theta^i)$$

$$\leq \mu^i_0(\Theta \setminus \Theta^i_\xi(\sigma)) \lim_{t \to \infty} e^{-t\alpha_\xi/2}$$

$$= 0,$$  

(24)

where the inequality follows from (22) and (23). Equations (21) and (24) imply that

$$\lim_{t \to \infty} A^i_t(h, \sigma, \xi)/B^i_t(h, \sigma, \eta_\xi) = 0 \text{ a.s.-P}$$

Lemma 2 only implies that the support of posteriors converges, but posteriors need not converge. We can always find, however, a subsequence of posteriors that converges. By continuity of behavior in beliefs and the assumption that players are myopic, the stable strategy profile must be statically optimal. Thus, we obtain the following characterization of the set of stable strategy profiles when players follow optimal policies.

**Theorem 2.** Suppose that a strategy profile $\sigma$ is stable under an optimal policy profile for a perturbed game. Then $\sigma$ is a Berk-Nash equilibrium of the game.

**Proof.** Let $\phi$ denote the optimal policy function under which $\sigma$ is stable. By Lemma 2, there exists $\mathcal{H} \subseteq \mathbb{H}$ with $\mathbf{P}^{\mu_0, \phi}(\mathcal{H}) > 0$ such that, for all $h \in \mathcal{H}$, $\lim_{t \to \infty} \sigma_t(h) = \sigma$ and $\lim_{t \to \infty} \mu^i_t(U^i) = 1$ for all $i \in I$ and all open sets $U^i \supseteq \Theta^i(\sigma)$; for the remainder of the proof, fix any $h \in \mathcal{H}$. For all $i \in I$, compactness of $\Delta(\Theta^i)$ implies the existence of a subsequence, which we denote as $(\mu^i_{(j)})$, such that $\mu^i_{(j)}$ converges (weakly) to $\mu^i_\infty$ (the limit could depend on $h$). We conclude by showing, for all $i \in I$:  

32
(i) $\mu^i_\infty \in \Delta(\Theta^i(\sigma))$: Suppose not, so that there exists $\hat{\theta}^i \in \text{supp}(\mu^i_\infty)$ such that $\hat{\theta}^i \notin \Theta^i(\sigma)$). Then, since $\Theta^i(\sigma)$ is closed (by Lemma ??(ii)), there exists an open set $U^i \supset \Theta^i(\sigma)$ with closure $\bar{U}^i$ such that $\hat{\theta}^i \notin \bar{U}^i$. Then $\mu^i_\infty(\bar{U}^i) < 1$, but this contradicts the fact that $\mu^i_\infty(\bar{U}^i) \geq \lim_{j \to \infty} \mu^i_{t(j)}(U^i) = 1$, where the first inequality holds because $\bar{U}^i$ is closed and $\mu^i_{t(j)}$ converges (weakly) to $\mu^i_\infty$.

(ii) $\sigma^i$ is optimal for the perturbed game given $\mu^i_\infty \in \Delta(\Theta^i)$:

$$\sigma^i(x^i \mid s^i) = \lim_{j \to \infty} \sigma^i_{t(j)}(h)(x^i \mid s^i)$$

$$= \lim_{j \to \infty} P_\xi \left( \xi^i : x^i \in \Psi^i(\mu^i_{t(j)}, s^i, \xi^i) \right)$$

$$= P_\xi \left( \xi^i : x^i \in \Psi^i(\mu^i_\infty, s^i, \xi^i) \right),$$

where the second line follows from optimality of $\phi^i$ and the (standard) fact that $\Psi^i$ is single-valued with respect to $\mu^i$, a.s.- $P_\xi$,33 and the third line follows from a standard continuity argument provided in Claim A.4 in the Appendix.

4.3 A converse result

Theorem 2 provides our main justification for focusing on Berk-Nash equilibria: any strategy profile that is not an equilibrium cannot represent the limiting behavior of optimizing players. Theorem 2, however, does not imply that behavior will stabilize. In fact, it is well known that there are cases where optimal behavior will not converge to Nash equilibrium, which is a special case of Berk-Nash equilibrium.34 Thus, some assumption needs to be relaxed in order to prove convergence for general games.

An insight due to Fudenberg and Kreps (1993) is that a converse for the case of Nash equilibrium can be obtained by relaxing optimality and allowing players to make vanishing optimization mistakes.

---

33 $\Psi^i$ is single-valued a.s.- $P_\xi$ because the set of $\xi^i$ such that $\#\Psi^i(\mu^i, s^i, \xi^i) > 1$ is of dimension lower than $\#X^i$ and, by absolute continuity of $P_\xi$, this set has measure zero.

34 Jordan (1993) shows that non-convergence is robust to the choice of initial conditions; Benaim and Hirsch (1999) replicate this finding for the perturbed version of Jordan’s game. In the game-theory literature, general global convergence results have only been obtained in special classes of games—e.g. zero-sum, potential, and supermodular games (Hofbauer and Sandholm, 2002).
Definition 9. A policy profile $\phi$ is **asymptotically optimal** if there exists a positive real-valued sequence $(\varepsilon_t)_t$ with $\lim_{t \to \infty} \varepsilon_t = 0$ such that, for all $i \in I$, all $(\mu^i, s^i, \xi^i) \in \Delta(\Theta^i) \times S^i \times \Xi^i$, all $t$, and all $x^i \in X^i$,

$$E \bar{Q}^i_{\mu^i} (|s^i, x^i_t|) [\pi^i(x^i_t, Y^i)] + \xi^i(x^i_t) \geq E \bar{Q}^i_{\mu^i} (|s^i, x^i|) [\pi^i(x^i, Y^i)] + \xi^i(x^i) - \varepsilon_t$$

where $x^i_t = \phi^i_t(\mu^i, s^i, \xi^i)$.

Fudenberg and Kreps’ (1993) insight is to construct a strategy profile with the property that players are somehow convinced early on that this is the right strategy to play, and they continue to play this strategy unless they have strong enough evidence to think otherwise. But, as they continue to play the strategy, they become increasingly convinced that it is the right thing to do. This idea, however, need not work in the case of Berk-Nash equilibrium. The reason is that, even if players always follow the same strategy, beliefs need not converge if the model is misspecified. The following example shows this point in a context without actions (or, equivalently, with fixed actions).

**Example (Berk, 1966).** An unbiased coin is tossed every period and the agent believes that the probability of heads is either 1/4 or 3/4, but not 1/2. The agent, who takes no actions in this example, observes the outcome of each coin toss and updates her (non-doctrinaire) prior. In this case, both 1/4 and 3/4 are equidistant to the true distribution 1/2, and it is straightforward to show that the agent’s beliefs do not settle down. □

In the coin toss example, there is a failure of weak identification (see Definition 3). We can show, however, that if the game is weakly identified, then Lemma 2 and Fudenberg and Kreps’ (1993) insight can be combined to obtain the following converse of Theorem 2.

**Theorem 3.** Suppose that $\sigma$ is an equilibrium of a perturbed game that is weakly identified given $\sigma$. Then there exists a profile of priors with full support and an asymptotically optimal policy profile $\phi$ such that $\sigma$ is strongly stable under $\phi$.

**Proof.** See the Online Appendix.\(^\text{35}\)

\(^{35}\)Note that the requirement that the priors have full support makes the statement non-trivial.
4.4 Discussion

Importance of Assumption 1. Consider the following example. A single agent chooses action \( x \in \{ A, B \} \) and obtains an outcome \( y \in \{ 0, 1 \} \). The agent’s model is parameterized by \( \theta = (\theta_A, \theta_B) \), where \( Q_\theta(y = 1 \mid A) = \theta_A \) and \( Q_\theta(y = 1 \mid B) = \theta_B \).

Suppose that the true model is \( \theta^0 = (1/4, 3/4) \). The agent, however, is misspecified and considers only \( \theta_1 = (0, 3/4) \) and \( \theta_2 = (1/4, 1/4) \) to be possible, i.e., \( \Theta = \{ \theta_1, \theta_2 \} \). In particular, Assumption 1(iii) fails for parameter \( \theta_1 \).\(^{36}\) Suppose that \( A \) is optimal for parameter \( \theta_1 \) and \( B \) is optimal for \( \theta_2 \) (further details about payoffs are not needed).

A Berk-Nash equilibrium does not exist in this example. If \( A \) is played with positive probability, then \( \theta_1 \) yields a wKL divergence of infinity (i.e., \( \theta_1 \) cannot rationalize \( y = 1 \) given \( A \)) and \( \theta_2 \) is the best fit. But then \( B \) is optimal. If pure strategy \( B \) is played, then \( \theta_1 \) is the best fit; but then \( A \) is optimal.

In addition, Lemma 2 fails: Suppose that the path of play converges to pure strategy \( B \). The best fit given \( B \) is \( \theta_1 \), but it is not necessarily true that the posterior converges weakly to a degenerate probability distribution on \( \theta_1 \). The reason is that it is possible that, along the path of play, the agent tried action \( A \) and observed \( y = 1 \), in which case the posterior would immediately assign probability 1 to \( \theta_2 \).

Forward-looking agents. In the dynamic model, we assumed for simplicity that all players are myopic and maximize current period’s payoffs given current beliefs. In the Online Appendix, we extend Theorem 2 to the case of non-myopic players who solve a dynamic optimization problem with beliefs as a state variable. A key fact used in the proof of Theorem 2 is that myopically optimal behavior is continuous in beliefs. Non-myopic optimal behavior is also continuous in beliefs, but the issue is that it may not coincide with myopic behavior in the steady state if players still have incentives to experiment. We prove the extension by requiring that the game is weakly identified, which guarantees that players have no incentives to experiment in steady state.

Stationarity of subjective models. An important assumption in our framework is that players act as if they were facing a stationary environment. But, except in single-agent settings, the environment may not be stationary if other players are learning and experimenting with different strategies. There a few reasons, however, why the assumption of stationary subjective models is sensible. The assumption is

\(^{36}\)Assumption 1 would be satisfied if, for example, the parameter values \( \theta = (\varepsilon, 3/4) \), for all \( 0 < \varepsilon \leq \bar{\varepsilon} \), were also included in \( \Theta \).
exactly correct in steady state and approximately correct near the steady state. Also, it is exactly correct in the construction we rely on to prove Theorem 3. Moreover, stationary models are a relatively simple and parsimonious way in which players can model their environment. Each player in our framework has a stationary model of other players’ actions, but does not have a model to explain why other players take certain actions. If, in contrast, a player had a non-stationary model of how other players make decisions, then the equilibrium concept would differ from Nash (or any related concept, such as Berk-Nash). For example, if one player correctly believes that all other players are like the players assumed in this paper, then she can influence their behavior and the right notion of steady state would be more akin to a Stackelberg equilibrium. Whether Nash (and Berk-Nash) or Stackelberg (and related concepts) are appropriate or not is likely to depend on the application. While our paper focuses on extending Nash to misspecified settings, it would be interesting to also do so for Stackelberg and related concepts.\footnote{Extending the Stackelberg approach to misspecified settings is challenging due to the well-known infinite regress problem that arises when a player is allowed to have models of other players’ reasoning.}

**LARGE POPULATION MODELS.** Our framework assumes that there is a fixed number of players but, by focusing on stationary subjective models, rules out aspects of “repeated games” where players attempt to influence each others’ play. An alternative setting, in which players have negligible incentives to influence each other’s play, is a model in which there is a population of a large number of agents in the role of each player \(i \in I\). We discuss some variants of population models that differ in the matching technology and feedback. The right variant of population model will depend on the specific application.\footnote{In some cases, it may be unrealistic to assume that players are able to observe the signals of previous generations, so some of these models might be better suited to cases with public, but not private, information.}

**Single pair model:** Each period a single pair of players is randomly selected from each of the \(i\) populations to play the game. At the end of the period, the signals, actions, and outcomes of their own population are revealed to everyone.\footnote{Alternatively, we can think of different incarnations of players born every period who are able to observe the history of previous generations.} Steady-state behavior in this case corresponds exactly to the notion of Berk-Nash equilibrium described in the paper.

**Random matching model:** Each period, all players are randomly matched and

---

\footnotetext[37]{Extending the Stackelberg approach to misspecified settings is challenging due to the well-known infinite regress problem that arises when a player is allowed to have models of other players’ reasoning.}

\footnotetext[38]{In some cases, it may be unrealistic to assume that players are able to observe the signals of previous generations, so some of these models might be better suited to cases with public, but not private, information.}

\footnotetext[39]{Alternatively, we can think of different incarnations of players born every period who are able to observe the history of previous generations.}
observe only feedback from their own match. We now modify the definition of Berk-Nash equilibrium to account for this random-matching setting. The idea is similar to Fudenberg and Levine’s (1993) definition of a heterogeneous self-confirming equilibrium. Now each agent in population \( i \) can have different experiences and, hence, have different beliefs and play different strategies in steady state.

For all \( i \in I \), define

\[
BR^i(\sigma^{-i}) = \{ \sigma^i : \sigma^i \text{ is optimal given } \mu^i \in \Delta(\Theta^i(\sigma^i, \sigma^{-i})) \}.
\]

Note that \( \sigma \) is a Berk-Nash equilibrium if and only if \( \sigma^i \in BR^i(\sigma^{-i}) \) for all \( i \in I \).

**Definition 10.** A strategy profile \( \sigma \) is a heterogeneous Berk-Nash equilibrium of game \( G \) if, for all \( i \in I \), \( \sigma^i \) is in the convex hull of \( BR^i(\sigma^{-i}) \).

Intuitively, a heterogeneous equilibrium strategy \( \sigma^i \) is the result of convex combinations of strategies that belong to \( BR^i(\sigma^{-i}) \), i.e., \( \{ \sigma^i_k : k \in \mathbb{K}^i \} \) where \( \mathbb{K}^i \) indexes the set of strategies in population \( i \). The idea is that, for each \( k \in \mathbb{K}^i \), there is a segment of the population \( i \) that follows strategy \( \sigma^i_k \) that is optimal given beliefs \( \mu^i_k \in \Delta(\Theta^i(\sigma^i_k, \sigma^{-i})) \), where \( \sigma^{-i} \) is the average strategy of agents in populations other than \( i \).\(^{40}\)

**Random-matching model with population feedback:** Each period all players are randomly matched; at the end of the period, each player in population \( i \) observes the signals, actions, and outcomes of their own population. Define

\[
BR^i(\sigma^i, \sigma^{-i}) = \{ \hat{\sigma}^i : \hat{\sigma}^i \text{ is optimal given } \mu^i \in \Delta(\Theta^i(\sigma^i, \sigma^{-i})) \}.
\]

**Definition 11.** A strategy profile \( \sigma \) is a heterogeneous Berk-Nash equilibrium with population feedback of game \( G \) if, for all \( i \in I \), \( \sigma^i \) is in the convex hull of \( BR^i(\sigma^i, \sigma^{-i}) \).

The main difference when players receive population feedback is that their beliefs no longer depend on their own strategies but rather on the aggregate population strategies.

\(^{40}\)Note that, unlike the case of self-confirming equilibrium, a definition where each action in the support of \( \sigma \) is supported by a (possibly different) belief would not be appropriate here. The reason is that \( BR^i(\sigma^{-i}) \) might contain only mixed, but not pure strategies (e.g., Example 1).
Equilibrium refinements. Theorems 2 and 3 provide a justification for our definition of equilibrium but leave open the possibility of refinements. One natural refinement is to require exact, not asymptotic, optimality, and to ask whether certain equilibria can be reached with positive probability. Another possible refinement is to follow Harsanyi (1973) and rule out those equilibria that are not regular in the sense that they might not be approachable by equilibrium sequences of perturbed games. All of these ideas have been extensively studied in related contexts (e.g., Benaim and Hirsch (1999), Fudenberg and Levine (1993b), Doraszelski and Escobar (2010)). Moreover, whether or not it is natural to pursue some of these refinements likely depends on the specific application under consideration.

Distance to true model. We showed that the assumption of Bayesian updating implies that the appropriate notion of “distance” in the definition of equilibrium is the (weighted) Kullback-Leibler divergence. It would be interesting to explore how other, non-Bayesian assumptions on the belief updating process yield other notions of distance.

Detecting misspecifications. In our setting, Bayesian agents have no reason to discover that they are misspecified. But, in practice, people who are aware of the possibility of misspecification might conduct tests to detect misspecification. These tests, which impose additional restrictions on beliefs, might provide a way to endogenize the types of misspecifications that agents can hold in equilibrium.\footnote{Such additional restrictions on beliefs are imposed, for example, by Arrow and Green (1973) and Esponda, 2008.}

5 Conclusion

We propose and provide a foundation for an equilibrium framework that allows players to have misspecified views of the game they are playing. By doing so, we highlight an implicit assumption in the concept of Nash equilibrium and considerably extend its domain of applicability. Our framework not only unifies an existing literature on bounded rationality and misspecified learning, but it also provides a systematic approach to studying certain aspects of bounded rationality, that, we hope, stimulates further developments in this area.
References


Appendix

Notation: Let $\mathcal{Z}^i = \{(s^i, x^i, y^i) \in \mathcal{S}^i \times \mathcal{X}^i \times \mathcal{Y}^i : y^i = f^i(\omega, x^i, x^{-i}), \ \omega \in \Omega, x^{-i} \in \mathcal{X}^{-i}\}$ be the set of feasible signals, actions, and consequences of player $i$. For each $z^i = (s^i, x^i, y^i) \in \mathcal{Z}^i$, define $\bar{P}_\sigma^i(z^i) = Q^i_\sigma(y^i | s^i, x^i)\sigma^i(x^i | s^i)p_{S^i}(s^i)$. We sometimes abuse notation and write $Q_\sigma^i(z^i) \equiv Q^i_\sigma(y^i | s^i, x^i)$, and similarly for $Q_\theta^i$. We begin by proving a few claims.

Claim A.1. There exists $\theta^i_* \in \Theta^i$ and $K < \infty$ such that, for all $\sigma \in \Sigma$, $K^i(\sigma, \theta^i_*) \leq K$.

Proof: By Assumption 1 and finiteness of $\mathcal{Z}^i$, there exists $\theta^i_* \in \Theta$ and $\alpha \in (0, 1)$ such that $Q^i_{\theta^i_*}(z^i) \geq \alpha$ for all $z^i \in \mathcal{Z}^i$. Thus, for all $\sigma \in \Sigma$, $K(\sigma, \theta^i_*) \leq -E_{\bar{P}^i_\theta} \ln Q^i_{\theta^i_*(Z^i)} \leq -\ln \alpha$. □

Claim A.2. Fix any $\theta^i \in \Theta^i$ and $(\sigma_n)_n$ such that $Q^i_{\theta^i}(z^i) > 0$ for all $z^i \in \mathcal{Z}^i$ and $\lim_n \sigma_n = \sigma$. Then $\lim_{n \to \infty} K^i(\sigma_n, \theta^i) = K^i(\sigma, \theta^i)$.

Proof: Note that

$$K^i(\sigma_n, \theta^i) - K(\sigma, \theta^i) = \sum_{z^i \in \mathcal{Z}^i} \left( \bar{P}^i_{\sigma_n}(z^i) \ln Q^i_{\sigma_n}(z^i) - \bar{P}^i_\sigma(z^i) \ln Q^i_\sigma(z^i) \right)$$

$$+ \sum_{z^i \in \mathcal{Z}^i} (\bar{P}^i_{\sigma_n}(z^i) - \bar{P}^i_\sigma(z^i)) \ln Q^i_{\theta^i}(z^i).$$

(25)

The first term in the RHS of (25) converges to zero because $\lim_n \sigma_n = \sigma$, $Q^i_\sigma$ is continuous, and $x \ln x$ is continuous for all $x \in [0, 1]$. The second term converges to zero because $\lim_n \sigma_n = \sigma$, $\bar{P}^i_\sigma$ is continuous, and $\ln Q^i_{\theta^i}(z^i) \in (-\infty, 0]$ for all $z^i \in \mathcal{Z}^i$.

□

Claim A.3. $K^i$ is (jointly) lower semicontinuous: Fix any $(\theta^i_n)_n$ and $(\sigma_n)_n$ such that $\lim_n \theta^i_n = \theta^i$, $\lim_n \sigma_n = \sigma$. Then $\liminf_{n \to \infty} K^i(\sigma_n, \theta^i_n) \geq K(\sigma, \theta^i)$.

Proof: Note that

$$K^i(\sigma_n, \theta^i_n) - K(\sigma, \theta^i) = \sum_{z^i \in \mathcal{Z}^i} \left( \bar{P}^i_{\sigma_n}(z^i) \ln Q^i_{\sigma_n}(z^i) - \bar{P}^i_\sigma(z^i) \ln Q^i_\sigma(z^i) \right)$$

$$+ \sum_{z^i \in \mathcal{Z}^i} (\bar{P}^i_{\sigma_n}(z^i) \ln Q^i_{\theta^i_n}(z^i) - \bar{P}^i_\sigma(z^i) \ln Q^i_{\theta^i_n}(z^i)).$$

(26)
The first term in the RHS of (26) converges to zero (same argument as equation 25). The proof concludes by showing that, for all $z' \in \mathbb{Z}$,

$$\lim_{n \to \infty} -\tilde{P}^i_{\sigma_n}(z') \ln Q^i_{\theta_n}(z') \geq -\tilde{P}^i_{\sigma}(z') \ln Q^i_{\theta}(z'). \quad (27)$$

Suppose $\lim_{n \to \infty} -\tilde{P}^i_{\sigma_n}(z') \ln Q^i_{\theta_n}(z') \leq M < \infty$ (if not, (27) holds trivially). Then either (i) $\tilde{P}^i_{\sigma_n}(z') \to \tilde{P}^i_{\sigma}(z') > 0$, in which case (27) holds with equality, or (ii) $\tilde{P}^i_{\sigma_n}(z') \to \tilde{P}^i_{\sigma}(z') = 0$, in which case (27) holds because its RHS is zero (by convention that $0 \ln 0 = 0$) and its LHS is always nonnegative. □

**Claim A.4.** Suppose that $\xi^i$ is a real-valued vector with dimension $\#X^i$ and $P_\xi$ is absolutely continuous. Then, for all $(s^i, x^i)$,

$$\sigma^i(\mu^i)(x^i | s^i) = P_\xi \left( \xi^i : \arg\max_{\bar{x}^i \in \overline{X}^i} \mathbb{E}_{Q^i_{\mu^i}(\cdot | s^i, \bar{x}^i)} \left[ \pi^i(\bar{x}^i, Y^i) \right] + \xi^i(\bar{x}^i) \right)$$

is continuous as a function of $\mu^i$.

Proof: Fix $s^i \in S^i$ and $x^i \in X^i$ and define $g^i(\theta^i, \bar{x}^i) = \mathbb{E}_{Q^i_{\mu^i}(\cdot | s^i, \bar{x}^i)} \left[ \pi^i(x^i, Y^i) - \pi^i(\bar{x}^i, Y^i) \right]$, $G^i(\mu^i, \bar{x}^i) = \int_{\Theta} g^i(\theta^i, \bar{x}^i) \mu^i(d\theta^i)$ and $\Delta \xi^i(\bar{x}^i) = \xi^i(\bar{x}^i) - \xi^i(x^i)$ for all $\bar{x}^i \in X^i$. Because $g^i(\cdot, \bar{x}^i)$ is continuous and bounded, then $G^i(\cdot, \bar{x}^i)$ is continuous under weak convergence. Then, for every $\varepsilon > 0$, there exists $N_\varepsilon$ such that, for all $n \geq N_\varepsilon$,

$$|\sigma^i(\mu^i_n)(x^i | s^i) - \sigma^i(\mu^i)(x^i | s^i)| \leq P_\xi \left( \Delta \xi^i(\bar{x}^i) \leq G^i(\mu^i, \bar{x}^i) + \varepsilon \ \forall \bar{x}^i \right) - P_\xi \left( \Delta \xi^i(\bar{x}^i) \leq G^i(\mu^i, \bar{x}^i) - \varepsilon \ \forall \bar{x}^i \right) \leq \sum_{\bar{x}^i \in X^i} P_\xi \left( G^i(\mu^i, \bar{x}^i) - \varepsilon \leq \Delta \xi^i(\bar{x}^i) \leq G^i(\mu^i, \bar{x}^i) + \varepsilon \right).$$

By absolute continuity of $P_\xi$, the above expression goes to zero as $\varepsilon$ goes to zero. □

**Proof of Lemma 1.** Part (i). Note that

$$K^i(\sigma, \theta^i) = - \sum_{(s^i, x^i) \in S^i \times X^i} E_{Q_{\sigma^i}(\cdot | s^i, x^i)} \ln \left[ \frac{Q^i_{\theta^i}(Y^i | s^i, x^i)}{Q^i_{\sigma}(Y^i | s^i, x^i)} \right] \sigma_i(x^i | s^i)p_{S^i}(s^i) \geq - \sum_{(s^i, x^i) \in S^i \times X^i} \ln \left( \frac{Q^i_{\mu^i}(Y^i | s^i, x^i)}{Q^i_{\sigma}(Y^i | s^i, x^i)} \right) \sigma_i(x^i | s^i)p_{S^i}(s^i) \quad (28)$$

$$= 0,$$
where Jensen’s inequality and the strict concavity of \( \ln(\cdot) \) imply the inequality in (28) as well as the fact that (28) holds with equality if and only if \( Q^i_{\hat{\sigma}}(\cdot \mid s^i, x^i) = Q^i_{\hat{\sigma}}(\cdot \mid s^i, x^i) \) for all \((s^i, x^i)\) such that \( \sigma_i(x^i \mid s^i) > 0 \) (recall that, by assumption, \( p_{si}(s^i) > 0 \)).

Part (ii). \( \Theta^i(\sigma) \) is nonempty: By Claim A.1, there exists \( K < \infty \) such that the minimizers are in the constraint set \( \{ \theta^i \in \Theta^i : K^i(\sigma, \theta^i) \leq K \} \). Because \( K^i(\sigma, \cdot) \) is continuous over a compact set, a minimum exists.

\( \Theta^i(\sigma) \) is uhc: Fix any \((\sigma_n)\) and \((\theta^i_n)\) such that \( \lim_n \sigma_n = \sigma \), \( \lim_n \theta^i_n = \theta^i \), and \( \theta^i_n \in \Theta^i(\sigma_n) \) for all \( n \). We want to show that \( \theta^i \in \Theta^i(\sigma) \) (so that \( \Theta^i(\cdot) \) has a closed graph and hence, by compactness of \( \Theta^i \), it is uhc). Suppose, in order to obtain a contradiction, that \( \theta^i \notin \Theta^i(\sigma) \). Then, by Claim A.1, there exists \( \hat{\theta}^i \in \Theta^i \) and \( \varepsilon > 0 \) such that \( K^i(\sigma, \hat{\theta}^i) \leq K^i(\sigma, \theta^i) - 3\varepsilon \) and \( K^i(\sigma, \hat{\theta}^i) < \infty \). By Assumption 1, there exists \( (\hat{\theta}^i_j) \) with \( \lim_j \hat{\theta}^i_j = \hat{\theta}^i \) and, for all \( j \), \( Q^i_{\hat{\theta}^i_j}(z^i) > 0 \) for all \( z^i \in \mathbb{Z}^i \). We will show that there is an element of the sequence, \( \hat{\theta}^i_j \), that “does better” than \( \theta^i_n \) given \( \sigma_n \), which is a contradiction. Because \( K^i(\sigma, \hat{\theta}^i) < \infty \), continuity of \( K^i(\sigma, \cdot) \) implies that there exists \( J \) large enough such that \( \left| K^i(\sigma, \hat{\theta}^i) - K^i(\sigma, \theta^i) \right| \leq \varepsilon/2 \). Moreover, Claim A.2 applied to \( \theta^i = \hat{\theta}^i_j \) implies that there exists \( N_{\varepsilon,J} \) such that, for all \( n \geq N_{\varepsilon,J} \),

\[
\left| K^i(\sigma_n, \hat{\theta}^i_j) - K^i(\sigma, \hat{\theta}^i) \right| \leq \left| K^i(\sigma_n, \hat{\theta}^i_j) - K^i(\sigma, \hat{\theta}^i) \right| + \left| K^i(\sigma, \hat{\theta}^i) - K^i(\sigma, \hat{\theta}^i) \right| \leq \varepsilon
\]

and, therefore,

\[
K^i(\sigma_n, \hat{\theta}^i_j) \leq K^i(\sigma, \hat{\theta}^i) + \varepsilon \leq K^i(\sigma, \theta^i) - 2\varepsilon. \tag{29}
\]

Suppose \( K^i(\sigma, \theta^i) < \infty \). By Claim A.3, there exists \( n_{\varepsilon} \geq N_{\varepsilon,J} \) such that \( K^i(\sigma_{n_{\varepsilon}}, \theta^i_{n_{\varepsilon}}) \geq K^i(\sigma, \theta^i) - \varepsilon \). This result together with expression (29), implies that \( K^i(\sigma_{n_{\varepsilon}}, \hat{\theta}^i_{n_{\varepsilon}}) \leq K(\sigma_{n_{\varepsilon}}, \theta^i_{n_{\varepsilon}}) - \varepsilon \). But this contradicts \( \theta^i_{n_{\varepsilon}} \in \Theta^i(\sigma_{n_{\varepsilon}}) \). Finally, if \( K^i(\sigma, \theta^i) = \infty \), Claim A.3 implies that there exists \( n_{\varepsilon} \geq N_{\varepsilon,J} \) such that \( K^i(\sigma_{n_{\varepsilon}}, \theta^i_{n_{\varepsilon}}) \geq 2K \), where \( K \) is the bound defined in Claim A.1. But this also contradicts \( \theta^i_{n_{\varepsilon}} \in \Theta^i(\sigma_{n_{\varepsilon}}) \).

\( \Theta^i(\sigma) \) is compact: As shown above, \( \Theta^i(\cdot) \) has a closed graph, and so \( \Theta^i(\sigma) \) is a closed set. Compactness of \( \Theta^i(\sigma) \) follows from compactness of \( \Theta^i \). □

**Proof of Theorem 1.** We prove the result in two parts. Part 1. We show existence of equilibrium in the perturbed game. Let \( \Gamma : \times_{i \in I} \Delta(\Theta^i) \to \times_{i \in I} \Delta(\Theta^i) \) be
a correspondence such that, for all $\mu = (\mu^i)_{i \in I} \in \times_{i \in I} \Delta(\Theta^i)$,

$$\Gamma(\mu) = \times_{i \in I} \Delta(\Theta^i(\sigma(\mu))),$$

where $\sigma(\mu) = (\sigma^i(\mu^i))_{i \in I} \in \Sigma$ and is defined as

$$\sigma^i(\mu^i)(x^i|s^i) = P^i_x \left( \xi^i : x^i \in \arg\max_{\bar{x}^i \in X^i} E_{Q^i_{\mu^i}}(\cdot|s^i,\bar{x}^i) \left[ \pi^i(\bar{x}^i, Y^i) \right] + \xi^i(\bar{x}^i) \right) \quad (30)$$

for all $(x, s) \in X^i \times S^i$. Note that if there exists a $\mu_* \in \times_{i \in I} \Delta(\Theta^i)$ such that $\mu_* \in \Gamma(\mu_*)$, then $\sigma_* \equiv (\sigma^i(\mu_*^i))_{i \in I}$ is an equilibrium of the perturbed game. We now show that such $\mu_*$ exists by checking the conditions of the Kakutani-Fan-Glicksberg fixed point theorem: (i) $\times_{i \in I} \Delta(\Theta^i)$ is compact, convex and locally convex Hausdorff: The set $\Delta(\Theta^i)$ is convex, and since $\Theta^i$ is compact $\Delta(\Theta^i)$ is also compact under the weak topology (Aliprantis and Border (2006), Theorem 15.11). By Tychonoff’s theorem, $\times_{i \in I} \Delta(\Theta^i)$ is compact too. Finally, the set is also locally convex under the weak topology.\(^{42}\); (ii) $\Gamma$ has convex, nonempty images: It is clear that $\Delta(\Theta^i(\sigma(\mu)))$ is convex valued for all $\mu$. Also, by Lemma 1, $\Theta^i(\sigma(\mu))$ is non-empty for all $\mu$; (iii) $\Gamma$ has a closed graph: Let $(\mu_n, \mu_n)$ be such that $\mu_n \in \Gamma(\mu_n)$ and $\mu_n \to \mu$ and $\mu_n \to \hat{\mu}$ (under the weak topology). By Claim A.4, $\mu^i \mapsto \sigma^i(\mu^i)$ is continuous. Thus, $\sigma_n \equiv (\sigma^i(\mu^i_n))_{i \in I} \to \sigma \equiv (\sigma^i(\mu^i))_{i \in I}$. In particular, $\hat{\mu}_n \in \times_{i \in I} \Delta(\Theta^i(\sigma_n))$. By Lemma 1, $\sigma \mapsto \Theta^i(\sigma)$ is uhc; thus, by Theorem 17.13 in Aliprantis and Border (2006), $\times_{i \in I} \Delta(\Theta^i(\sigma))$ is also uhc. Therefore, $\hat{\mu} \in \times_{i \in I} \Delta(\Theta^i(\sigma)) = \Gamma(\mu)$.

Part 2. Fix a sequence of perturbed games indexed by the probability of perturbations $(P^i_{\xi,n})_n$. By Part 1, there is a corresponding sequence of fixed points of $\Gamma$, $(\mu_n)_n$, such that $\mu_n \in \times_{i \in I} \Delta(\Theta^i(\sigma_n))$ for all $n$, where $\sigma_n \equiv \sigma^i(\mu^i_n, P^i_{\xi,n})(x^i|s^i)$ (see equation (30), where we now explicitly account for the dependance on $P^i_{\xi,n}$). By compactness, there exist subsequences of $(\mu_n)_n$ and $(\sigma_n)_n$ that converge to $\mu$ and $\sigma$, respectively. Since $\times_{i \in I} \Delta(\Theta^i(\sigma))$ is uhc, then $\mu \in \times_{i \in I} \Delta(\Theta^i(\sigma))$. We now show that if we choose $(P^i_{\xi,n})_n$ such that, for all $\epsilon > 0$,

$$\lim_{n \to \infty} P^i_{\xi,n} \left( \left\| \xi^i_n \right\| \geq \epsilon \right) = 0, \quad (31)$$

then $\sigma$ is optimal given $\mu$ in the unperturbed game—this establishes existence of equilib-
rium in the unperturbed game. Suppose not, so that there exists \( i, s^i, x^i, \tilde{x}^i \), and \( \varepsilon > 0 \) such that \( \sigma^i(x^i \mid s^i) > 0 \) and \( E_{\tilde{Q}_{\mu_i}^i(\cdot \mid s^i)}[\pi^i(x^i, Y^i)] + 4\varepsilon \leq E_{\tilde{Q}_{\mu_i}^i(\cdot \mid s^i, \tilde{x}^i)}[\pi^i(\tilde{x}^i, Y^i)] \).

By continuity of \( \mu^i \mapsto \tilde{Q}_{\mu_i}^i \) and the fact that \( \lim_n \mu_{n_i}^i = \mu^i \), there exists \( n_1 \) such that, for all \( n \geq n_1 \), \( E_{Q_{\mu_n^i}^i(\cdot \mid s^i)}[\pi^i(x^i, Y^i)] + 2\varepsilon \leq E_{Q_{\mu_n^i}^i(\cdot \mid s^i, \tilde{x}^i)}[\pi^i(\tilde{x}^i, Y^i)] \). It then follows from (30) and (31) that \( \lim_n \sigma^i(\mu_{n_i}^i; P_{\xi,n})(x^i \mid s^i) = 0 \). But this contradicts \( \lim_n \sigma^i(\mu_{n_i}^i; P_{\xi,n})(x^i \mid s^i) = \sigma^i(x^i \mid s^i) > 0 \). □

**Proof of Proposition 3.** In the next paragraph, we prove the following result: For all \( \sigma \) and \( \tilde{\theta}_i^j \in \Theta^i(\sigma) \), it follows that

\[
P_{\Omega, \tilde{\theta}_i^j}(\omega' \mid s^i(\omega)) = p_{\Omega}(\omega' \mid s^i(\omega)) \tag{32}
\]

for all \( \omega, \omega' \) and

\[
P_{X^{-i}, \tilde{\theta}_i^j}(x^{-i} \mid t^i(\omega')) = \sum_{\omega'' \in \Omega} p_{\Omega}(\omega'' \mid \alpha^i(\omega')) \prod_{j \neq i} \sigma^j(x^j \mid s^j(\omega'')) \tag{33}
\]

for all \( \omega', x^{-i} \). Equivalence between Berk-Nash and ABEE is then immediate by using (32) and (33) and by noticing that expected utility of player \( i \) with signal \( s^i(\omega) \) and beliefs \( \tilde{\theta}_i^j \) is given by \( \sum_{\omega' \in \Omega} P_{\Omega, \tilde{\theta}_i^j}(\omega' \mid s^i(\omega)) \sum_{x^{-i} \in X^{-i}} P_{X^{-i}, \tilde{\theta}_i^j}(x^{-i} \mid t^i(\omega)) \pi^i(x^i, \omega', x^{-i}) \).

Proof of (32) and (33): note that the (negative of) wKL divergence of player \( i \) given \( \sigma \) can be written, up to a constant, as

\[
= \sum_{\omega'} \sum_{x, x^{-i}} \ln \left( P_{\Omega, \tilde{\theta}_i^j}(\omega' \mid s^i(\omega')) P_{X^{-i}, \tilde{\theta}_i^j}(x^{-i} \mid t^i(\omega)) \right) \prod_{j \neq i} \sigma^j(x^j \mid s^j(\omega')) p(\omega, s^i(\omega'))
\]

\[
= \sum_{\omega'} \ln \left( P_{\Omega, \tilde{\theta}_i^j}(\omega' \mid s^i(\omega')) \right) p_{\Omega}(\omega' \mid s^i(\omega')) p(s^i(\omega'))
+ \sum_{t^i \in T^i} \sum_{x^{-i}} \ln \left( P_{X^{-i}, \tilde{\theta}_i^j}(x^{-i} \mid t^i) \right) \sum_{\tilde{\omega}} \prod_{j \neq i} \sigma^j(x^j \mid s^j(\tilde{\omega})) p(\tilde{\omega}).
\]

It is straightforward to check that any parameter that maximizes the above expression satisfies (32) and (33).

**Proof of Claims in Lemma 2.** Here, we prove claims 2.1-3 used in the proof of Lemma 2 in the text. For each \( z^i \in \mathbb{Z}^i \), define \( freq^i(z^i) = \frac{1}{t} \sum_{\tau=1}^t \mathbf{1}_{z^i}(z^i_\tau) \).
function \( K^i_t \) can be written as

\[
K^i_t(h, \theta^i) = \kappa^i_{1t}(h) + \kappa^i_{2t}(h) + \kappa^i_{3t}(h, \theta^i),
\]

where

\[
\kappa^i_{1t}(h) = -\frac{1}{t} \sum_{\tau=1}^{t} \sum_{z^i \in \mathbb{Z}^i} (1_{z^i}(z^i_\tau) - \bar{P}^i_{\sigma^i}(z^i)) \ln Q_{\sigma^i}^i(z^i),
\]

\[
\kappa^i_{2t}(h) = -\frac{1}{t} \sum_{\tau=1}^{t} \sum_{z^i \in \mathbb{Z}^i} \bar{P}^i_{\sigma^i}(z^i) \ln Q_{\sigma^i}^i(z^i),
\]

and

\[
\kappa^i_{3t}(h, \theta^i) = \sum_{z^i \in \mathbb{Z}^i} \text{freq}_t^i(z^i) \ln Q_{\theta^i}^i(z^i).
\]

The claims made below hold almost surely in \( \mathcal{H} \), but we omit this qualification in the proof.

**Claim 2.1.** \( \lim_{t \to \infty} K^i_t(h, \theta^i) = -K^i(\sigma, \theta^i) \) for all \( \theta^i \in \Theta^i \), a.s. in \( \mathcal{H} \).

**Proof:** First, consider \( \kappa^i_{1t}(h) \). Define, for all \( z^i \in \mathbb{Z}^i \),

\[
l^i_t(h, z^i) = (1_{z^i}(z^i_\tau) - \bar{P}^i_{\sigma^i}(z^i)) \ln Q_{\sigma^i}^i(z^i)
\]

and \( L^i_t(h, z^i) = \sum_{\tau=1}^{t} l^i_t(h, z^i) \). Fix any \( z^i \in \mathbb{Z}^i \). We now show that \( L^i_t(h, z^i) \) converges a.s. to an integrable, and, therefore, finite \( L^i_\infty(h, z^i) \). In order to show this, we use martingale convergence results. First, we show that \( (L^i_t(h, z^i))_t \) is a martingale with respect to \( \mathbb{P}^{\mu_0, \phi} \). Let \( h^t \) denote the partial history until time \( t \). Observe that

\[
E_{\mathbb{P}^{\mu_0, \phi}_t(h^t)} \left[ L^i_{t+1}(h, z^i) \right] = \sum_{\tau=1}^{t} \tau^{-1} l^i_\tau(h, z^i) + \frac{1}{t+1} E_{\mathbb{P}^{\mu_0, \phi}_t(h^t)} \left[ l^i_{t+1}(h, z^i) \right]
\]

\[
= \sum_{\tau=1}^{t} \tau^{-1} l^i_\tau(h, z^i) + \frac{1}{t+1} \left( \ln Q_{\sigma^i_{t+1}}^i(z^i) \right) E_{\mathbb{P}^{\mu_0, \phi}_t(h^t)} \left[ 1_{z^i_{t+1}}(z^i_{t+1}) - \bar{P}^i_{\sigma^i_{t+1}}(z^i) \right]
\]

\[
= L^i_t(h, z^i).
\]
Second, we show that \((L_i(t, z^i))_t\) is uniformly integrable, i.e.,

\[
\sup_t E_{P_{\mu_0, \phi}} [|L_i^t|1\{|L_i^t| \geq M\}] \to 0
\]  \hspace{1cm} (35)

as \(M \to \infty\). By the Markov inequality,

\[
E_{P_{\mu_0, \phi}} [|L_i^t|1\{|L_i^t| \geq M\}] \leq M^{-1} E_{P_{\mu_0, \phi}} \left[ (L_i^t(h, z^i))^2 \right]
\]

\[
= M^{-1} \sum_{\tau = 1}^t \tau^{-2} \left( l_i^{\tau}(h, z^i) \right)^2 + 2 \sum_{\tau' > \tau} \frac{1}{\tau'^2} E_{P_{\mu_0, \phi}} \left[ l_i^{\tau'}(h, z^i) l_i^{\tau}(h, z^i) \right]
\]

\[
= M^{-1} \left\{ \sum_{\tau = 1}^t \tau^{-2} E_{P_{\mu_0, \phi}} \left[ \left( l_i^{\tau}(h, z^i) \right)^2 \right] + \sum_{\tau' > \tau} \frac{2}{\tau'^2} E_{P_{\mu_0, \phi}} \left[ l_i^{\tau'}(h, z^i) l_i^{\tau}(h, z^i) \right] \right\}
\]

\[
\leq M^{-1} \sum_{\tau = 1}^t \tau^{-2} E_{P_{\mu_0, \phi}} \left[ \left( \ln Q_i^{\tau}(z^i) \right)^2 Q_i^{\tau}(z^i) \right]
\]

\[
\leq M^{-1}
\]

where the fourth line follows from the fact that, for \(\tau' > \tau\), \(E_{P_{\mu_0, \phi}} \left[ l_i^{\tau'}(h, z^i) l_i^{\tau}(h, z^i) \right] = 0\) because \((l_i^t)_t\) is a martingale difference sequence; the fifth line follows from the law of iterated expectations and the fact that

\[
E_{P_{\mu_0, \phi}(|h\tau - 1|)} \left[ \left( l_i^{\tau}(h, z^i) \right)^2 \right] = \left( \ln Q_i^{\tau}(z^i) \right)^2 \left( \bar{P}_i^{\tau}(z^i) - (\bar{P}_i^{\tau}(z^i))^2 \right)
\]

\[
\leq \left( \ln Q_i^{\tau}(z^i) \right)^2 \bar{P}_i^{\tau}(z^i)
\]

\[
\leq \left( \ln Q_i^{\tau}(z^i) \right)^2 Q_i^{\tau}(z^i);
\]

and the last inequality follows because \((\ln x)^2 x \leq 1\) for all \(x \in [0, 1]\), where we use the convention that \((\ln 0)^2 0 = 0\). Therefore, (35) holds and, by theorem 36 in Pollard (2001), page 154, \(L_i^t(h, z^i)\) converges a.s.-\(P_{\mu_0, \phi}\) to a finite \(L_i^\infty(h, z^i)\). Thus, by
Kronecker’s lemma (Pollard (2001), page 105)\footnote{This lemma implies that for a sequence \((\ell_t)_t\) if \(\sum_t \ell_t < \infty\), then \(\sum_{t=1}^{t} \frac{b_t}{b_t} \ell_t \to 0\) where \((b_t)_t\) is a non-decreasing positive real valued that diverges to \(\infty\). We can apply the lemma with \(\ell_t \equiv t^{-1} \ell_t\) and \(b_t = t\).}, it follows that

\[
\lim_{t \to \infty} \sum_{z^i \in \mathbb{Z}^i} \left\{ t^{-1} \sum_{\tau=1}^{t} \ln Q_{\sigma^i}(z^i) \left( 1_{z^i}(z^i) - \bar{P}_{\sigma^i}(z^i) \right) \right\} = 0
\]
a.s.-\(P^{\mu_0, \phi}\). Therefore,

\[
\lim_{t \to \infty} \kappa^i_{\ell_t}(h) = 0.
\] (36)
a.s.-\(P^{\mu_0, \phi}\). Next, consider \(\kappa^i_{2\ell_t}(h)\). The assumption that \(\lim_{t \to \infty} \sigma_t = \sigma\) and continuity of \(Q_{\sigma}^i \ln Q_{\sigma}^i\) in \(\sigma\) imply that

\[
\lim_{t \to \infty} \kappa^i_{2\ell_t}(h) = - \sum_{(s^i,x^i) \in S^i \times X^i} E_{Q_{\sigma}^i(s^i,x^i)} \left[ \ln Q_{\sigma}^i(Y^i \mid s^i,x^i) \right] \sigma^i(x^i \mid s^i) p_{S^i}(s^i). \] (37)

Finally, consider \(\kappa^i_{3\ell_t}(h, \theta^i)\). For all \(z^i \in \mathbb{Z}^i\),

\[
|freq_t(z^i) - P^i_{\sigma}(z^i)| = \left| \frac{1}{t} \sum_{\tau=1}^{t} \left( \left[1_{z^i}(z^i) - P^i_{\sigma}(z^i) \right] \right) \right| + \left| \frac{1}{t} \sum_{\tau=1}^{t} \left( P^i_{\sigma}(z^i) - P^i_{\sigma}(z^i) \right) \right|. \] (38)

The first term in the RHS of (38) goes to zero (the proof is essentially identical to the proof above that \(\kappa^i_{1}\) goes to zero, where equation (34) is replaced by \(l^i_t(h,z^i) = (1_{z^i}(z^i) - \bar{P}^i_{\sigma}(z^i))\). The second term goes to zero because \(\lim_{t \to \infty} \sigma_t = \sigma\) and \(\bar{P}_i\) is continuous. Thus, for all \(\zeta > 0\), there exists \(\tilde{t}_\zeta\) such that, for all \(t \geq \tilde{t}_\zeta\),

\[
|freq_t(z^i) - P^i_{\sigma}(z^i)| < \zeta
\] (39)

for all \(z^i \in \mathbb{Z}^i\). Therefore,

\[
\lim_{t \to \infty} \kappa^i_{3\ell_t}(h, \theta^i) = \sum_{(s^i,x^i) \in S^i \times X^i} E_{Q_{\sigma}^i(s^i,x^i)} \left[ \ln Q_{\theta^i}^i(Y^i \mid s^i,x^i) \right] \sigma^i(x^i \mid s^i) p_{S^i}(s^i). \] (40)

Equations (36), (37), and (40) imply (?). Claim 2.2. There exists \(\xi^* > 0\) and \(T_{\xi^*}\).
such that, for all $t \geq T_{\xi^*}$,

$$K_i^t(h, \theta^i) < -(K^i(\sigma, \theta^i) + (3/2)\alpha_\varepsilon)$$

for all $\theta^i \notin \Theta^i_{\sigma, \xi}$.

Proof: Equations (36) and (37) imply that, for all $\gamma > 0$, there exists $\hat{t}_\gamma$ such that, for all $t \geq \hat{t}_\gamma$,

$$\left| \kappa^i_{1t}(h) + \kappa^i_{2t}(h) + \sum_{(s^i, x^i) \in \mathcal{S}^i \times \mathcal{X}^i} E_{Q_\sigma(\cdot|s^i, x^i)} \left[ \ln Q_\sigma^i(Y^i | s^i, x^i) \right] \sigma^i(x^i | s^i)p_{S^i}(s^i) \right| \leq \gamma. \quad (41)$$

For all $\theta^i \notin \Theta^i_{\sigma, \xi}$, let $z^i_{\theta^i}$ be such that $\hat{P}^i_\sigma(z^i_{\theta^i}) > 0$ and $Q^i_{\theta^i}(z^i_{\theta^i}) < \xi$. By (39), there exists $t_{p_L/2}$ such for all $t \geq t_{p_L/2}$,

$$\kappa^i_{3t}(h, \theta^i) \leq \text{freq}_t(z^i_{\theta^i}) \ln Q^i_{\theta^i}(z^i_{\theta^i}) \leq (p^i_{L}/2) \ln \xi, \quad (42)$$

for all $\theta^i \notin \Theta^i_{\sigma, \xi}$, where $p^i_L = \min_{\mathcal{Z}^i} \{ \hat{P}^i_\sigma(z^i) : \hat{P}^i_\sigma(z^i) > 0 \}$. Then (41) and (42) imply that, for all $t \geq t_1 \equiv \max \{ t_{p^i_{L}/2}, \hat{t}_1 \}$,

$$K^i_t(h, \theta^i) \leq - \sum_{(s^i, x^i) \in \mathcal{S}^i \times \mathcal{X}^i} E_{Q_\sigma(\cdot|s^i, x^i)} \left[ \ln Q_\sigma^i(Y^i | s^i, x^i) \right] \sigma^i(x^i | s^i)p_{S^i}(s^i) + 1 + (p^i_{L}/2) \ln \xi$$

$$\leq \#\mathbb{Z}^i + 1 + (q^i_{L}/2) \ln \xi \quad (43)$$

for all $\theta^i \notin \Theta^i_{\sigma, \xi}$, where the second line follows from the facts that

$$- \sum_{(s^i, x^i) \in \mathcal{S}^i \times \mathcal{X}^i} E_{Q_\sigma(\cdot|s^i, x^i)} \left[ \ln Q_\sigma^i(Y^i | s^i, x^i) \right] \sigma^i(x^i | s^i)p_{S^i}(s^i) \leq$$

$$- \sum_{(s^i, x^i) \in \mathcal{S}^i \times \mathcal{X}^i} \sum_{y^i \in \mathcal{Y}^i} (\ln Q_\sigma^i(y^i | s^i, x^i)) Q_\sigma^i(y^i | s^i, x^i)$$

and $x \ln(x) \in [-1, 0]$ for all $x \in [0, 1]$. In addition, the fact that $K^i_0(\sigma) < \infty$ and $\alpha_\varepsilon \leq \bar{\alpha} < \infty$ for all $\varepsilon \leq \bar{\varepsilon}$ implies that the RHS of (43) can be made lower than $-(K^i_0(\sigma) + (3/2)\alpha_\varepsilon)$ for some sufficiently small $\xi^*$. 

52
Claim 2.3. For all $\xi > 0$, there exists $\hat{T}_\xi$ such that, for all $t \geq \hat{T}_\xi$,

$$K_i^i(h, \theta^i) < -K^i(\sigma, \theta^i) + \alpha\varepsilon/2$$

for all $\theta^i \in \Theta_{i,\sigma,\xi}$.

Proof: For any $\xi > 0$, let $\zeta_\xi = -\alpha\varepsilon/\left(\#\mathbb{Z}^i 4 \ln \xi\right) > 0$ and consider any $\theta^i \in \Theta_{i,\sigma,\xi}$, so that $Q_{\theta^i}(z^i) \geq \xi$ for all $z^i$ such that $\bar{P}_{\theta^i}(z^i) > 0$. Then, by (39), there exists $\hat{t}_{\zeta_\xi}$ such that, for all $t \geq \hat{t}_{\zeta_\xi}$,

$$K_i^i(h, \theta^i) \leq \sum_{\{z^i : \bar{P}_{\theta^i}(z^i) > 0\}} \text{freq}_i^i(z^i) \ln Q_{\theta^i}(z^i)$$

$$\leq \sum_{\{z^i : \bar{P}_{\theta^i}(z^i) > 0\}} \left(\bar{P}_{\theta^i}(z^i) - \zeta_\xi\right) \ln Q_{\theta^i}(z^i)$$

$$\leq \sum_{(s^i, x^i) \in \mathbb{S}^i \times \mathbb{X}^i} E_{Q_{\sigma^i}(\cdot | s^i, x^i)} \left[\ln Q_{\theta^i}(Y^i | s^i, x^i)\right] \sigma^i(x^i | s^i) p_{s^i}(s^i) - \#\mathbb{Z}^i \zeta_\xi \ln \xi,$n

$$= \sum_{(s^i, x^i) \in \mathbb{S}^i \times \mathbb{X}^i} E_{Q_{\sigma^i}(\cdot | s^i, x^i)} \left[\ln Q_{\theta^i}(Y^i | s^i, x^i)\right] \sigma^i(x^i | s^i) p_{s^i}(s^i) + \alpha\varepsilon/4. \quad (44)$$

Then (41) and (44) imply the result for $\hat{T}_\xi = \max\{\hat{t}_{\zeta_\xi}, \hat{t}_{\alpha\varepsilon/4}\}. \quad \square$