An Equilibrium Framework for Players with Misspecified Models*

Ignacio Esponda
(NYU Stern)

Demian Pouzo
(UC Berkeley)

March 15, 2014

Abstract

We introduce an equilibrium framework that relaxes the standard assumption that people have a correctly-specified view of their environment. Players repeatedly play a simultaneous-move game where they potentially face both strategic and payoff uncertainty. Each player has a potentially misspecified view of the environment and uses Bayes' rule to update her views based on the (possibly partial) feedback obtained at the end of each period. We show that steady-state behavior of this multi-player decision and learning problem is captured by a generalized notion of *equilibrium*: a strategy profile such that each player optimizes given certain beliefs and where these beliefs put probability one on those subjective distributions over consequences that are closest—in terms of relative entropy—to the correct, equilibrium distribution. Standard solution concepts such as Nash equilibrium and self-confirming equilibrium constitute special cases where players learn with correctly-specified models. Our framework unifies an existing literature on bounded rationality and misspecified learning and provides a systematic approach to modeling certain aspects of bounded rationality.

*We thank Pierpaolo Battigalli, Larry Blume, Aaron Bodoh-Creed, Emilio Espino, Erik Eyster, Drew Fudenberg, Philippe Jehiel, Kristóf Madarász, Matthew Rabin, Ariel Rubinstein, Joel Sobel, Jörg Stoye, and several seminar participants for helpful comments. Esponda: Stern School of Business, New York University, 44 W Fourth Street KMC 7-76, New York, NY 10012, iesponda@stern.nyu.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 theorists 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 (or, by the decision maker, in the special case where there is only one economic agent in the environment). 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. For example, a subjective model might arise from a player's underlying belief over the distribution of payoff relevant states and a belief of the strategies of other players. Alternatively, a player might not even understand that she is facing other players. A key feature is that we allow the subjective model to be misspecified. 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. 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 people's limiting behavior when they 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.

The fact that our notion of equilibrium includes standard equilibrium concepts as special cases helps to put our contribution into context. Suppose that whether or not players get feedback about consequences does not depend on their own actions; we call this property *full feedback*. Suppose, in addition, that the subjective model is *correctly specified*, which roughly means that the support of each player's prior contains the true, objective model. Then, our notion of equilibrium is equivalent to Nash equilibrium. More generally, the framework forces the modeler to make explicit 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 pa-

per, there are examples form 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 learning. 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) provide the only general treatment in games that we are aware of, and also 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), and sparse Nash equilibrium (Gabaix (2012)). In particular, analogy-based expectation, (fully) cursed, and behavioral equilibrium can all be integrated into our framework, thus

[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 over-reaction 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)).

clarifying the underlying misspecification in each of these cases.

This literature can also be viewed as continuing the tradition of equilibrium analysis in non-cooperative game theory. The initial notion of Nash equilibrium (Nash, 1951) assumed that players only faced strategic uncertainty. Subsequently, the work of Vickrey (1961), Harsanyi (1967-8), and others extended Nash equilibrium to also account for payoff uncertainty. The previous literature and our paper relax yet another assumption of the standard framework, which is the assumption that players have a correctly-specified view of the game they are playing. We hope that the introduction of misspecified models into an otherwise standard format can stimulate further development of ideas that have so far been studied under different, specific frameworks. Of course, we would like to emphasize that there are several, interesting aspects of bounded rationality that cannot be naturally viewed as the outcome of a (misspecified) learning process.[2]

Our paper is also related to the literature that shows that agents in a decision problem might optimally end up with incorrect beliefs if they are not patient enough to experiment (e.g., Rothschild (1974), McLennan (1984), Easley and Kiefer (1988)). A similar insight emerges in a game theoretic context via the notion of a self-confirming equilibrium (Battigalli (1987), Fudenberg and Levine (1993a), Dekel et al. (2004)), which requires players to have beliefs that are consistent with observed past play, though not necessarily correct when feedback is coarse. In our framework, equilibrium is equivalent to self-confirming equilibrium when we drop the assumption of full feedback but maintain the assumption that the subjective model is correctly specified.[3] More generally, we stress that misspecified learning can make beliefs endogenously incorrect even if there is full experimentation; in particular, behavior might be characterized as a fixed point (whereas behavior determines beliefs and beliefs, in turn, determine behavior) even in single-agent settings with full feedback.[4]

From a technical point of view, our results extend and combine results from two

[2]For example, there is a 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-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).

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 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]

In Section 2, we illustrate the equilibrium concept in the context of a simple example. We introduce the equilibrium framework in Section 3 and provide a learning foundation in Section 4. In Section 5, we illustrate the applicability of the framework with examples that come from a variety of different fields. We conclude in Section 6.

[5]See Fudenberg and Levine (1998, 2009) for a survey of this literature.

[6]Some 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).

2 Example: monopolist with unknown demand

The problem of a monopolist trying to learn the demand function was originally considered by Rothschild (1974) and McLennan (1984). The monopolist starts with a prior over a set possible demand functions that includes the true demand function– hence, the model is correctly specified. The monopolist faces a trade-off between exploitation and experimentation and one of the main insights is that an impatient monopolist might not learn the true demand function because it finds it optimal not to fully experiment. Here, we focus instead on the case where the monopolist has a misspecified demand model.

The following example serves four purposes. First, it illustrates, via a simple graphical approach, the notion of equilibrium proposed in this paper. Second, it illustrates that beliefs can be incorrect even if there is persistent experimentation. Third, it highlights a feature that cannot occur in correctly-specified settings: an agent's beliefs might endogenously depend on her own actions even if her own actions do not affect the amount of feedback that she obtains. In particular, equilibrium might be characterized as a fixed point even in a single-agent, full-feedback setting. Finally, given the fixed point nature of beliefs, we must allow the agent to follow mixed strategies to allow for the existence of equilibrium (and to rule out some types of non-convergent behavior).

The setup of the example is taken from Nyarko (1991). The monopolist chooses, at every period $t = 0, 1, ...$, a price $x_t \in \mathbb{X} = \{2, 10\}$ and then sells quantity y_t according to the demand function

$$y_t = a^0 - b^0 x_t + \varepsilon_t,$$

where $(\varepsilon_t)_t$ is an i.i.d. normally distributed process with mean zero and unit variance.[7] The monopolist observes sales y_t but it does not observe the random shocks ε_t; it does know, however, the distribution of $(\varepsilon_t)_t$. The monopolist has no costs of production and, therefore, her profits in period t are $\pi(x_t, y_t) = x_t y_t$. The monopolist wishes to maximize the discounted expected profits, where her discount factor is $\delta \in [0, 1)$.

The monopolist does not know the true demand intercept and slope (a^0, b^0). It starts with a prior μ_0 with full support over the set $\Theta \subset \mathbb{R}^2$ of parameters that it views as possible, where $\theta = (a, b)$ denotes a demand intercept and slope, and it

[7] As mentioned by Nyarko, sales can be negative with positive probability but the normal distribution is nevertheless chosen for simplicity.

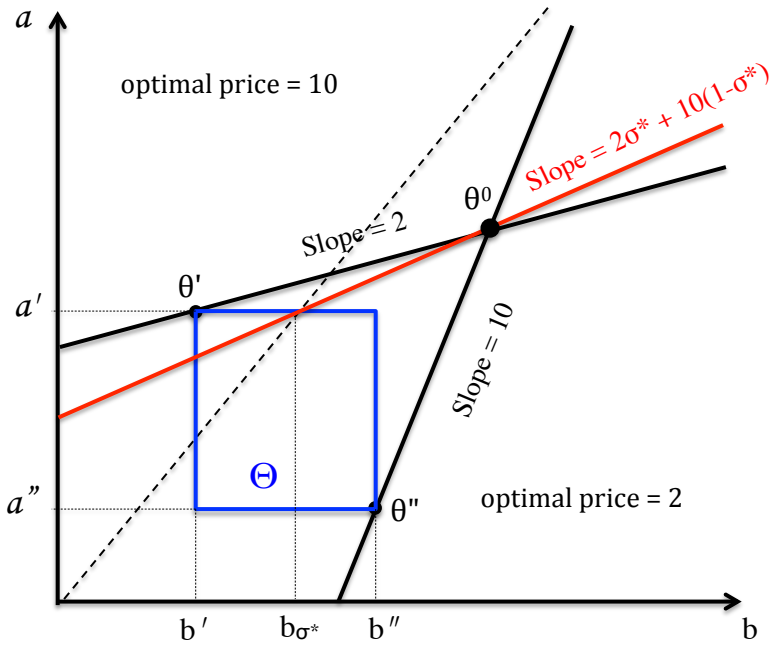

Figure 1: Monopolist with unknown demand

updates its prior using Bayes' rule. As is well known, the monopolist's problem can be represented recursively by a value function which is defined over the set of beliefs over Θ. Our objective is to characterize the limiting behavior of the monopolist.

Figure 1 shows the true demand parameter θ^0 and the set of parameters Θ, which is given by the rectangle with vertices at the points θ', θ'', (a', b'') and (a'', b'). In this case $\theta^0 \notin \Theta$ and, therefore, we say that the monopolist has a misspecified model. The dashed line separates the space of all parameters into two regions. If a parameter lies to the left of the dashed line, then the unique optimal price of a monopolist who is certain of that parameter is 10; if a parameter lies to the right, then the unique optimal price is 2; and if a parameter lies on the line, then the monopolist is indifferent between these two prices.

Nyarko (1991) shows that the monopolist's actions do not converge. To see the intuition behind this result, suppose that the monopolist were to always choose price 2. Then, it would observe that average sales are $a^0 - b^0 2$. It would eventually believe that any (a, b) that also gives rise to such average sales can explain the data. The set of all such (a, b)'s is given by the line with slope 2 passing through the true parameter θ^0. Moreover, $\theta' = (a', b')$ is the only point on that line that also belongs to Θ.

Therefore, the probability that the monopolist puts on any neighborhood of θ' will converge to 1. But θ' lies to the left of the dashed line, so that the monopolist would eventually strictly prefer to charge price 10, contradicting our initial assumption that it always charges 2. Similarly, if the monopolist were to always charge a price of 10, then it would eventually become very confident that the true parameter is θ'', but then, since θ'' is to the right of the dashed line, it would prefer to deviate and charge 2. Thus, the monopolist's behavior cycles forever between prices 2 and 10.

Non-convergent behavior, however, is due to the fact that actions are not continuous in beliefs. This feature is well known to possibly lead to non-convergence in correctly-specified settings. To avoid this issue, we extend Nyarko's example by allowing the monopolist to choose mixed strategies. Our interpretation and justification of mixed strategies will be standard and follows Harsanyi (1973) and Fudenberg and Kreps (1993).[8]

Figure 1 depicts the mixed strategy where the monopolist chooses price 2 with probability σ^*. If the monopolist chooses σ^*, then it is not too difficult to show that it will eventually become very confident that the true parameter is given by the point (a', b_{σ^*}) in Figure 1. This is the point on the set Θ that is closest to θ^0 when distance is measured along the line with slope $2\sigma^* + 10(1 - \sigma^*)$ passing through θ^*. If the monopolist were to be certain that the parameter is (a', b_{σ^*}), then, because (a', b_{σ^*}) lies on the dashed line, the monopolist would be indifferent between both prices and mixing would be optimal. A strategy such as σ^*, with the property that it is optimal given beliefs that are generated by playing it, is said to be an *equilibrium* strategy. This equilibrium strategy, which is unique in this example, represents the steady-state of this dynamic environment.

Next, we introduce a general framework, provide the definition of equilibrium, and finally show that this definition captures the steady-state of a corresponding dynamic learning environment.

[8] A similar point can be made by allowing the monopolist to choose price from the interval [2,10].

3 The framework

3.1 The environment and the equilibrium concept

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

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

where: I is a finite set of players; Ω is a finite set of payoff-relevant states; $\mathbb{S} = \times_{i \in I} \mathbb{S}^i$ is a finite set of profiles of signals, where $\mathbb{S}^i$ is the set of signals of player i; p is a full-support probability distribution over $\Omega \times \mathbb{S}$; $\mathbb{X} = \times_{i \in I} \mathbb{X}^i$ is a set of profiles of actions, where $\mathbb{X}^i$ is the finite set of actions of player i; $\mathbb{Y} = \times_{i \in I} \mathbb{Y}^i$ is a finite set of profiles of (observable) consequences, where $\mathbb{Y}^i$ is the set of consequences of player i; $f = (f^i)_{i \in I}$ is a profile of *feedback functions*, where the feedback function of player i, $f^i : \Omega \times \mathbb{X} \to \mathbb{Y}^i$, maps outcomes in $\Omega \times \mathbb{X}$ into consequences of player i; and $\pi = (\pi^i)_{i \in I}$, where $\pi^i : \mathbb{X}^i \times \mathbb{Y}^i \to \mathbb{R}$ is a bounded payoff function of player i. This environment includes the special case where there is a single player; we sometimes refer to that case as a *decision problem*, rather than a game.

The timing of the objective game is as follows: First, $(\omega, s) \in \Omega \times \mathbb{S}$ are drawn according to the probability distribution p. Second, each player i privately observes s^i. Third, each player i simultaneously chooses an action from $\mathbb{X}^i$, resulting in a profile of actions $x \in \mathbb{X}$. Fourth, each player i observes her own consequence $y^i = f^i(w, x) \in \mathbb{Y}^i$ and obtains payoff $\pi^i(x^i, y^i)$. This description of the game is standard; the explicit use of a feedback function is borrowed from the self-confirming equilibrium literature and allows us to capture situations where players might not get perfect feedback about the outcome of the game.[9]

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

For a fixed objective game, if we also fix a strategy profile σ, then we obtain an *objective distribution over player i's consequences*, Q^i_σ, where, for each $(s^i, x^i) \in$

[9]For simplicity, we implicitly assumes that players observe their own payoffs and actions.

$\mathbb{S}^i \times \mathbb{X}^i$, $Q^i_\sigma(\cdot \mid s^i, x^i) \in \Delta(\mathbb{Y}^i)$ is defined as follows:

$$Q^i_\sigma(y^i \mid s^i, x^i) = \sum_{\{(\omega, x^{-i}) : f^i(\omega, x^i, x^{-i}) = y^i\}} \sum_{s^{-i}} \prod_{j \neq i} \sigma^j(x^j \mid s^j) p(\omega, s^{-i} \mid s^i). \tag{1}$$

The objective distribution represents the true distribution over consequences given the objective game and a strategy profile followed by the players.

For a fixed objective game, a *subjective model* is a tuple

$$\mathcal{Q} = \langle \Theta, (Q_\theta)_{\theta \in \Theta} \rangle,$$

where: $\Theta = \times_{i \in I} \Theta^i$ is a parameter space, where Θ^i, the parameter set of player i, is a compact subset of a (finite-dimensional) Euclidean space; and $Q_\theta = (Q^i_{\theta^i})_{i \in I}$ is a profile of distributions, where $Q^i_{\theta^i}$ is a distribution over player i's consequences parameterized by $\theta^i \in \Theta^i$ that satisfies two properties: (i) $Q^i_{\theta^i}(y^i \mid s^i, x^i)$ is continuous as a function of θ^i and all $(y^i, s^i, x^i) \in \mathbb{Y}^i \times \mathbb{S}^i \times \mathbb{X}^i$; and (ii) for each player $i \in I$, there exists $\theta^i \in \Theta^i$ such that, for all $(s^i, x^i) \in \mathbb{S}^i \times \mathbb{X}^i$, $Q^i_{\theta^i}(y^i \mid s^i, x^i) > 0$ for all $y^i \in f^i(\Omega, x^i, \mathbb{X}^{-i})$. The second property rules out a stark form of misspecification by guaranteeing that there exists a parameter that can rationalize every feasible observation; otherwise, we would expect players to re-consider their subjective models.[10]

While the objective game represents the true environment, the subjective model represents the players' perception of their environment. Each player desires to learn the distribution over own consequences conditional on each private signal and on each action they can take. In particular, $\mathcal{Q}^i = (\Theta^i, (Q^i_{\theta^i})_{\theta^i \in \Theta^i})$ represents all the distributions, i.e., all the models of the world, that player i views as possible. This separation between objective and subjective models, which is often implicit in standard treatments of games, is crucial in this paper.

Remark. The subjective model is fairly general. In the standard case, it can arise from an underlying model of players' beliefs about (σ^{-i}, p), i.e., the strategies followed by other players and the distribution over $\Omega \times \mathbb{S}$. And, following Fudenberg and Levine (1993a), we allow players to have correlated beliefs about their opponents' strategies. But the subjective model also allows us to model situations where players do not even

[10] In terms of the dynamic model of Section 4, condition (ii) guarantees that Bayesian updating is always well defined. Also, for simplicity, this definition of a subjective model assumes that players know the distribution over their own signals.

understand that they are facing other players.

The equilibrium concept that we propose places two requirements on strategy profiles. The first requirement is that players optimize given some beliefs: A *strategy* σ^i *for player* i *is optimal given* $\mu^i \in \Delta(\Theta^i)$ if $\sigma^i(x^i \mid s^i) > 0$ implies that[11]

$$x^i \in \arg\max_{\bar{x}^i \in \mathbb{X}^i} E_{\bar{Q}^i_{\mu^i}(\cdot \mid s^i, \bar{x}^i)} \left[\pi^i(\bar{x}^i, Y^i)\right] \tag{2}$$

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

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

is defined as the *distribution over consequences of player* i *induced by* μ^i.

The second requirement places restrictions on equilibrium beliefs. In particular, beliefs must 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 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) distance between the objective distribution over i's consequences given $\sigma \in \Sigma$ and the distribution as parameterized by $\theta^i \in \Theta^i$:

$$K^i(\sigma, \theta^i) = \sum_{(s^i, x^i) \in \mathbb{S}^i \times \mathbb{X}^i} E_{Q^i_\sigma(\cdot \mid s^i, x^i)} \left[\ln\left(\frac{Q^i_\sigma(Y^i \mid s^i, x^i)}{Q^i_{\theta^i}(Y^i \mid s^i, x^i)}\right)\right] \sigma^i(x^i \mid s^i) p_{\mathbb{S}^i}(s^i), \tag{3}$$

where we use the convention that $-\ln 0 = \infty$. The *set of closest parameters of player* i *given* $\sigma \in \Sigma$ 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 σ.

Remark. 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

[11] The notation E_Q denotes expectation with respect to the probability measure Q.

when players are Bayesian.

Remark. 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.

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 $Q_{\theta^i}(\cdot \mid s^i, x^i) = Q^i_\sigma(\cdot \mid s^i, x^i)$ for all (s^i, x^i) such that $\sigma^i(x^i \mid s^i) > 0$. (ii) For every $i \in I$, $\Theta^i(\cdot)$ is non-empty, compact valued, and upper hemicontinuous.*

Proof. See the Appendix. □

Before stating the definition of equilibrium, we restrict the types of misspecified models that players are allowed to have. We say that a subjective model $\mathcal{Q}$ is *identifiable* if the following condition is satisfied for all $i \in I$ and $\sigma \in \Sigma$: 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 $(s^i, x^i) \in \mathbb{S}^i \times \mathbb{X}^i$ such that $\sigma^i(x^i \mid s^i) > 0$. As the examples throughout the paper illustrate, this condition is fairly mild and relatively easy to check.[12]

Identifiability says that if, for some player, two parameters are closest to the objective distribution given some strategy profile σ, then the distributions associated with these parameters must be indistinguishable given the feedback observed by the player. There are two reasons why a player might not be able to distinguish between two subjective distributions. The first is that she does not take a particular action and, therefore, fails to learn in some dimension. This is the key idea behind the notion of self-confirming equilibrium, where a player can entertain different beliefs, as long as these beliefs cannot be distinguished given her feedback. This first situation is permitted by our framework. The second reason why a player might not be able to distinguish between two distributions is that these distributions are different but equidistant (in terms of the wKLD function) to the objective distribution, in the sense that they provide an equally good fit of the data. In this case, beliefs might never

[12] We can weaken this condition by requiring that what players can identify are their expected profit functions, rather than distributions over consequences.

settle down. We rule out this knife-edge case by requiring the subjective model to be identifiable.[13]

We summarize the primitives that define the environment studied in this paper in the following definition.

Definition 1. A *(simultaneous-move) game* $\mathcal{G} =< \mathcal{O}, \mathcal{Q} >$ is composed of a (simultaneous-move) objective game $\mathcal{O}$ and a corresponding subjective model $\mathcal{Q}$ that is identifiable.

We propose the following solution concept.

Definition 2. A strategy profile $\sigma \in \Sigma$ is an *equilibrium of the game* $\mathcal{G}$ if, for all players $i \in I$, there exists $\mu^i \in \Delta(\Theta^i)$ such that

(i) σ^i is optimal given μ^i, and

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

Definition 2 places two standard types of restrictions on equilibrium behavior: (i) optimization given beliefs, and (ii) endogenous restrictions on beliefs.[14] It is insightful to compare this definition to the definition of a Nash equilibrium, which is identical to Definition 2 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.[15]

The following example illustrates the previous definitions in the context of a misspecification captured by (fully) cursed equilibrium (Eyster and Rabin, 2005) and analogy-based expectation equilibrium (Jehiel (2005), Jehiel and Koessler (2008)).[16]

[13] To illustrate, consider the following example by 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 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. The failure of identifiability is not robust, however, in the sense that, if the agent were to entertain the possibility of any parameter strictly between $1/4$ and $3/4$, the model would be identifiable and beliefs would converge to a unique parameter.

[14] 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).

[15] This is equivalent to the standard definition of Nash equilibrium and it allows for both strategic and payoff uncertainty (e.g., Vickrey, 1961). It is also what Harsanyi (1967-8) refers to as a Bayesian Nash equilibrium with a (correct) common prior, although we prefer to interpret the "prior" as an equilibrium belief, in the same manner as the belief over others' strategies is often interpreted as an equilibrium belief (see Esponda (2013) for further discussion).

[16] The example is taken from Fudenberg and Levine (2009, page 408).

Example. *Objective game.* There are two players, $I = \{1, 2\}$, and two states of the world, $\Omega = \{0, 1\}$. State 0 is chosen with probability $p_\Omega(1) = 2/3$. Players obtain perfect signals about the state: $\mathbb{S}^1 = \mathbb{S}^2 = \mathbb{S} = \{0, 1\}$, where $p_{\mathbb{S}|\Omega}(s \mid \omega) = 1$ for $s = \omega$. Player 1 and 2 choose actions from the set $\mathbb{X}^1 = \mathbb{X}^2 = \{0, 1\}$. Player 1 gets a payoff of 1 if he matches the action of player 2 and a payoff of 0 otherwise. Player 2 gets a payoff of 1 if she matches the state and zero otherwise. Player 1 receives feedback about the state and the action of player 2, $f^1(\omega, x) = (\omega, x^2) \in \mathbb{Y}^1 = \Omega \times \mathbb{X}^2$. Player 2 receives feedback about the state, $f^2(\omega, x) = \omega \in \mathbb{Y}^2 = \Omega$. A strategy for player i is a mapping $\sigma^i : \mathbb{S} \to \Delta(\mathbb{X}^i)$. Given a strategy profile σ, the objective distribution over consequences are $Q^1_\sigma(\omega, x^2 \mid s, x^1) = p_{\mathbb{S}|\Omega}(\omega \mid s)\sigma^2(x^2 \mid s)$ for player 1 and $Q^2_\sigma(\omega \mid s, x^2) = p_{\mathbb{S}|\Omega}(\omega \mid s)$ for player 2.

Subjective game. The players know that they obtain a perfect signal of the state. Player 2 has nothing to learn in this game. Player 1 wants to learn the strategy of player 2. He believes, perhaps incorrectly, that player 2's action does not depend on player 2's signal, $\sigma^2(1 \mid 1) = \sigma^2(1 \mid 0) = \theta$. Player 1 wants to learn the parameter $\theta \in [0, 1]$, which is his perceived probability that player 2 plays $x^2 = 1$, and his subjective model is determined by

$$Q^1_\theta(\omega, x^2) = p_{\mathbb{S}|\Omega}(\omega \mid s) \left(\theta x^2 + (1 - \theta)(1 - x^2)\right).$$

Analysis. The wKLD function of player 1 is

$$\begin{aligned} K^1(\sigma, \theta) &= \sum_{s \in \mathbb{S}} \sum_{x^2 \in \mathbb{X}^2} \ln \left(\frac{\sigma^2(x^2 \mid s)}{\theta x^2 + (1 - \theta)(1 - x^2)} \right) p_{\mathbb{S}}(s) \\ &= -\bar{\sigma}^2(1) \ln \theta - (1 - \bar{\sigma}^2(1)) \ln(1 - \theta) + C, \end{aligned}$$

where C is a constant term that does not depend on θ and $\bar{\sigma}^2(1) = \sum_s \sigma^2(1 \mid s) p_{\mathbb{S}}(s)$ is the average probability that player 2 plays $x^2 = 1$. The unique minimizer of $K^1(\sigma, \theta)$ is $\theta(\sigma) = \bar{\sigma}^2(1)$. Since it is dominant for player 2 to play $x^2 = 1$ if $s = 1$ and $x^1 = 0$ if $s = 0$, then $\bar{\sigma}^2(1) = p_{\mathbb{S}}(1) = 2/3$. Then, in equilibrium, player 1 must play $x^1 = 1$ irrespective of his signal because this is a best response to her belief that player 2 chooses $x^2 = 1$ with probability $2/3$ irrespective of her signal. In contrast, in a Nash equilibrium, player 1 would realize that player 2 plays differently in different states and would, therefore, best respond by matching his action to his own signal. □

3.2 Relationship to Nash equilibrium

To highlight the relationship between Nash equilibrium and our equilibrium concept, we introduce two properties of the environment. We say that a subjective model has *full feedback* if, for all i, there exists a set $\mathbb{Z}^i$, a function $g^i : \mathbb{X}^i \times \mathbb{Z}^i$, and probability distributions $P^i_{\theta^i}$ over $\mathbb{Z}^i$ for all $\theta^i \in \Theta^i$ with the properties that: (a) $g^i(x^i, z^i) = f^i(\omega, x^i, x^{-i})$ for all (x, ω, z^i), (b) $Q^i_{\theta^i}(y^i \mid s^i, x^i) = P_{\theta^i}(\{z^i : y^i = g(x^i, z^i) \mid s^i, x^i\})$ for all (s^i, x^i, y^i) and all θ^i, and (c) $P^i_{\theta^i}(z^i \mid s^i, x^i) = P^i_{\theta^i}(z^i \mid s^i, \hat{x}^i)$ for all (s^i, z^i), all $x^i \neq \hat{x}^i$, and all θ^i.[17] Conditions (a) and (b) simply say that there is an alternative, equivalent way to represent consequences by the space $\mathbb{Z}^i$ rather than $\mathbb{Y}^i$. Condition (c) requires that the probability distribution over the space of consequences does not depend on the player's own action. This condition is usually straightforward to check, since (c) is either satisfied for the original consequence space $\mathbb{Y}^i$ or for an equivalent representation of the consequence space. The full feedback condition implies that lack of experimentation can play no role in justifying incorrect equilibrium beliefs.[18]

Moreover, we say that a subjective model is *correctly specified in steady state* if for all $i \in I$ and $\sigma \in \Sigma$, there exists $\theta^i \in \Theta^i$ such that $Q^i_{\theta^i}(y^i \mid s^i, x^i) = Q^i_{\sigma}(y^i \mid s^i, x^i)$ for all $(s^i, x^i) \in \mathbb{S}^i \times \mathbb{X}^i$ and $y^i \in \mathbb{Y}^i$; otherwise, the subjective model is *misspecified in steady state*. In decision problems, where there is a single "player", the qualification "in steady state" is unnecessary and this definition coincides with the standard definition of misspecification in the statistics literature.[19]

The next result shows that Nash equilibrium is a special case of our solution concept that arises when the subjective model is correctly specified in steady state and there is full feedback.

[17] Alternatively, we can define full feedback by assuming that $f^i(\omega, x) \neq f^i(\omega', x')$ for all $(\omega, x) \neq (\omega', x')$ and then requiring that the subjective model be "consistent" with feedback function f^i and an underlying parameterized belief over the distribution $p \in \Delta(\Omega \times \mathbb{S})$ and over other players' strategies σ^{-i}. As mentioned earlier, we prefer to make assumptions directly on subjective model because it yields a more general framework.

[18] Arrow and Green (1973) defined a similar condition and restricted their entire setup to satisfy this condition.

[19] With more than one player, the qualification "in steady-state" is necessary because the subjective model of each player is always potentially misspecified in the dynamic model of Section 4 (since each player believes that they face a stationary environment, when in fact players' strategies might be changing over time).

Proposition 1. *Suppose that the subjective model has full feedback and is correctly specified in steady state. Then σ is an equilibrium if and only if it is a Nash equilibrium.*

Proof. By the assumption that the model is correctly specified in steady state, for all i, let $\theta^i_* \in \Theta^i$ be such that $Q^i_{\theta^i_*} = Q^i_\sigma$. In particular, $K^i(\theta^i_*, \sigma) = 0$. By Lemma 1(i), $K^i(\theta^i, \sigma) \geq 0$ for all $\theta^i \in \Theta^i$; therefore, $\theta^i_* \in \Theta^i(\sigma)$. Now suppose that σ is an equilibrium and consider any $\hat{\theta}^i \in \Theta^i(\sigma)$, where $\hat{\theta}^i \neq \theta^i_*$. Since $K^i(\theta^i_*, \sigma) = 0$, it must also be true that $K^i(\hat{\theta}^i, \sigma) = 0$. Lemma 1(i) then implies that $Q^i_{\hat{\theta}^i}(\cdot \mid s^i, x^i) = Q_{\theta^i_*}(\cdot \mid s^i, x^i)$ for all (s^i, x^i) such that $\sigma^i(x^i \mid s^i) > 0$. Moreover, the assumption of full feedback implies, without loss of generality, that, for all θ^i, $Q^i_{\theta^i}(\cdot \mid s^i, x^i) = Q^i_{\theta^i}(\cdot \mid s^i, \hat{x}^i)$ for all s^i and all $x^i \neq \hat{x}^i$. Thus, $Q^i_{\hat{\theta}^i} = Q^i_{\theta^i_*} = Q^i_\sigma$. and, since the equality holds for all $\hat{\theta}^i \in \Theta^i(\sigma)$, it follows that $\bar{Q}^i_{\mu^i} = Q^i_\sigma$ for all $\mu^i \in \Delta(\Theta^i(\sigma))$. Since σ is optimal given some $\mu^i \in \Delta(\Theta^i(\sigma))$, it then follows that σ must be optimal given the objective distribution Q^i_σ, which is the definition of a Nash equilibrium. For the converse result, suppose that σ is a Nash equilibrium. Then, by definition, σ is optimal given Q^i_σ. By the assumption that the model is correctly specified, for all i, there exists $\theta^i_* \in \Theta^i$ such that $Q^i_{\theta^i_*} = Q^i_\sigma$. Moreover, by the same argument as before, $\theta^i_* \in \Theta^i(\sigma)$. Then σ is also optimal given $Q^i_{\theta^i_*}$ and, since $\theta^i_* \in \Theta^i(\sigma)$, it follows that σ is an equilibrium. □

Remark. Proposition 1 can be particularly useful because a subjective model can be correctly specified in steady state even if the model is inherently wrong. In this case, a wrong model of the world is inconsequential. For example, a firm might incorrectly believe that its demand function does not depend on the price of a competitor, but it might nevertheless be able to learn the correct consequences of its price choices in equilibrium (see Sections 5.2 and 5.4 for examples).

If we relax the assumption of full feedback but maintain that the subjective model is correctly specified, then Definition 2 is equivalent to the definition of a (unitary) self-confirming equilibrium (Dekel et al., 2004). It is well known that a Nash equilibrium is always a self-confirming equilibrium (though the converse is not necessarily true). Thus, if the subjective model is correctly specified in steady-state, then a Nash equilibrium is also an equilibrium according to Definition 2.[20] But, if the subjective

[20]For a formal proof, notice that the relevant direction of Proposition 1 does not rely on the full feedback assumption.

model is misspecified in steady-state, then a Nash equilibrium is not necessarily an equilibrium according to Definition 2, as illustrated by several examples in the paper. The novelty of our equilibrium concept consists of relaxing the assumption that players have a correctly-specified subjective model.[21]

Unfortunately, existence of equilibrium cannot be established using standard methods when the subjective model is misspecified in steady-state.[22] It is straightforward to generalize the notion of a best response by letting

$$BR^i(\sigma) = \left\{\hat{\sigma} \in \Sigma : \forall i \in I,\, \exists \mu^i \in \Delta(\Theta^i(\sigma)) \text{ such that } \hat{\sigma}^i \text{ is optimal given } \mu^i\right\} \quad (4)$$

and then showing that σ is an equilibrium if and only if it is a fixed point of $BR = (BR^i)_{i \in I}$. The problem, however, is that BR is not necessarily convex-valued because there might be multiple beliefs, each of which justifies a different best response.[23]This issue is also present for self-confirming equilibrium, but that literature can sidestep this problem because existence is easily established by noting that a Nash equilibrium is always a self-confirming equilibrium. In the next subsection, we directly tackle the existence problem by studying a perturbed game. These perturbations also turn out to be important for the learning analysis of Section 4.

3.3 Perturbed game and existence of equilibrium

In this section, we introduce a perturbed game and then use it to show existence of equilibrium of the unperturbed game defined in the previous section. We first perturb the payoffs of the game and establish that equilibrium exists under a class of perturbations. We then consider a sequence of equilibria of perturbed games where the perturbations go to zero and establish that the limit is an equilibrium of the

[21]Arrow and Green (1973) impose a condition that requires the subjective model to be "correctly specified" but only on the equilibrium path. This is equivalent to the requirement that the wKLD function 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.

[22]When the subjective model is correctly specified in steady-state, i.e., in a self-confirming equilibrium, then existence follows easily from existence of a Nash equilibrium.

[23]For example, fix σ and suppose that $\Theta^i(\sigma) = \{\theta_1^i, \theta_2^i\}$. Then it is possible that σ_1^i is optimal if player i puts probability 1 on θ_1^i and that σ_2^i is optimal if she puts probability 1 on θ_2^i, but that a convex combination of σ_1^i and σ_2^i is not optimal for any belief satisfying $\mu^i \in \Delta(\Theta^i(\sigma))$.

unperturbed game.[24]

A *perturbation structure* is a tuple $\mathcal{P} = \langle \mathbb{V}, P_{\mathbb{V}} \rangle$ where: $\mathbb{V} = \times_{i \in I} \mathbb{V}^i$ and $\mathbb{V}^i \subseteq \mathbb{R}^{|\mathbb{X}^i|}$ is a set of payoff perturbations for each action of player i; $P_{\mathbb{V}} = (P_{\mathbb{V}^i})_{i \in I}$, where $P_{\mathbb{V}^i} \in \Delta(\mathbb{V}^i)$ is a distribution over payoff perturbations of player i that is absolutely continuous with respect to the Lebesgue measure, satisfies $\int_{\mathbb{V}^i} ||\eta^i|| P_{\mathbb{V}^i}(d\eta^i) < \infty$, and is independent from the perturbations of other players.

Definition 3. A *perturbed game* $\mathcal{G}^{\mathcal{P}} = \langle \mathcal{G}, \mathcal{P} \rangle$ is composed of a game $\mathcal{G}$ and a perturbation structure $\mathcal{P}$. A *fully-perturbed game* is a perturbed game where the support of $P_{\mathbb{V}^i}$ is $\mathbb{R}^{|\mathbb{X}^i|}$ for all $i \in I$.

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

A *strategy σ^i for player i is optimal for the perturbed game given $\mu^i \in \Delta(\Theta^i)$* if, for all $(s^i, x^i) \in \mathbb{S}^i \times \mathbb{X}^i$,[25]

$$\sigma^i(x^i \mid s^i) = P_{\mathbb{V}^i}\left(\eta^i : x^i \in \arg\max_{\bar{x}^i \in \mathbb{X}^i} E_{\bar{Q}^i_{\mu^i}(\cdot|s^i,\bar{x}^i)}\left[\pi^i(\bar{x}^i, Y^i)\right] + \eta^i(\bar{x}^i)\right). \tag{5}$$

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

Games that are not just perturbed but also fully perturbed have the following convenient properties.

Lemma 2. *(i) If σ^i is an optimal strategy of player i in a fully-perturbed game, then $\sigma^i(x^i \mid s^i) > 0$ for all $(s^i, x^i) \in \mathbb{S}^i \times \mathbb{X}^i$; (ii) For all $\sigma \in \Sigma$, if $\sigma^i(x^i \mid s^i) > 0$ for*

[24]The idea of perturbations and the strategy of the existence proof dates back to Harsanyi (1973); Selten (1975) and Kreps and Wilson (1982) also used these ideas to prove existence of perfect and sequential equilibrium, respectively.

[25]Equation (5) is well defined by the assumption of absolute continuity of perturbations and the fact that the set of η's such that a player is indifferent between any two actions lies in a lower-dimensional hyperplane.

all $(s^i, x^i) \in \mathbb{S}^i \times \mathbb{X}^i$ *and some* $i \in I$*, then* $\bar{Q}^i_{\mu^i_1}(\cdot \mid s^i, x^i) = \bar{Q}^i_{\mu^i_2}(\cdot \mid s^i, x^i)$ *for all* $(s^i, x^i) \in \mathbb{S}^i \times \mathbb{X}^i$ *and all* $\mu^i_1, \mu^i_2 \in \Delta(\Theta(\sigma))$.

Proof. See the Appendix. □

The first claim in Lemma 2 says that, in fully-perturbed games, optimality implies that all actions are chosen with positive probability. The second claim then follows by the assumption that the subjective model is identifiable.

The definition of equilibrium of a perturbed game $\mathcal{G}^{\mathcal{P}}$ is analogous to Definition 2, with the obvious difference that (i) is replaced by: (i') σ^i is optimal *for the perturbed game* given μ^i. The next result establishes existence of equilibrium for fully-perturbed games.

Theorem 1. *Every fully-perturbed game has an equilibrium.*

Proof. See the Appendix. □

The proof of Theorem 1 follows by characterizing equilibrium as a fixed point of a continuous function and then applying Brouwer's fixed point theorem. Continuity is obtained from identifiability, the upper hemicontinuity of $\Theta^i(\cdot)$, continuity of $Q^i_{\theta^i}$ as a function of θ^i, and the absolute continuity of $P_{\mathbb{V}^i}$.

We now consider sequences of perturbed games where the payoff perturbations go to zero.

Definition 4. A *sequence of vanishing [fully] perturbed games* is a sequence of [fully] perturbed games where each game in the sequence shares the same primitives, except possibly for the perturbation structure $\langle \mathbb{V}_\xi, P_{\mathbb{V}_\xi} \rangle$, where $\xi \in \mathbb{N}$ indexes the element of the sequence, and where, for all $i \in I$, (η^i_ξ) converges in probability to 0 as $\xi \to \infty$, i.e., for all $\varepsilon > 0$,

$$\lim_{\xi \to \infty} P_{\mathbb{V}_\xi{}^i} \left(\left\| \eta^i_\xi \right\| \geq \varepsilon \right) = 0. \tag{6}$$

The next result applies standard continuity arguments.

Theorem 2. *Fix a sequence of vanishing perturbed games and a corresponding sequence* $(\sigma_\xi)_\xi$ *of equilibria such that* $\lim_{\xi\to\infty}\sigma_\xi = \sigma$. *Then* σ is an equilibrium of the (unperturbed) game.

Proof. See the Appendix. □

Existence of equilibrium of (unperturbed) games follows as a corollary of the previous results.

Corollary 1. *Every (unperturbed) game has an equilibrium.*

Proof. Fix a sequence of vanishing fully-perturbed games where the (unperturbed) game remains fixed.[26] By Theorem 1, there exists a corresponding sequence of equilibria. Since equilibria live in a compact space, then there exists a subsequence of equilibria that converges. Theorem 2 says that this limit point is an equilibrium of the (unperturbed) game.

□

4 Learning foundation for equilibrium

In this section, we provide a learning foundation for the notion of equilibrium defined in Section 3. We do so in the context of perturbed games because, as highlighted by Fudenberg and Kreps (1993), behavior need not be continuous in beliefs in an unperturbed game. Thus, even if beliefs were to converge, behavior would not necessarily settle down in an unperturbed game. One important role of the perturbations is to make sure that if beliefs converge then behavior also converges.

Throughout this section, we fix a perturbed game $\mathcal{G}^{\mathcal{P}}$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, (ω_t, s_t), are independently drawn from the same distribution p every period. In addition, each player i has a prior μ_0^i with full support over her (finite-dimensional) parameter space, Θ^i.[27] At the

[26]For example, suppose that for every player and action, each perturbation is drawn independently from a normal distribution with mean zero and variance that goes to zero.

[27]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 a decision problem with a correctly-specified model and full feedback (Freedman (1963), Diaconis and Freedman (1986)).

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 maximize discounted expected payoffs, where $\delta^i \in [0,1)$ is the discount factor of player i. In particular, players can be forward looking and decide to experiment. Players believe, however, that they face a stationary environment and, therefore, have no incentives to influence the future behavior of other players.[28] Finally, we assume for simplicity that players know the distribution of their own payoff perturbations.

Let $B^i : \Delta(\Theta^i) \times \mathbb{S}^i \times \mathbb{X}^i \times \mathbb{Y}^i \to \Delta(\Theta^i)$ denote the Bayesian operator of player i: for all $A \subseteq \Theta$ Borel measurable and all (μ^i, s^i, x^i, y^i),

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

Because players believe that they face a stationary environment, they solve a (subjective) dynamic optimization problem that can be cast recursively as follows. By the Principle of Optimality, $V^i(\mu^i, s^i)$ is a function that denotes the maximum expected discounted payoffs (i.e., the value function) of player i if she starts a period by observing signal s^i and by holding belief μ^i if and only if

$$V^i(\mu^i, s^i) = \int_{\mathbb{V}^i} \left\{ \max_{x^i \in \mathbb{X}^i} E_{\bar{Q}_{\mu^i}(\cdot|s^i,x^i)} \left[\pi^i(x^i, Y^i) + \eta^i(x^i) + \delta E_{p_{\mathbb{S}^i}} \left[V^i(\hat{\mu}^i, S^i) \right] \right] \right\} P_{\mathbb{V}^i}(d\eta^i), \tag{7}$$

where $\hat{\mu}^i = B^i(\mu^i, s^i, x^i, Y^i)$ is the updated belief. For all (μ^i, s^i, η^i), let

$$\Phi^i(\mu^i, s^i, \eta^i) = \arg\max_{x^i \in \mathbb{X}^i} E_{\bar{Q}_{\mu^i}(\cdot|s^i,x^i)} \left[\pi^i(x^i, Y^i) + \eta^i(x^i) + \delta E_{p_{\mathbb{S}^i}} \left[V^i(\hat{\mu}^i, S^i) \right] \right].$$

We use standard arguments to prove the following properties of the value function and optimal solutions.[29]

Lemma 3. *There exists a unique solution V^i to the Bellman equation (7); this solution is bounded in $\Delta(\Theta^i) \times \mathbb{S}^i$ and continuous as a function of μ^i. Moreover, Φ^i is single-valued and continuous with respect to μ^i, a.s.- $P_{\mathbb{V}^i}$.*

[28] Kalai and Lehrer (1993) study learning in the repeated game and Fudenberg and Levine (1993b) and Fudenberg and Takahashi (2011) look at large populations where the "no influence" assumption is more accurate.

[29] Doraszelski and Escobar (2010) study a similarly perturbed version of the Bellman equation.

Proof. See the Appendix. □

Without loss of generality, we restrict behavior to depend on the state of the recursive problem.

Definition 5. A *policy of player i is a sequence of functions $\phi^i = (\phi^i_t)_t$, where $\phi^i_t : \Delta(\Theta^i) \times \mathbb{S}^i \times \mathbb{V}^i \to \mathbb{X}^i$. A policy ϕ^i is optimal if $\phi^i_t \in \Phi$ for all t.* A policy profile $\phi = (\phi^i)_{i \in I}$ *is optimal* if ϕ^i is optimal for all $i \in I$.

Let $\mathbb{H} = (\mathbb{S} \times \Omega \times \text{graph}\,(\times_{i \in I} f^i(\Omega, \cdot)))^\infty$ denote the set of infinite (feasible) histories, where any history $h = (s_0, \eta_0, x_0, y_0, ..., s_t, \eta_t, x_t, y_t ...) \in \mathbb{H}$ must satisfy the feasibility restriction that $y_t \in \times_{i \in I} f^i(\Omega, x_t)$ for all t. Let $\mathbf{P}^{\mu_0, \phi}$ denote the (objective) probability distribution over $\mathbb{H}$ 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 $(\mu_t)_t$ denote the *sequence of beliefs* $\mu_t : \mathbb{H} \to \times_{i \in I} \Delta(\Theta^i)$ such that, for all $t \geq 1$ and all $i \in I$, μ^i_t is the posterior at time t defined recursively by $\mu^i_t(h) = B(\mu^i_{t-1}(h), s^i_{t-1}(h), x^i_{t-1}(h), y^i_{t-1}(h))$ for all $h \in \mathbb{H}$.

Definition 6. The *sequence of intended strategy profiles given policy profile* $\phi = (\phi^i)_{i \in I}$ is the sequence $(\sigma_t)_t$ of random variables $\sigma_t : \mathbb{H} \to \times_{i \in I} \Delta(\mathbb{X}^i)^{\mathbb{S}^i}$ such that, for all t and all $i \in I$,

$$\sigma^i_t(h)(x^i \mid s^i) = P_{\mathbb{V}^i}\left(\eta^i : \phi^i_t(\mu^i_t(h), s^i, \eta^i) = x^i\right). \tag{8}$$

An intended strategy profile σ_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, μ_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.

Definition 7. A strategy profile $\sigma \in \Sigma$ is *stable [or strongly stable] under policy profile* ϕ if the sequence of intended strategies, $(\sigma_t)_t$, converges to σ with positive probability [or with probability one], i.e.,

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

The next result extends results from the statistics of misspecified learning (Berk (1966), Bunke and Milhaud (1998)) to establish that, if behavior stabilizes to a strategy profile σ, then the support of the posterior beliefs converges to $\Theta^i(\sigma)$ for each player i. The proof clarifies the origin of the wKLD function in the definition of equilibrium in Section 3.

Lemma 4. *Suppose that, for a policy profile ϕ, the sequence of intended strategies, $(\sigma_t)_t$, converges to σ for all histories in a set $\mathcal{H} \subseteq \mathbb{H}$ such that $\mathbf{P}^{\mu_0,\phi}(\mathcal{H}) > 0$. Then, for all open sets $U^i \supseteq \Theta^i(\sigma)$,*

$$\lim_{t\to\infty} \mu_t^i\left(U^i\right) = 1$$

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

Proof. It is sufficient to establish that $\lim_{t\to\infty} \int_{\Theta^i} d^i(\sigma,\theta^i)\mu_{t+1}^i(d\theta^i) = 0$ a.s. in $\mathcal{H}$, where $d^i(\sigma,\theta^i) = \inf_{\hat{\theta}^i\in\Theta^i(\sigma)} \left\|\theta^i - \hat{\theta}^i\right\|$. Fix $i \in I$ and $h \in \mathbb{H}$. Then

$$\begin{aligned}
\int_{\Theta^i} d^i(\sigma,\theta^i)\mu_{t+1}^i(d\theta^i) &= \frac{\int_{\Theta^i} d^i(\sigma,\theta^i)\prod_{\tau=1}^t Q_{\theta^i}^i(y_\tau^i \mid s_\tau^i,x_\tau^i)\mu_0^i(d\theta^i)}{\int_{\Theta^i}\prod_{\tau=1}^t Q_{\theta^i}^i(y_\tau^i \mid s_\tau^i,x_\tau^i)\mu_0^i(d\theta^i)} \\
&= \frac{\int_{\Theta^i} d^i(\sigma,\theta^i)\prod_{\tau=1}^t \frac{Q_{\theta^i}^i(y_\tau^i|s_\tau^i,x_\tau^i)}{Q_{\sigma_\tau}^i(y_\tau^i|s_\tau^i,x_\tau^i)}\mu_0^i(d\theta^i)}{\int_{\Theta^i}\prod_{\tau=1}^t \frac{Q_{\theta^i}^i(y_\tau^i|s_\tau^i,x_\tau^i)}{Q_{\sigma_\tau}^i(y_\tau^i|s_\tau^i,x_\tau^i)}\mu_0^i(d\theta^i)} \\
&= \frac{\int_{\Theta^i} d^i(\sigma,\theta^i)\exp\left\{tK_t^i(h,\theta^i)\right\}\mu_0^i(d\theta^i)}{\int_{\Theta^i}\exp\left\{tK_t^i(h,\theta^i)\right\}\mu_0^i(d\theta^i)},
\end{aligned}$$

where the first line is well-defined because the definition of a subjective model (i.e., properties (i) and (ii)), where the second line is well-defined because $\mathbf{P}^{\mu_0,\phi}(\mathcal{H}) > 0$ implies that all the terms we divide by are positive, and where we define $K_t^i(h,\theta^i) = -\frac{1}{t}\sum_{\tau=1}^t \ln\frac{Q_{\sigma_\tau}^i(y_\tau^i|s_\tau^i,x_\tau^i)}{Q_{\theta^i}^i(y_\tau^i|s_\tau^i,x_\tau^i)}$.[30] Then, for all $\varepsilon > 0$ and $\eta > 0$,

$$\int_{\Theta^i} d^i(\sigma,\theta^i)\mu_{t+1}^i(d\theta^i) \leq \varepsilon + C\frac{A_t^i(h,\varepsilon)}{B_t^i(h,\eta)}, \tag{9}$$

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

$$A_t^i(h,\varepsilon) = \int_{\{\theta^i : d^i(\sigma,\theta^i)\geq\varepsilon\}} \exp\left\{tK_t^i(h,\theta^i)\right\}\mu_0^i(d\theta^i)$$

[30] If, for some θ^i, $Q_{\theta^i}^i(y_\tau^i \mid s_\tau^i,x_\tau^i) = 0$ for some $\tau \in \{1,...,t\}$, then we define $K_t^i(h,\theta^i) = -\infty$ and $\exp\left\{tK_t^i(h,\theta^i)\right\} = 0$.

and

$$B_t^i(h,\eta) = \int_{\{\theta^i : d^i(\sigma,\theta^i)\leq\eta\}} \exp\left\{tK_t^i(h,\theta^i)\right\}\mu_0^i(d\theta^i).$$

In particular, expression (9) is well defined because the assumption that the priors have full support implies that $B_t^i(h,\eta) > 0$.[31] The proof concludes by showing that for every (sufficiently small) $\varepsilon > 0$, there exists $\eta_\varepsilon > 0$ such that $\lim_{t\to\infty} A_t^i(h,\varepsilon)/B_t^i(h,\eta_\varepsilon) = 0$. This result is achieved in several steps. First, we show in the Online Appendix, Section A, that

$$\lim_{t\to\infty} K_t^i(h,\theta^i) = -K^i(\sigma,\theta^i) \tag{10}$$

for all $\theta^i \in \Theta^i$, a.s. in $\mathcal{H}$. Next, fix $\varepsilon > 0$. Recall that $K^i(\sigma,\theta^i)$ denotes the wKLD function (see equation 3) and define $K_\varepsilon^i(\sigma) = \inf\{K^i(\sigma,\theta^i) \mid \theta^i \in \Theta^i,\ d^i(\sigma,\theta^i) \geq \varepsilon\}$ and $\alpha_\varepsilon = \left(K_\varepsilon^i(\sigma) - K_0^i(\sigma)\right)/3$. 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, fix $\varepsilon \leq \bar{\varepsilon}$. It follows that

$$K^i(\sigma,\theta^i) > K_0^i(\sigma) + 2\alpha_\varepsilon \tag{11}$$

for all θ^i such that $d^i(\sigma,\theta^i) \geq \varepsilon$. Also, by continuity of $K^i(\sigma,\cdot)$, we can find $\eta_\varepsilon > 0$ such that

$$K^i(\sigma,\theta^i) < K_0^i(\sigma) + \alpha_\varepsilon/2 \tag{12}$$

for all θ^i such that $d^i(\sigma,\theta^i) \leq \eta_\varepsilon$. Then

$$\begin{aligned}
\liminf_{t\to\infty} B_t^i(h,\eta_\varepsilon)\exp\left\{t(K_0^i(\sigma)+\frac{\alpha_\varepsilon}{2})\right\} &= \liminf_{t\to\infty}\int_{\{\theta^i : d^i(\sigma,\theta^i)\leq\eta_\varepsilon\}} \exp\left\{t(K_0^i(\sigma)+\frac{\alpha_\varepsilon}{2}+K_t^i(h,\theta^i))\right\}\mu_0^i(d\theta^i) \\
&\geq \int_{\{\theta^i : d^i(\sigma,\theta^i)\leq\eta_\varepsilon\}} \exp\left\{\lim_{t\to\infty} t(K_0^i(\sigma)+\frac{\alpha_\varepsilon}{2}-K^i(\sigma,\theta^i))\right\}\mu_0^i(d\theta^i) \\
&= \infty,
\end{aligned}$$

a.s.-$\mathbf{P}^{\mu_0,\phi}$, where the second line follows from Fatou's Lemma and (10), and the third line follows from (12).

[31] Note that this result is also true if we replace the assumption that the priors have full support with the assumption that $\mu_0^i(\Theta^i(\sigma)) > 0$ and $\Theta^i(\sigma) \neq \Theta^i$.

Finally, it suffices to show

$$\lim_{t\to\infty} A_t^i(h,\varepsilon)\exp\left\{t(K_0^i(\sigma)+\alpha_\varepsilon)\right\} = \lim_{t\to\infty}\int_{\{\theta^i:d^i(\sigma,\theta^i)\geq\varepsilon\}} \exp\left\{t(K_0^i(\sigma)+\alpha_\varepsilon+K_t^i(h,\theta^i))\right\}\mu_0^i(d\theta^i)$$
$$= 0, \tag{13}$$

a.s.-$\mathbf{P}^{\mu_0,\phi}$. The intuition is that equation (13) follows from (10) and (11); this result, however, requires a rather tedious proof because it is not obvious that the the limit and the integral can be interchanged in equation (13). We prove equation (13) in the Online Appendix, Section A. □

Lemma 4 only implies that the *support* of posteriors converges, but posteriors need not converge.[32] We can always find, however, a subsequence of posteriors that converges. By continuity of behavior in beliefs, the stable strategy profile is dynamically optimal (in the sense of solving the dynamic optimization problem) given this convergent posterior. By the assumption that the subjective model is identifiable, the convergent posterior is a fixed point of the Bayesian operator. Thus, the players understand that their limiting strategies will provide no new information. Since the value of experimentation is non-negative, it follows that the stable strategy profile must also be myopically optimal (in the sense of solving the optimization problem that ignores the future), which is the definition of optimality used in the equilibrium model of Section 3. Thus, we obtain the following characterization of the set of stable strategy profiles when players follow optimal policies.

Theorem 3. *If $\sigma\in\Sigma$ is stable under an optimal policy profile, then σ is an equilibrium of the perturbed game.*

Proof. Let ϕ denote the optimal policy function under which σ is stable. By Lemma 4, 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 $\lim_{t\to\infty}\mu_t^i(U^i)=1$ 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 $(\mu^i_{t(j)})_j$, such that $\mu^i_{t(j)}\to\mu^i_\infty$ (the limit could

[32]One obvious way of obtaining convergence of player i's posteriors is to restrict attention to settings where $\Theta^i(\sigma)$ is a singleton for all σ. But this assumption rules out some interesting cases of multiplicity, such as those captured by the idea of self-confirming equilibrium. Instead, we rely on the weaker assumption of identifiability of subjective models.

depend on h). We now show that $\mu^i_\infty \in \Delta(\Theta^i)$. Suppose not, so that there exists $\hat{\theta}^i \in supp(\mu^i_\infty)$ such that $\hat{\theta}^i \notin \Theta^i(\sigma)$. Then, since $\Theta^i(\sigma)$ is closed (by Lemma 1(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 $\mu^i_\infty\left(\bar{U}^i\right) \geq \lim_{t\to\infty} \mu^i_t\left(\bar{U}^i\right) \geq \lim_{t\to\infty} \mu^i_t\left(U^i\right) = 1$.

Given that $\lim_{j\to\infty} \sigma_{t(j)} = \sigma$ and $\mu^i_\infty \in \Delta(\Theta^i(\sigma))$ for all i, it remains to show that, for all i, σ^i is optimal for the perturbed game given $\mu^i_\infty \in \Delta(\Theta^i)$, *i.e.,* for all (s^i, x^i),

$$\sigma^i(x^i \mid s^i) = P_{\mathbb{V}^i}\left(\eta^i : \psi^i(\mu^i_\infty, s^i, \eta^i) = \{x^i\}\right), \tag{14}$$

where $\psi^i(\mu^i_\infty, s^i, \eta^i) \equiv \arg\max_{x^i \in X^i} E_{\bar{Q}^i_{\mu^i_\infty}(\cdot|s^i,x^i)}\left[\pi^i(x^i, Y^i)\right] + \eta^i(x^i)$.

To establish (14), fix $i \in I$ and $s^i \in \mathbb{S}^i$. Then

$$\begin{aligned}\lim_{j\to\infty} \sigma^i_{t(j)}(h)(x^i|s^i) &= \lim_{j\to\infty} P_{\mathbb{V}^i}\left(\eta^i : \phi^i_{t(j)}(\mu^i_{t(j)}, s^i, \eta^i) = x^i\right) \\ &= P_{\mathbb{V}^i}\left(\eta^i : \Phi^i(\mu^i_\infty, s^i, \eta^i) = \{x^i\}\right),\end{aligned}$$

where the second line follows by optimality of ϕ^i and Lemma 3. This implies that $\sigma^i(x^i|s^i) = P_{\mathbb{V}^i}\left(\eta^i : \Phi^i(\mu^i_\infty, s^i, \eta^i) = \{x^i\}\right)$. Thus, it remains to show that

$$P_{\mathbb{V}^i}\left(\eta^i : \Phi^i(\mu^i_\infty, s^i, \eta^i) = \{x^i\}\right) = P_{\mathbb{V}^i}\left(\eta^i : \psi^i(\mu^i_\infty, s^i, \eta^i) = \{x^i\}\right) \tag{15}$$

for all x^i such that $P_{\mathbb{V}^i}\left(\eta^i : \Phi^i(\mu^i_\infty, s^i, \eta^i) = \{x^i\}\right) > 0$. From now on, fix any such x^i. Since $\sigma^i(x^i \mid s^i) > 0$, the assumption that the subjective model is identifiable implies that $Q^i_{\theta^i_1}(\cdot \mid x^i, s^i) = Q^i_{\theta^i_2}(\cdot \mid x^i, s^i)$ for all $\theta_1, \theta_2 \in \Theta(\sigma)$. The fact that $\mu^i_\infty \in \Delta(\Theta^i(\sigma))$ then implies that

$$B^i(\mu^i_\infty, s^i, x^i, y^i) = \mu^i_\infty \tag{16}$$

for all $y^i \in \mathbb{Y}^i$. Thus, $\Phi^i(\mu^i_\infty, s^i, \eta^i) = \{x^i\}$ is equivalent to

$$\begin{aligned}&E_{\bar{Q}_{\mu^i_\infty}(\cdot|s^i,x^i)}\left[\pi^i(x^i, Y^i) + \eta^i(x^i) + \delta E_{p_{\mathbb{S}^i}}\left[V^i(\mu^i_\infty, S^i)\right]\right] \\ &\quad > E_{\bar{Q}_{\mu^i_\infty}(\cdot|s^i,\tilde{x}^i)}\left[\pi^i(\tilde{x}^i, Y^i) + \eta^i(\tilde{x}^i) + \delta E_{p_{\mathbb{S}^i}}\left[V^i(B^i(\mu^i_\infty, s^i, \tilde{x}^i, Y^i), S^i)\right]\right] \\ &\quad \geq E_{\bar{Q}_{\mu^i_\infty}(\cdot|s^i,\tilde{x}^i)}\left[\pi^i(\tilde{x}^i, Y^i) + \eta^i(\tilde{x}^i)\right] + \delta E_{p_{\mathbb{S}^i}}\left[V^i(E_{\bar{Q}_{\mu^i_\infty}(\cdot|s^i,\tilde{x}^i)}\left[B^i(\mu^i_\infty, s^i, \tilde{x}^i, Y^i)\right], S^i)\right] \\ &\quad = E_{\bar{Q}_{\mu^i_\infty}(\cdot|s^i,\tilde{x}^i)}\left[\pi^i(\tilde{x}^i, Y^i) + \eta^i(\tilde{x}^i)\right] + \delta E_{p_{\mathbb{S}^i}}\left[V^i(\mu^i_\infty, S^i)\right]\end{aligned}$$

for all $\tilde{x}^i \in \mathbb{X}^i$, where the first line follows by equation (16) and definition of Φ, the

second line follows by the convexity[33] of V^i as a function of μ^i and Jensen's inequality, and the last line by the fact that Bayesian beliefs have the martingale property. In turn, the above expression is equivalent to $\psi(\mu^i_\infty, s^i, \eta^i) = \{x^i\}$. □

Theorem 3 provides our main justification for focusing on equilibria of perturbed games: any strategy profile that is not an equilibrium cannot represent the limiting behavior of optimizing players. Theorem 3, however, does not imply that behavior will stabilize in a perturbed game. In fact, we know that there are cases where optimal behavior will not converge to Nash equilibrium, which is a special case of the equilibrium concept in this paper.[34] Thus, some assumption needs to be relaxed in order to prove convergence for general games.

In the remaining of this section, we adapt an idea due to Fudenberg and Kreps (1993) and provide a sort of converse to Theorem 3. The following definition relaxes optimality by allowing players to make optimization mistakes that nevertheless vanish with time.

Definition 8. A *policy profile ϕ is asymptotically optimal if there exists a sequence $(\varepsilon_t)_t$ with $\lim_{t\to\infty} \varepsilon_t = 0$ such that, for all $i \in I$, all $(\mu^i, s^i, \eta^i) \in \Delta(\Theta^i) \times \mathbb{S}^i \times \mathbb{V}^i$, and all t,*

$$U^i(\mu^i, s^i, \eta^i, \phi^i_t(\mu^i, s^i, \eta^i)) \geq U^i(\mu^i, s^i, \eta^i, x^i) - \varepsilon_t$$

for all $x^i \in \mathbb{X}^i$, where

$$U^i(\mu^i, s^i, \eta^i, x^i) \equiv E_{\bar{Q}_{\mu^i}(\cdot|s^i,x^i)}\left[\pi^i(x^i, Y^i) + \eta^i(x^i) + \delta E_{p_{\mathbb{S}^i}}\left[V^i(B^i(\mu^i, s^i, x^i, Y^i), S^i)\right]\right].$$

Theorem 4. *Suppose that σ is an equilibrium of the perturbed game such that $\Theta^i(\sigma) \neq \Theta^i$ for all i. If $\delta^i = 0$ for all $i \in I$, then there exists a profile of priors with the property that $\mu^i_0(\Theta^i(\sigma)) < 1$ and an asymptotically optimal policy profile ϕ such that σ is strongly stable under ϕ.*

[33] See, for example, Nyarko (1994), for a proof of convexity of the value function.

[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).

Proof. See the Appendix. □

Theorem 4 says that, for any equilibrium, we can always find policy profiles that are asymptotically optimal, in the sense that players make vanishing optimization mistakes, and myopic, in the sense that players maximize current payoffs, such that behavior converges to that equilibrium with probability one.[35]

The main idea of the proof is due to Fudenberg and Kreps (1993). We construct a strategy profile with the property that players play optimally except when their belief is in a neighborhood of the belief supporting the equilibrium. In such neighborhood, players play as if the belief were exactly the equilibrium belief. We choose $(\varepsilon_t)_t$ to make sure that beliefs always remain in the neighborhood; thus, players always play the equilibrium strategy. Lemma 4 then implies that, as time goes by, players become increasingly confident of the equilibrium belief. We can then decrease the size of the neighborhood and, therefore, we can take ε_t to zero and guarantee that the region in which players are not optimizing vanishes. Intuitively, players are somehow convinced early on about 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. Finally, the requirement of myopia is necessary to reach certain equilibria that rely on incorrect beliefs due to lack of experimentation.

Theorems 3 and 4 provide our justification for our definition of equilibrium.[36] Of course, our results leave open the possibility of refinements of equilibria. One natural refinement is to require exact, not asymptotic, optimality, and to ask whether certain equilibria can be reached with positive probability (e.g., Benaim and Hirsch, 1999). Another possible refinement is to make assumptions on the discount factor: the higher the discount factor, the higher the incentives to experiment. Thus, certain outcomes can be ruled out by considering patient agents (e.g., Easley and Kiefer (1988) in single-agent settings). A final possible refinement is to follow Harsanyi (1973) and rule out those equilibria that are not regular in the sense that they might not be

[35]The assumption that $\mu_0^i(\Theta^i(\sigma)) < 1$ highlights that we are not picking the prior simply to match the equilibrium belief, in which case the statement would hold trivially. Similarly, we consider the arguably interesting case where $\Theta^i(\sigma) \neq \Theta^i$ because, otherwise, there might be no learning and the prior might never get updated.

[36]It is straightforward to check that Theorem 3 also holds if we replace the assumption that profiles are optimal with the assumption that they are asymptotically optimal.

approachable by some sequence of perturbed games (e.g., Doraszelski and Escobar (2010) in dynamic stochastic games). All of these ideas have been extensively studied in the literature. Moreover, we believe that the issue of refinements or equilibrium selection is probably best left for specific applications.

5 Additional examples

We illustrate the applicability of the framework by discussing several additional examples taken from different fields. For all of our examples, we describe the game and characterize equilibrium, but we omit the description of the dynamic learning environment for which, according to the results in Section 4, equilibrium represents steady-state behavior. In some cases, we restrict attention to pure strategies but allow the players to choose from a continuum of actions because this is the standard way in which these examples are often described.[37]

5.1 Non-linear pricing

Sobel (1984) considered the problem of a consumer who faces a possibly non-linear pricing menu from a monopolist but acts as if she faces a linear price.[38] A consumer decides to purchase quantity $x \in \mathbb{X} = \{x_1, ..., x_k\} \subset \mathbb{R}_+$ from a monopolist at a *unit* cost of $y \in \mathbb{Y} \subset \mathbb{R}$ and obtains a payoff of $\pi(x, y) = u(x) - yx$, where yx is the total cost from purchasing quantity x. The total cost is determined by a (possibly nonlinear) pricing menu $(x, r(x))_{x \in \mathbb{X}}$, where $r(x)$ is the amount charged by the monopolist when the consumer purchases quantity x. Thus, the set of all possible *unit* costs is $\mathbb{Y} = \{y \in \mathbb{R} : y = r(x)/x, \, x \in \mathbb{X}\}$, where $r(x)/x$ is the true unit cost of purchasing quantity x. The consumer incorrectly believes that she faces a (possibly random) linear price $y \in \mathbb{Y}$ that does not depend on her choice. Her misspecified subjective model is the set of all probability distributions over $\mathbb{Y}$, i.e., $\Theta = \Delta(\mathbb{Y})$. For

[37] For convenience, in some examples we also omit to restrict Θ to be a compact space or the space of signals or consequences to be finite when it is otherwise clear that existence of equilibrium will not be jeopardized. Also, in some cases, we assume that the subjective model is linear with error terms that are normally distributed. These assumptions are convenient because, as we show in the Online Appendix B, the solution of the minimization of the wKLD function is akin to the estimand of a linear regression model.

[38] More recently, Ito (2012) provides evidence that households facing nonlinear electricity price schedules respond to average price rather than marginal or expected marginal price.

a parameter vector $\theta = (\theta_1, ..., \theta_k)$, we let θ_j denote the probability that the linear price is $r(x_j)/x_j$. A strategy is denoted by $\sigma = (\sigma_1, ..., \sigma_k) \in \Delta(\mathbb{X})$, where σ_j is the probability that the consumer chooses quantity x_j.

We now characterize the set of equilibria in this environment. For a strategy σ, the wKLD function is

$$K(\sigma, \theta) = \sum_{j=1}^{k} \sigma_j \ln \frac{1}{\theta_j}$$

and the unique minimizer is $\theta(\sigma) = (\sigma_1, ..., \sigma_k)$. In other words, if behavior stabilizes to σ, then the consumer eventually believes that the linear price is $r(x_j)/x_j$ with probability σ_j and, therefore, that the expected cost of purchasing any quantity x is $\left(\sum_{j=1}^{k} (r(x_j)/x_j)\, \sigma_j\right) x$. Thus, σ is an equilibrium if and only if every x in the support of σ maximizes

$$u(x) - \left(\sum_{j=1}^{k} (r(x_j)/x_j)\, \sigma_j\right) x. \tag{17}$$

Sobel (1984) shows that, for a broad class of economies, the monopolist is worse off and some consumers are better off when consumers follow the boundedly rational equilibrium strategy.

5.2 Misspecified market structure

In complex environments, it is sometimes unrealistic or too hard for firms to take into account the actions of every other firm.[39] Alternatively, some firms might make the opposite simplification and act as if they are in a perfectly competitive market, thus failing to take into account their effect on the market price. Arrow and Green (1973) and Kirman (1975) were among the first to formally study such settings.

Example 1. The following example is studied by Arrow and Green (1973). The inverse demand function is

$$y = \alpha - \omega \sum_{i=1}^{I} x^i \in \mathbb{Y} \subset \mathbb{R}, \tag{18}$$

[39] According to Phillips (2005), firms usually make pricing decisions by applying revenue management models that often do not forecast the response of competitors.

where y is the market price, α is the demand intercept, ω is the realization of a random variable with exponential distribution with parameter $1/\bar{\theta}$, and $x^i \in \mathbb{X}^i \subset \mathbb{R}$ is the quantity chosen by firm i. To simplify notation, let $X^S \equiv \sum_{i=1}^{I} x^i$.

Each of I firms competes by simultaneously choosing quantity. The only feedback firms observe is the market price. The profit of firm i given quantity x^i and market price y is $\pi^i(x^i, y) = yx^i - c(x^i)$, where $c(x^i) = .5\,(x^i)^2$ is the cost of producing quantity x^i. Moreover, each firm i believes that the price is not given by (18) but rather by

$$y = \alpha - \omega,$$

where ω is the realization of a random variable with exponential distribution with parameter $1/\theta^i$, where $\theta^i \in \Theta$. Thus, firms believe that the market price is a random variable unaffected by their actions. As usual, we let $Q^i_{\theta^i}$ denote the subjective distribution of the consequence (i.e., the market price) given $\theta^i \in \Theta$.

If firm i believes that the parameter is θ^i, then its optimal quantity equates marginal cost to the expected price, i.e.,

$$x^i = \alpha - \theta^i, \tag{19}$$

assuming that $\alpha \geq \theta^i$.

A noteworthy aspect of this setting is that, for every $x = (x^i)_{i \in I}$ and every firm i, there exists $\hat{\theta}^i \in \Theta$ such that the subjective distribution coincides with the objective distribution, conditional on x^i, i.e., $Q^i_{\hat{\theta}^i}(\cdot \mid x^i) = Q^i_x(\cdot \mid x^i)$.[40] To see this claim, note that, because the firms know α, it is sufficient to compare the distributions over ω and $\omega \sum_{i=1}^{I} x^i$. The former is an exponential distribution with parameter $1/\theta^i$ and the latter is exponential with parameter $1/\bar{\theta}(\sum_{i=1}^{I} x^i)$. Thus, the claim follows by setting

$$\hat{\theta}^i(x) = \bar{\theta} X^S. \tag{20}$$

One implication of this result is that, if firms play x, then firms will not only believe that the true parameter is $\hat{\theta}^i$, but their misspecified model will also provide a perfect fit (i.e., the wKLD is zero at $\hat{\theta}^i$). Despite this perfect fit in equilibrium, notice that firms still have incorrect beliefs about off-equilibrium actions, which explains why an

[40] Arrow and Green (1973) impose this restriction on the subjective game. As we show in this paper, this restriction is not required to carry out the analysis of misspecified learning.

equilibrium might not constitute a Nash equilibrium in this case.[41]

From (19) and (20), it follows that x is an equilibrium if and only if

$$x^i = \alpha - \bar{\theta}(\sum_{i=1}^{I} x^i)$$

for all $i = 1, ..., I$. It follows from some algebra that there is a unique equilibrium and it is given by the symmetric strategy $x^i = \alpha/(1 + \bar{\theta}I)$. This equilibrium is different from the Nash equilibrium, which is easily verified to be unique and given by $x^i_{NE} = \alpha/(1 + \bar{\theta}(I + 1))$. Under both equilibrium concepts, profits go to zero as the number of firms go to infinity. But profits per firm are always higher in a Nash equilibrium compared to the equilibrium of the misspecified model. Thus, every firm ends up worse off when all firms ignore their market power.

Example 2. The following example is from Kirman (1975). Each of I firms now compete by simultaneously choosing *price*. Demand is given by the differentiated demand system

$$y^i = \alpha - \beta x^i + \gamma \sum_{k \neq i} x^k + \varepsilon^i, \tag{21}$$

where y^i the firm i's quantity, $x^i \in \mathbb{X} \subset \mathbb{R}$ is firm i's price, $x^j \in \mathbb{X}$ is firm j's price, ε^i is independently drawn from a standard normal distribution, and (α, β, γ) are the true parameters. Firm i believes, in contrast, that quantity demanded is

$$y_i = \theta^i - \beta x^i + \varepsilon^i, \tag{22}$$

where ε^i is drawn from a standard normal distribution and $\theta^i \in \Theta$ is the unknown demand intercept. In particular, each firm fails to take into account the presence of the other firms. Let $Q^i_{\theta^i}$ denote the subjective distribution over the quantity y^i of firm i given parameter θ^i.

We now verify that the subjective model is correctly specified in steady state. Fix any strategy profile x^* and firm i. From (21), for all x^i, the objective distribution $Q^i_{x^*}(\cdot \mid x^i)$ is normal with mean $\alpha - \beta x^i + \gamma \sum_{k \neq i} x^k$. From (22), for all x^i, the subjective distribution $Q^i_{\theta^i}(\cdot \mid x^i)$ parameterized by θ^i is normal with mean $\theta^i - \beta x^i$

[41] This situation is different than the situation in a self-confirming equilibrium, where players *might* have incorrect beliefs about off-equilibrium actions, but where Nash equilibrium is a special case because players can always hold the correct counterfactual beliefs.

and unit variance. Thus, there exists $\theta^i(x^*) = \alpha + \gamma \sum_{k \neq i} x^k$ such that $Q^i_{\theta^i(x^*)}(\cdot \mid x^i) = Q^i_{x^*}(\cdot \mid x^i)$ for all x^i.

It is trivial to check that the subjective model also has full feedback.[42] Thus, by Proposition 1, any equilibrium must also be a Nash equilibrium. Consequently, we obtain the same equilibrium outcome whether or not firms fail to account for each others' presence.

Example 3. Arrow and Green (1973) and Kirman (1975) also studied examples where firms do not know the slope of their demand functions and obtained multiple equilibria. This multiplicity, however, arises because it is not possible to identify the slope without variation in actions. We now consider a richer example where variation in costs leads naturally to variation in actions. The example also illustrates that biases can result from ignoring competition even if all actions are strategically independent.

A vector of costs $s = (s^1, s^2) \in \mathbb{S}^1 \times \mathbb{S}^2 \subset \mathbb{R}^2_+$ is drawn according to the probability distribution $p_S \in \Delta(\mathbb{S}^1 \times \mathbb{S}^2)$. Each firm $i = 1, 2$ privately observes its marginal cost s^i and then simultaneously chooses price $x^i \in \mathbb{X} \subset \mathbb{R}_+$. The quantity sold by firm i is given by the demand system

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

where $\gamma^* > 1$ is the demand elasticity (in absolute value) and the error terms are independent (of each other and also of costs) and standard normal, $\varepsilon^i \sim N(0, 1)$ for $i = 1, 2$. As a benchmark, it is straightforward to check that the best response of a firm does not depend on the choice of the other firm. Thus, there is a unique Nash equilibrium and it is in dominant strategies, $\sigma^{NE}(s_i) = [\gamma^*/(\gamma^* - 1)]\, s_i$.

Suppose that each firm $i = 1, 2$ (incorrectly) believes that they are a monopolist in this market and that the demand function they face is

$$\ln y^i = \alpha - \gamma \ln x^i + \varepsilon^i, \tag{23}$$

where $\varepsilon^i \sim N(0, 1)$. Formally, the subjective model is Θ, where $\theta = (\alpha, \gamma)$, but the only parameter of interest is the demand elasticity γ, since, once elasticity is known, it is optimal to set price $x^i = \sigma(s^i) = [\gamma/(\gamma - 1)]\, s^i$ when cost is s^i. Since the error term

[42] From equation (22), we can write $y^i = g(x^i, z^i)$, where player i believes that $z^i = \theta^i + \varepsilon^i$ has a distribution that does not depend on x^i.

is normally distributed, the minimizer of the wKLD function is given by estimand of equation (23). Thus, for all $\sigma = (\sigma^1, \sigma^2)$,

$$\begin{aligned}\gamma^{OLS}(\sigma) &= -\frac{Cov(\ln \sigma^i(S^i), \ln y^i)}{Var(\ln \sigma^i(S^i))} \\ &= \gamma^* - \beta^* \frac{Cov(\ln \sigma^i(S^i), \ln \sigma^j(S^j))}{Var(\ln \sigma^i(S^i))}. \end{aligned} \tag{24}$$

It follows that a strategy profile σ is an equilibrium if and only if

$$\sigma_1(s) = \sigma_2(s) = \left[\gamma^{OLS}(\sigma)/(\gamma^{OLS}(\sigma) - 1)\right] s. \tag{25}$$

By replacing (25) into (24), we obtain that, in equilibrium, $\gamma^{OLS}(\sigma)$ is independent of σ and given by

$$\gamma^{OLS} = \gamma^* - \beta^* \frac{Cov(\ln S^i, \ln S^j)}{Var(\ln S^i)}. \tag{26}$$

Thus, there is a unique equilibrium and it is supported by belief γ^{OLS}.[43] Moreover, (26) 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 a Nash equilibrium.

5.3 Regression towards the mean

Tversky and Kahneman (1973) argue that people often fail to understand the notion of regression towards the mean. One of their examples is about experienced flight instructors who note that praise for a good landing is usually followed by a poorer landing on the next try, while criticism for a bad landing is usually followed by a better landing. The instructors conclude that praise hurts performance while criticism

[43] This is true as long as we make an assumption on the primitives that makes $\gamma^{OLS} > 1$; otherwise firms would like to price infinity because of the assumption of constant elasticity of demand.

improves performance.

We provide a simple model that formalizes the potential misspecification underlying the instructor's reasoning. The instructor observes the initial performance s_1 of a student and decides whether to praise or criticize, $x \in \{C, P\}$. The student then performs again and the instructor observes this final performance, s_2. The truth is that performances $Y = (S_1, S_2)$ are independent, standard normal random variables. The instructor believes, however, that

$$s_2 = s_1 + \varepsilon_x,$$

where $\varepsilon_x \sim N(\theta_x, 1)$. Thus, the instructor believes that the final performance depends on the initial performance and an error term with a mean that potentially depends on the decision to praise or criticize the initial performance. The instructor's parameter space is $\Theta = \mathbb{R}^2$, where a parameter vector $\theta = (\theta_C, \theta_P)$ represents the mean of the error term after criticism and praise, respectively.

The instructor's payoff is $\pi(x, (s_1, s_2)) = s_2 - c(x, s_1)$, where $c(x, s_1) = \kappa |s_1| > 0$ if either $s_1 > 0, x = C$ or $s_1 < 0, x = P$, and, in all other cases, $c(x, s_1) = 0$. 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.

We now characterize the equilibrium of this decision problem. 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 performances that are above σ and criticizes the remaining performances.

The wKLD function is

$$K(\sigma, \theta) = \int_{\mathbb{R}} \left(1_{s_1 < \sigma}(s_1) E \ln \frac{f(S_2)}{f(S_2 - (\theta_C + s_1))} + 1_{s_1 > \sigma}(s_1) E \ln \frac{f(S_2)}{f(S_2 - (\theta_P + s_1))} \right) f(s_1) ds_1,$$

where f is the density of the standard normal distribution and expectations are with respect to the true distributions. It is straightforward to show that, for each σ, the

unique parameter vector that minimizes $K(\sigma, \cdot)$ is

$$\begin{aligned}\theta_C(\sigma) &= E(S_2 - S_1 \mid S_1 < \sigma) \\ &= 0 - E(S_1 \mid S_1 < \sigma) > 0\end{aligned}$$

and, similarly, $\theta_P(\sigma) = 0 - E(S_1 \mid S_1 > \sigma) < 0$. The intuition is that instructors are 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 σ, 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. Also, by optimality, she chooses a cutoff that makes her indifferent between praising and criticizing. Thus, $\sigma^* > 0$ is an equilibrium cutoff if and only if

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

Similar steps establish that there is no equilibrium with $\sigma^* \leq 0$. Thus, instructors are excessively critical in equilibrium because they incorrectly believe that criticizing a student improves her performance and that praising a student worsens it.

5.4 Classical and Keynesian monetary policy

This example is based on Sargent (1999, Chapter 7). There are two players, the government (G) and the public (P). The government chooses monetary policy $x^G \in \mathbb{R}$ and the public chooses inflation forecasts $x^P \in \mathbb{R}$. Inflation, e, and unemployment, U, are determined as follows:

$$\begin{aligned}e &= x^G + v\varepsilon_e \\ U &= u^* - \phi(e - x^P) + \varepsilon_U,\end{aligned}$$

where ε_e and ε_U are independent and have standard normal distributions, and $v^2 > 0$ is the variance of inflation. In other words, 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 $\phi > 0$ measures the effect of surprise inflation on unemployment and $u^* > 0$ is the natural rate of unemployment.

The government's payoff is $\pi(x^G, e, U) = -(U^2 + e^2)$ and the public's payoff is $\pi(x^P, e) = -(e - x^P)^2$. It is straightforward to check that there is a unique Nash equilibrium given by

$$x^G_{NE} = x^P_{NE} = \phi u^*. \tag{27}$$

In a Nash equilibrium, the government inflates the economy because it cannot commit to avoid inflation surprises. Of course, in equilibrium, there are no inflation surprises, and, therefore, everyone is worse off compared to the situation where the government could commit to $x^G = 0$ (Kydland and Prescott, 1977).

We consider two types of subjective models. In the "classical model",

$$\begin{aligned} e &= \theta^C_o + x^G + v^C_e \varepsilon_e \\ U &= \theta^C_1 - \theta^C_2 e + v^C_U \varepsilon_U. \end{aligned}$$

Thus, the classical 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. Formally, the government's subjective model is $\Theta^C \subseteq \mathbb{R}^5$, where $\theta^C = (\theta^C_0, \theta^C_1, \theta^C_2, v^C_e, v^C_U)$.

In the "Keynesian model",

$$\begin{aligned} U &= \theta^C_0 - x^G + v^K_U \varepsilon_U \\ e &= \theta^K_1 - \theta^K_2 U + v^K_e \varepsilon_e. \end{aligned}$$

Thus, the Keynesian government believes that its monetary policy affects unemployment, not inflation, and that unemployment in turn affects inflation. Formally, the government's subjective model is $\Theta^K \subseteq \mathbb{R}^5$, where $\theta^K = (\theta^K_1, \theta^K_2, \theta^K_2, v^K_e, v^K_U)$.

In both cases, we assume that the public believes that $e \sim N(\theta^P, v^2)$. Formally, the public's model is $\Theta^P \subseteq \mathbb{R}$.

First, note that the classical model is correctly specified in steady state. To see this claim, fix any (x^G_*, x^P_*). For a given, x^G, the objective distribution for inflation is $e \sim N\left(x^G, v^2\right)$, while the government believes it is $e \sim N\left(\theta^C_0 + x^G, \left(v^C_e\right)^2\right)$. Similarly, the objective unemployment distribution given x^G is $U \sim N(u^* - \phi(x^G - x^P_*), (v\phi)^2 + 1)$, while the government believes that a choice of x^G would lead to $U \sim N(\theta^C_1 -$

$\theta_2^C x^G, \left(v_e^C \theta_2^C\right)^2 + \left(v_U^C\right)^2$). It follows, by setting $\theta_0^C = 0$, $\theta_1^C = u^* + \phi x_*^P$, $\theta_2^C = \phi$, $v_e^C = v$, and $v_U^C = 1$, that these objective and subjective distributions coincide for all x^G. Similarly, the public's belief about e coincides with the true distribution by setting $\theta^P = x_*^G$.

It is trivial to check that the classical model also has full feedback.[44] Thus, by Proposition 1, the unique equilibrium of this game is given by the Nash equilibrium in equation (27). In other words, it is irrelevant whether or not the government realizes that unemployment is driven by surprise, not actual, inflation.

The Keynesian model, in contrast, is not correctly specified in steady state. To find the equilibrium, we begin by finding the optimal strategies given fixed parameter values. For the government, the optimal strategy given θ^K can be shown to be

$$x^G(\theta^K) = -\frac{\theta_1^K \theta_2^K}{\left(\theta_2^K\right)^2 + 1} + \theta_0^K.$$

For the public, the optimal strategy given θ^P is $x^P(\theta^P) = \theta^P$.

Next, we minimize the wKLD function for every strategy profile. Because of the normality assumptions, this is equivalent to computing the estimands of a linear regression model. Fix any strategy profile $x = (x^G, x^P)$. Then, the unique minimizers of the wKLD are given by the estimands

$$\begin{aligned}
\theta_0^K(x) &= EU + x^G = \left(u^* - \phi\left(x^G - x^P\right)\right) + x^G, \\
\theta_1^K(x) &= Ee + \theta_2^K(x) EU = x^G + \theta_2^K(x)\left(u^* - \phi\left(x^G - x^P\right)\right), \\
\theta_2^K(x) &= -Cov(e, U)/Var(U) = \phi v^2/(\phi^2 v^2 + 1),
\end{aligned}$$

for the government, and $\theta^P(x) = x^G$ for the public.[45] From the previous characterization of beliefs and optimal strategies, it follows that $x = (x^G, x^P)$ is an equilibrium if and only if

$$x^P = x^G = -\frac{\theta_1^K(x)\theta_2^K(x)}{\left(\theta_2^K(x)\right)^2 + 1} + \theta_0^K(x),$$

or, equivalently,

$$x^P = x^G = \frac{(\phi v)^2 + 1}{\phi v^2} u^*. \tag{28}$$

[44] Let $z = (z_U, z_e)$, where $z_U = \theta_0^C + v_U^K \varepsilon_U$ and $z_e = \theta_1^K - \theta_2^K U + v_e^K \varepsilon_e$, and note that the distribution of z is believed not to depend on x^i.

[45] The estimates for the variances are irrelevant for computing the equilibrium.

A comparison of (27) and (28) reveals that the equilibrium policy—hence, expected inflation—is always higher for a Keynesian government compared to the Nash or classical equilibrium policy.

5.5 Trading under adverse selection

Several equilibrium concepts have been proposed to model people who fail to account for the information content of other people's actions: cursed equilibrium (Eyster and Rabin (2005)), analogy-based expectation equilibrium (Jehiel (2005); Jehiel and Koessler (2008)), and behavioral equilibrium (Esponda (2008)).[46] Applications include auctions, elections, and games of strategic information transmission. We illustrate, using a simple lemons problem, how these three equilibrium concepts fit into our framework.[47]

A (risk-neutral) buyer and a seller simultaneously submit a (bid) price $x \in \mathbb{X} \subset \mathbb{R}$ and an ask price $a \in \mathbb{A} \subset \mathbb{R}$, respectively. If $a \leq x$, then the buyer pays x to the seller and receives the seller's object, which the buyer values at $v \in \mathbb{V} \subset \mathbb{R}$. 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. Suppose that the seller's ask price and the buyer's value are drawn from the same probability distribution $p \in \Delta(\mathbb{A} \times \mathbb{V})$.[48]

We consider two different feedback functions. Under *full feedback*, the buyer observes the ask price and her own value at the end of each period. Under *partial feedback*, the buyer continues to observe the ask price, but she only observes her own value if she trades in that period.

We also consider two types of misspecified models for the buyer. In the first case, the buyer believes that her valuation V is independent of the seller's ask price: $\Theta_I = \Delta(\mathbb{A}) \times \Delta(\mathbb{V})$. The second type generalizes the first type. Consider a partition of the set $\mathbb{V}$ into k "analogy classes" $(\mathbb{V}_j)_{j=1,...,k}$, where $\cup_j \mathbb{V}_j = V$ and $\mathbb{V}_i \cap \mathbb{V}_j = 0$ for all $i \neq j$. The buyer believes that (A, V) are independent conditional on $V \in \mathbb{V}_i$, for each $i = 1, ..., k$. The parameter space is $\Theta_A = \times_j \Delta(\mathbb{A}) \times \Delta(\mathbb{V})$, where, for

[46] For experimental evidence, 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).

[47] See Esponda (2008) and Spiegler (2011) for additional results and discussion.

[48] 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.

a parameter $\theta = (\theta_1, ..., \theta_k, \theta_{\mathbb{V}}) \in \Theta_A$, $\theta_{\mathbb{V}}$ parameterizes the marginal distribution over $\mathbb{V}$ and, for each $j = 1, ..., k$, $\theta_j \in \Delta(\mathbb{A})$ parameterizes the distribution over $\mathbb{A}$ conditional on $V \in \mathbb{V}_j$. Detailed computations for these examples are provided in Online Appendix C.

As a benchmark, the Nash equilibrium (NE) price maximizes

$$\Pi^{NE}(x) = \Pr(A \leq x) \left(E(V \mid A \leq x) - x\right),$$

where Pr and E denote probability and expectation with respect to the true distribution.

Cursed equilibrium. With full feedback and model Θ_I, 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 steady-state belief about her expected profit from choosing any price x is

$$\Pi^{CE}(x) = \Pr(A \leq x) \left(E(V) - x\right). \tag{29}$$

A price is an equilibrium if and only if it maximizes (29) over $x \in \mathbb{X}$, i.e., a (fully) cursed equilibrium price in the terminology of Eyster and Rabin (2005).[49]

Behavioral equilibrium. With partial feedback and model Θ_I, 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 steady-state belief about her expected profit from choosing any other price x is

$$\Pi^{BE}(x, x^*) = \Pr(A \leq x) \left(E(V \mid A \leq x^*) - x\right). \tag{30}$$

Equilibrium is a fixed point: x^* is an equilibrium price if and only if $x = x^*$ maximizes (30), i.e., a *naive* behavioral equilibrium in the terminology of Esponda (2008).[50]

Analogy-based expectation equilibrium. With full feedback and model Θ_A, beliefs are as in a cursed equilibrium conditional on each analogy class, which implies that the buyer's steady-state belief about her expected profit from choosing any price x

[49] This misspecified model was first discussed in the lemons context by Kagel and Levin (1986).

[50] Behavioral equilibrium is more generally defined for any possible feedback and also allows for players who are sophisticated in the sense of having a correctly-specified model.

is[51]

$$\Pi^{ABEE}(x) = \sum_{j=1}^{k} \Pr(V \in \mathbb{V}_j) \left\{ \Pr(A \leq x \mid V \in \mathbb{V}_j) \left(E(V \mid V \in \mathbb{V}_j) - x \right) \right\}. \tag{31}$$

A price is an equilibrium if and only if it maximizes (31) over x, i.e., an analogy-based expectation equilibrium in the terminology of Jehiel and Koessler (2008).

Behavioral equilibrium with analogy classes. We obtain a new concept by combining partial feedback with model Θ_A. Suppose that the buyer's behavior has stabilized to some price x^*. Due to the possible correlation across analogy classes, the buyer might now believe that deviating to a different price $x \neq x^*$ affects her valuation. In particular, the buyer might have multiple beliefs at x^*. To obtain a natural equilibrium refinement, we assume that the buyer also observes the analogy class that contains her realized valuation, whether she trades or not, and that $\Pr(V \in \mathbb{V}_j,\ A \leq x) > 0$ for all $j = 1, ..., k$ and $x \in \mathbb{X}$.[52] Then, the buyer's steady-state belief about her expected profit from choosing any price x is

$$\Pi^{BEA}(x, x^*) = \sum_{i=j}^{k} \Pr(V \in \mathbb{V}_j) \left\{ \Pr(A \leq x \mid V \in \mathbb{V}_j) \left(E(V \mid V \in \mathbb{V}_j, A \leq x^*) - x \right) \right\}. \tag{32}$$

A price x^* is an equilibrium if and only if $x = x^*$ maximizes (32).

6 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

[51] 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.

[52] Alternatively, and more naturally, we could require the equilibrium to be the limit of a sequence of mixed strategy equilibria with the property that all prices are chosen with positive probability.

further developments in this area.

There are several natural directions for further research. One direction is to extend the tools developed in this paper to other environments that are of interest to economists, such as extensive-form games or dynamic environments and also market settings with price-taking agents. A second direction is to explore non-Bayesian models of belief updating. In our paper, 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.[53]

References

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

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

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

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

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

Barberis, N., A. Shleifer, and R. Vishny, "A model of investor sentiment," *Journal of financial economics*, 1998, *49* (3), 307–343.

Battigalli, P., *Comportamento razionale ed equilibrio nei giochi e nelle situazioni sociali*, Universita Bocconi, Milano, 1987.

[53] An example of one such type of restriction is provided by the concept of behavioral equilibrium (Esponda, 2008).

_ , **S. Cerreia-Vioglio, F. Maccheroni, and M. Marinacci**, "Selfconfirming equilibrium and model uncertainty," Technical Report 2012.

Bénabou, Roland and Jean Tirole, "Self-confidence and personal motivation," *The Quarterly Journal of Economics*, 2002, *117* (3), 871–915.

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

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

Billingsley, P., *Probability and Measure*, Wiley, 1995.

Blume, L.E. and D. Easley, "Learning to be Rational," *Journal of Economic Theory*, 1982, *26* (2), 340–351.

Bray, M., "Learning, estimation, and the stability of rational expectations," *Journal of economic theory*, 1982, *26* (2), 318–339.

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

Charness, G. and D. Levin, "The origin of the winner's curse: a laboratory study," *American Economic Journal: Microeconomics*, 2009, *1* (1), 207–236.

Compte, Olivier and Andrew Postlewaite, "Confidence-enhanced performance," *American Economic Review*, 2004, pp. 1536–1557.

Dekel, E., D. Fudenberg, and D.K. Levine, "Learning to play Bayesian games," *Games and Economic Behavior*, 2004, *46* (2), 282–303.

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

Doraszelski, Ulrich and Juan F Escobar, "A theory of regular Markov perfect equilibria in dynamic stochastic games: Genericity, stability, and purification," *Theoretical Economics*, 2010, *5* (3), 369–402.

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

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

_ , "Rationalizable conjectural equilibrium: A framework for robust predictions," *Theoretical Economics*, 2013, *8* (2), 467–501.

_ **and E.I. Vespa**, "Hypothetical Thinking and Information Extraction in the Laboratory," *American Economic Journal: Microeconomics, forthcoming*, 2013.

Evans, G. W. and S. Honkapohja, *Learning and Expectations in Macroeconomics*, Princeton University Press, 2001.

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

Eyster, Erik and Michele Piccione, "An approach to asset-pricing under incomplete and diverse perceptions," *Econometrica*, 2013, *81* (4), 1483–1506.

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

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

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

_ **and** _ , "Steady state learning and Nash equilibrium," *Econometrica*, 1993, pp. 547–573.

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

_ **and** _ , "Learning and Equilibrium," *Annual Review of Economics*, 2009, *1*, 385–420.

_ **and D.M. Kreps**, "A Theory of Learning, Experimentation, and Equilibrium in Games," Technical Report, mimeo 1988.

_ **and** _ , "Learning in extensive-form games I. Self-confirming equilibria," *Games and Economic Behavior*, 1995, *8* (1), 20–55.

_ **and S. Takahashi**, "Heterogeneous beliefs and local information in stochastic fictitious play," *Games and Economic Behavior*, 2011, *71* (1), 100–120.

Gabaix, X., "Game Theory with Sparsity-Based Bounded Rationality," *Working Paper*, 2012.

Harsanyi, J.C., "Games with incomplete information played by 'Bayesian' players, parts i-iii," *Management science*, 1967-8, *14*, 159–182, 320–334, and 486–502.

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

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

Ito, Koichiro, "Do Consumers respond to marginal or average price," *American Economic Review, forthcoming*, 2012.

Ivanov, A., D. Levin, and M. Niederle, "Can relaxation of beliefs rationalize the winner's curse?: an experimental study," *Econometrica*, 2010, *78* (4), 1435–1452.

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

_ **and D. Samet**, "Valuation equilibrium," *Theoretical Economics*, 2007, *2* (2), 163–185.

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

Jordan, J. S., "Three problems in learning mixed-strategy Nash equilibria," *Games and Economic Behavior*, 1993, *5* (3), 368–386.

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

_ **and** _ , *Common value auctions and the winner's curse*, Princeton University Press, 2002.

Kalai, E. and E. Lehrer, "Rational learning leads to Nash equilibrium," *Econometrica*, 1993, pp. 1019–1045.

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

Kreps, D. M. and R. Wilson, "Sequential equilibria," *Econometrica*, 1982, pp. 863–894.

Kullback, S. and R. A. Leibler, "On Information and Sufficiency," *Annals of Mathematical Statistics*, 1951, *22* (1), 79–86.

Kydland, Finn E and Edward C Prescott, "Rules rather than discretion: The inconsistency of optimal plans," *The Journal of Political Economy*, 1977, pp. 473–491.

McLennan, A., "Price dispersion and incomplete learning in the long run," *Journal of Economic Dynamics and Control*, 1984, *7* (3), 331–347.

Nash, J., "Non-cooperative games," *The Annals of Mathematics*, 1951, *54* (2), 286–295.

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

_ , "On the convexity of the value function in Bayesian optimal control problems," *Economic Theory*, 1994, *4* (2), 303–309.

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

Phillips, R., *Pricing and revenue optimization*, Stanford University Press, 2005.

Piccione, M. and A. Rubinstein, "Modeling the economic interaction of agents with diverse abilities to recognize equilibrium patterns," *Journal of the European economic association*, 2003, *1* (1), 212–223.

Pollard, D., *A User's Guide to Measure Theoretic Probability*, Cambridge University Press, 2001.

Rabin, M., "Inference by Believers in the Law of Small Numbers," *Quarterly Journal of Economics*, 2002, *117* (3), 775–816.

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

Radner, R., *Equilibrium Under Uncertainty*, Vol. II of *Handbook of Mathematical Economies*, North-Holland Publishing Company, 1982.

Rothschild, M., "A two-armed bandit theory of market pricing," *Journal of Economic Theory*, 1974, *9* (2), 185–202.

Rubinstein, A. and A. Wolinsky, "Rationalizable conjectural equilibrium: between Nash and rationalizability," *Games and Economic Behavior*, 1994, *6* (2), 299–311.

Rubinstein, Ariel, "Finite automata play the repeated prisoner's dilemma," *Journal of economic theory*, 1986, *39* (1), 83–96.

Salant, Y., "Procedural analysis of choice rules with applications to bounded rationality," *The American Economic Review*, 2011, *101* (2), 724–748.

Sargent, T. J., "Bounded rationality in macroeconomics," 1993.

_ , *The Conquest of American Inflation*, Princeton University Press, 1999.

Schwartzstein, J., "Selective Attention and Learning," *working paper*, 2009.

Selten, R., "Reexamination of the perfectness concept for equilibrium points in extensive games," *International journal of game theory*, 1975, *4* (1), 25–55.

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

Spiegler, R., "The Market for Quacks," *Review of Economic Studies*, 2006, *73*, 1113–1131.

_ , *Bounded Rationality and Industrial Organization*, Oxford University Press, 2011.

_ , "Placebo reforms," *The American Economic Review*, 2013, *103* (4), 1490–1506.

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

Vickrey, W., "Counterspeculation, auctions, and competitive sealed tenders," *The Journal of finance*, 1961, *16* (1), 8–37.

Wilson, A., "Bounded Memory and Biases in Information Processing Job Market Paper," *Princeton University*, 2003.

Appendix

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

$$\begin{aligned} K^i(\sigma, \theta^i) &= -\sum_{(s^i,x^i)\in\mathbb{S}^i\times\mathbb{X}^i} E_{Q^i_\sigma(\cdot|s^i,x^i)}\left[\ln\left(\frac{Q^i_{\theta^i}(Y^i \mid s^i,x^i)}{Q^i_\sigma(Y^i \mid s^i,x^i)}\right)\right]\sigma_i(x^i \mid s^i)p_{\mathbb{S}^i}(s^i) \\ &\geq -\sum_{(s^i,x^i)\in\mathbb{S}^i\times\mathbb{X}^i} \ln\left(E_{Q^i_\sigma(\cdot|s^i,x^i)}\left[\frac{Q^i_{\theta^i}(Y^i \mid s^i,x^i)}{Q^i_\sigma(Y^i \mid s^i,x^i)}\right]\right)\sigma_i(x^i \mid s^i)p_{\mathbb{S}^i}(s^i) \qquad (33) \\ &= 0, \end{aligned}$$

where Jensen's inequality and the strict concavity of $\ln(\cdot)$ imply the inequality in (33) as well as the fact that (33) holds with equality if and only if $Q^i_{\theta^i}(\cdot \mid s^i, x^i) = Q^i_{\theta^i}(\cdot \mid s^i, x^i)$ for all (s^i, x^i) such that $\sigma_i(x^i \mid s^i)p_{\mathbb{S}^i}(s^i)$, or, equivalently, by the assumption that $p_{\mathbb{S}^i}(s^i) > 0$, $\sigma_i(x^i \mid s^i) > 0$.

Part (ii). Fix $i \in I$. By assumption, there exists θ^i_* such that $M_\sigma \equiv K^i(\sigma, \theta^i_*) < \infty$. This implies that the set $\{\theta \in \Theta^i : K^i(\sigma, \theta^i) \leq M_\sigma\}$ is nonempty for all $\sigma \in \Sigma$. Thus, $\Theta^i(\sigma)$ can be equivalently defined as the set of minimizers over this new constraint set. Moreover, continuity of $K^i(\sigma, \cdot)$ implies that this constraint set is closed, compactness of Θ^i further implies that it is compact, and continuity of $K^i(\cdot, \theta)$ implies that it is upper hemicontinuous in σ. The result then follows by the Theorem of the Maximum.□

Proof of Lemma 2. Part (i). By the assumption that payoffs are bounded, let $\bar{\pi}$ and $\underline{\pi}$ denote the upper and lower bounds, respectively. Then $\sigma^i(x_i \mid s_i) \geq P_{\mathbb{V}^i}\left(\eta^i : \eta^i(x^i) - \eta^i(\hat{x}^i) \geq \underline{\pi} - \bar{\pi}\ \forall \hat{x}^i\right) > 0$, where the strict inequality follows by the

assumption that the support of η^i is unbounded. Part (ii) follows trivially from the assumption that the model is identifiable. $\square$

Proof of Theorem 1. By Lemma 2(i) and finiteness of $\mathbb{X}$, there exists $c \in (0,1)$ such that the set $\Sigma^* = \{\sigma \in \Sigma : \sigma^i(x^i \mid s^i) \in [c, 1-c]\ \forall (s^i, x^i) \in \mathbb{S}^i \times \mathbb{X}^i\ , \forall i \in I\}$ contains all strategy profiles that can be optimal. For all $\sigma \in \Sigma^*$, Lemma 2(ii) implies that, for all $i \in I$, $\bar{Q}^i_{\mu^i_1} = \bar{Q}^i_{\mu^i_2}$ for all $\mu^i_1, \mu^i_2 \in \Delta(\Theta^i(\sigma))$. Thus, we can define $\tilde{Q}^i(\sigma) = \int_{\Theta^i} Q_{\theta^i} \mu^i(d\theta)$ (not to be confused with the objective distribution Q^i_σ) where μ^i is any belief that belongs to $\Delta(\Theta^i(\sigma))$ and $\sigma \in \Sigma^*$. Moreover, it is straightforward to see that σ is an equilibrium of a fully-perturbed game if and only if it is a fixed point of the function $g : \Sigma^* \to \Sigma^*$ defined by

$$g^i(x^{i} \mid s^i) = P_{\mathbb{V}^i}\left(\eta^i : x^i \in \arg\max_{\bar{x}^i \in \mathbb{X}^i} E_{\tilde{Q}^i(\sigma)(\cdot \mid s^i, \bar{x}^i)}\left[\pi^i(\bar{x}^i, Y^i)\right] + \eta^i(\bar{x}^i)\right).$$

The space Σ^* is a compact and convex subset of an Euclidean space. By Brouwer's fixed point theorem, a fixed point exists if g is continuous. To show that g is continuous, we first show that $\tilde{Q}^i$ is continuous for all $\sigma \in \Sigma^*$. Let $\sigma_* \in \Sigma^*$ and suppose that $(\sigma_n)_n$ is a sequence of strategies in Σ^* that converges to σ_*. For each element in the sequence, the fact that $\Theta^i(\sigma_n)$ is non-empty (by Lemma 1(ii)) and that Θ^i is compact implies that we can pick a subsequence $\theta^i_{n_k} \in \Theta^i(\sigma_{n_k})$ that converges to some θ^i_*. Then $\theta^i_* \in \Theta^i(\sigma_*)$ by the upper hemicontinuity of $\Theta^i(\cdot)$ established in Lemma 1(ii). Thus, the facts that $\theta^i_{n_k} \in \Theta^i(\sigma_{n_k})$ and $\theta^i_* \in \Theta^i(\sigma_*)$ imply that $\tilde{Q}^i(\sigma_{n_k}) = Q_{\theta^i_{n_k}}$ and $\tilde{Q}^i(\sigma_*) = Q_{\theta^i_*}$. Continuity of $\tilde{Q}^i$ then follows because, by assumption, $Q^i_{\theta^i}$ is continuous as a function of θ^i. Together with the absolute continuity of $P_{\mathbb{V}^i}$, a standard argument shows that g is continuous. $\square$

Proof of Theorem 2. By assumption, there is a sequence $(\sigma_\xi, \mu_\xi)_\xi$ such that, for all ξ and all $i \in I$, (i) σ^i_ξ is optimal for the perturbed game *given* μ^i_ξ, (ii) $\mu^i_\xi \in \Delta(\Theta^i(\sigma_\xi))$, and (iii) $\lim_{\xi \to \infty} \sigma_\xi = \sigma$. By compactness of $\Delta(\Theta)$, we can fix a subsequence $\mu_{\xi(j)}$ that converges to some μ. By Lemma 1(ii), $\Theta^i(\cdot)$ is upper hemicontinuous and compact valued; hence, by Theorem 17.13 of Aliprantis and Border (2006), the correspondence $\Delta(\Theta^i(\cdot))$ inherits the same properties. Therefore, $\mu^i \in \Delta(\Theta^i(\sigma))$ for all $i \in I$. Thus, to show that σ is an equilibrium of the (unperturbed) game, it remains to show that, for all i, σ^i is an optimal strategy for the (unperturbed) game given μ^i. We proceed by contradiction. Suppose not, so that there exists $i \in I$ and

$(s^i, x^i) \in \mathbb{S}^i \times \mathbb{X}^i$ such that $\sigma^i(x^i \mid s^i) = C > 0$, but there exists $\hat{x}^i$ such that

$$E_{\bar{Q}^i_{\mu^i}(\cdot|s^i,\hat{x}^i)}\left[\pi^i(\hat{x}^i, Y^i)\right] - E_{\bar{Q}^i_{\mu^i}(\cdot|s^i,x^i)}\left[\pi^i(x^i, Y^i)\right] = A > 0. \tag{34}$$

Let J be such that, for all $j \geq J$, the following two conditions are satisfied: (i) $\left|\sum_{y^i \in Y^i} \pi^i(\bar{x}^i, y^i)\left(\bar{Q}^i_{\mu^i_{\xi(j)}}(y^i \mid s^i, \bar{x}^i) - \bar{Q}^i_{\mu^i}(y^i \mid s^i, \bar{x}^i)\right)\right| \leq A/4$ for all $\bar{x}^i \in \mathbb{X}^i$ and (ii) $P_{\mathbb{V}^i_{\xi(j)}}\left(\eta^i : \max_{x^i \in \mathbb{X}^i} |\eta^i(x^i)| < A/4\right) \geq 1 - C/2$. Condition (i) is possible because payoffs are bounded and $Q^i_{\mu^i}$ is continuous as a function of μ^i and $\lim_{j\to\infty} \mu^i_{\xi(j)} = \mu^i$. Condition (ii) is possible by the assumption that perturbations vanish (equation 6). Let

$$N^i_j(x^i, \hat{x}^i) \equiv \left\{\eta^i : \eta^i(x^i) - \eta^i(\hat{x}^i) < E_{\bar{Q}^i_{\mu^i_{\xi(j)}}(\cdot|s^i,\hat{x}^i)}\left[\pi^i(\hat{x}^i, Y^i)\right] - E_{\bar{Q}^i_{\mu^i_{\xi(j)}}(\cdot|s^i,x^i)}\left[\pi^i(x^i, Y^i)\right]\right\}.$$

Then, conditions (i) and (ii) and equation (34) imply that, for all $j \geq J$,

$$\begin{aligned} P_{\mathbb{V}^i_{\xi(j)}}\left(N^i_j(x^i, \hat{x}^i)\right) &\geq P_{\mathbb{V}^i_{\xi(j)}}\left(\eta^i : \eta^i(x^i) - \eta^i(\hat{x}^i) < A/2\right) \\ &\geq P_{\mathbb{V}^i_{\xi(j)}}\left(\eta^i : \max_{x^i \in \mathbb{X}^i} \left|\eta^i(x^i)\right| < A/4\right) \\ &\geq 1 - C/2. \end{aligned}$$

Finally, inspection of (5) reveals that the event that defines the optimal strategy $\sigma^i_{\xi(j)}(x^i \mid s^i)$ is contained in the complement of the event $N^i_j(x^i, \hat{x}^i)$. Thus, for all $j \geq J$, $\sigma^i_{\xi(j)}(x^i \mid s^i) \leq C/2$, which contradicts the facts that $\sigma^i(x^i \mid s^i) = C$ and $\lim_{j\to\infty} \sigma^i_{\xi(j)} = \sigma^i$. $\square$

Proof of Lemma 3. We first show that

$$\eta^i \mapsto \max_{x^i \in \mathbb{X}^i} E_{\bar{Q}_{\mu^i}(\cdot|s^i,x^i)}\left[\pi^i(x^i, Y^i) + \eta^i(x^i) + \delta E_{p_{\mathbb{S}^i}}\left[H^i(B^i(\mu^i, s^i, x^i, Y^i), S^i)\right]\right]$$

is measurable for any $H^i \in L^\infty(\Delta(\Theta^i) \times \mathbb{S}^i)$ and any (μ^i, s^i). It suffices to check that set of the form $\left\{\eta^i : \max_{x^i \in \mathbb{X}^i} E_{\bar{Q}_{\mu^i}(\cdot|s^i,x^i)}\left[\pi^i(x^i, Y^i) + \eta^i(x^i) + \delta E_{p_{\mathbb{S}^i}}\left[H^i(B^i(\mu^i, s^i, x^i, Y^i), S^i)\right]\right] < a\right\}$ is measurable for any $a \in \mathbb{R}$. It is easy to see that this set is of the form

$$\cap_{x^i \in \mathbb{X}^i}\left\{\eta^i : E_{\bar{Q}_{\mu^i}(\cdot|s^i,x^i)}\left[\pi^i(x^i, Y^i) + \eta^i(x^i) + \delta E_{p_{\mathbb{S}^i}}\left[H^i(B^i(\mu^i, s^i, x^i, Y^i), S^i)\right]\right] < a\right\}.$$

Each set in the intersection is trivially (Borel) measurable, therefore the intersection

(of finitely many) of them is also measurable.

We now define the Bellman operator $H^i \in L^\infty(\Delta(\Theta^i) \times \mathbb{S}^i) \mapsto T^i[H^i]$ where

$$T^i[H^i](\mu^i, s^i) \equiv \int_{\mathbb{V}^i} \left\{ \max_{x^i \in \mathbb{X}^i} E_{\bar{Q}_{\mu^i}(\cdot|s^i,x^i)} \left[\pi^i(x^i, Y^i) + \eta^i(x^i) + \delta E_{p_{\mathbb{S}^i}} \left[H^i(B^i(\mu^i, s^i, x^i, Y^i), S^i) \right] \right] \right\} P_{\mathbb{V}^i}(d\eta^i).$$

By our first result, the operator is well-defined. Moreover, since $\int ||\eta^i|| P_{\mathbb{V}^i}(d\eta^i) < \infty$ and π^i is uniformly bounded, it follows that T^i maps $L^\infty(\Delta(\Theta^i) \times \mathbb{S}^i)$ into itself. By Blackwell's sufficient conditions, there exists a unique $V^i \in L^\infty(\Delta(\Theta^i) \times \mathbb{S}^i)$ such that $V^i = T^i[V^i]$.

In order to establish continuity of V^i, by standard arguments it suffices to show that T^i maps $C(\Delta(\Theta^i) \times \mathbb{S}^i)$ into itself, where $C(\Delta(\Theta^i) \times \mathbb{S}^i) \equiv \{f \in L^\infty(\Delta(\Theta^i) \times \mathbb{S}^i) \colon \mu^i \mapsto f(\mu^i, s^i) \text{ is continuous, for all } s^i\}$. Suppose that $H^i \in C(\Delta(\Theta^i) \times \mathbb{S}^i)$. Since $\mu^i \mapsto B^i(\mu^i, s^i, x^i, y^i)$ is also continuous for all (s^i, x^i, y^i), by the Dominated convergence theorem, it follows that $\mu^i \mapsto \int_{\mathbb{S}^i} H^i(B^i(\mu^i, s^i, x^i, y^i), \hat{s}^i) p_{\mathbb{S}^i}(d\hat{s}^i)$ is continuous, for all (s^i, x^i, y^i). This result and the fact that $\theta^i \mapsto E_{Q^i_{\theta^i}(y^i|s^i,x^i)} \left[\int_{\mathbb{S}^i} H^i(B^i(\tilde{\mu}^i, s^i, x^i, y^i), \hat{s}^i) p_{\mathbb{S}^i}(d\hat{s}^i) \right]$ is bounded and continuous (for a fixed $\tilde{\mu}^i$), readily implies that

$$\mu^i \mapsto E_{\bar{Q}_{\mu^i}(\cdot|s^i,x^i)} \left[E_{p_{\mathbb{S}^i}} \left[H^i(B^i(\mu^i, s^i, x^i, Y^i), S^i) \right] \right]$$

is also continuous. This result and the fact that $\mu^i \mapsto E_{\bar{Q}_{\mu^i}(\cdot|s^i,x^i)} [\pi^i(x^i, Y^i)]$ is continuous ($\theta^i \mapsto \sum_{y^i \in \mathbb{Y}^i} \pi^i(x^i, y^i) Q_{\theta^i}(y^i|s^i, x^i)$ is continuous and bounded), imply that T^i maps $C(\Delta(\Theta^i) \times \mathbb{S}^i)$ into itself.

The fact that Φ^i single-valued $a.s.-P_{\mathbb{V}^i}$, i.e., for all (μ^i, s^i), $P_{\mathbb{V}^i}(\eta^i \colon |\Phi^i(\mu^i, s^i, \eta^i)| > 1) = 0$, follows because the set of η^i such that $|\Phi^i(\mu^i, s^i, \eta^i)| > 1$ is of dimension lower than $|\mathbb{X}^i|$ and, by absolute continuity of $P_{\mathbb{V}^i}$, this set has measure zero.

To show continuity of $\mu^i \mapsto \Phi^i(\mu^i, s^i, \eta^i)$, observe that, by the previous calculations, $(\mu^i, x^i) \mapsto E_{\bar{Q}_{\mu^i}(\cdot|s^i,x^i)} \left[\pi^i(x^i, Y^i) + \eta^i(x^i) + \delta E_{p_{\mathbb{S}^i}} [V^i(\hat{\mu}^i, S^i)] \right]$ is continuous (under the product topology) for all s^i and $a.s. - P_{\mathbb{V}^i}$. Also, $\mathbb{X}^i$ is compact. Thus by the theorem of the maximum, $\mu^i \mapsto \Phi^i(\mu^i, s^i, \eta^i)$ is continuous, $a.s. - P_{\mathbb{V}^i}$. $\square$

Proof of Theorem 4. Let $(\bar{\mu}^i)_{i \in I}$ be a belief profile that supports σ as an equilibrium. Consider the following policy profile $\phi = (\phi^i_t)_{i,t}$: For all $i \in I$ and all t,

$$(\mu^i, s^i, \eta^i) \mapsto \phi^i_t(\mu^i, s^i, \eta^i) \equiv \begin{cases} \varphi^i(\bar{\mu}^i, s^i, \eta^i) & \text{if } \max_{i \in I} ||\bar{Q}^i_{\mu^i} - \bar{Q}^i_{\bar{\mu}^i}|| \leq \frac{1}{2C} \varepsilon_t \\ \varphi^i(\mu^i, s^i, \eta^i) & \text{otherwise,} \end{cases}$$

where φ^i is an arbitrary selection from Φ^i, $C \equiv \max_I \left\{|\mathbb{Y}^i| \times \sup_{\mathbb{X}^i \times \mathbb{Y}^i} |\pi^i(x^i, y^i)|\right\} < \infty$, and the sequence $(\varepsilon_t)_t$ will be defined below. Fix any prior profile μ_0 such that $0 < \mu_0^i(\Theta^i(\sigma)) < 1$ and $\mu_0^i(\cdot|\Theta^i(\sigma)) = \bar{\mu}^i$ for all $i \in I$ (where for any $A \subset \Theta$ Borel, $\mu(\cdot|A)$ is the conditional probability given A). This is possible because $\Theta^i(\sigma) \neq \Theta^i$ for all $i \in I$.

We now show that if $\varepsilon_t \geq 0$ for all t and $\lim_{t\to\infty} \varepsilon_t = 0$, then ϕ is asymptotically optimal. Throughout this argument, we fix an arbitrary $i \in I$. Abusing notation, let $U^i(\mu^i, s^i, \eta^i, x^i) = E_{\bar{Q}_{\mu^i}(\cdot|s^i,x^i)}[\pi^i(x^i, Y^i) + \eta^i(x^i)]$. Since $\delta^i = 0$, it suffices to show that

$$U^i(\mu^i, s^i, \eta^i, \phi_t^i(\mu^i, s^i, \eta^i)) \geq U^i(\mu^i, s^i, \eta^i, x^i) - \varepsilon_t \tag{35}$$

for all (i, t), all (μ^i, s^i, η^i), and all x^i. By construction of ϕ, equation (35) is satisfied if $\max_{i\in I} ||\bar{Q}^i_{\mu^i} - \bar{Q}^i_{\bar{\mu}^i}|| > \frac{1}{2C}\varepsilon_t$. If, instead, $\max_{i\in I} ||\bar{Q}^i_{\mu^i} - \bar{Q}^i_{\bar{\mu}^i}|| \leq \frac{1}{2C}\varepsilon_t$, then

$$U^i(\bar{\mu}^i, s^i, \eta^i, \phi_t^i(\mu^i, s^i, \eta^i)) = U^i(\bar{\mu}^i, s^i, \eta^i, \varphi^i(\bar{\mu}^i, s^i, \eta^i)) \geq U^i(\bar{\mu}^i, s^i, \eta^i, x^i), \tag{36}$$

for all $x^i \in \mathbb{X}^i$. Moreover, for all x^i,

$$\begin{aligned}
\left|U^i(\bar{\mu}^i, s^i, \eta^i, x^i) - U^i(\mu^i, s^i, \eta^i, x^i)\right| &= \left|\sum_{y^i \in \mathbb{Y}^i} \pi(x^i, y^i)\left\{\bar{Q}^i_{\bar{\mu}^i}(y^i \mid s^i, x^i) - \bar{Q}^i_{\mu^i}(y^i \mid s^i, x^i)\right\}\right| \\
&\leq \sup_{\mathbb{X}^i\times\mathbb{Y}^i} |\pi^i(x^i, y^i)| \sum_{y^i\in\mathbb{Y}^i} \left|\left\{\bar{Q}^i_{\bar{\mu}^i}(y^i \mid s^i, x^i) - \bar{Q}^i_{\mu^i}(y^i \mid s^i, x^i)\right\}\right| \\
&\leq \sup_{\mathbb{X}^i\times\mathbb{Y}^i} |\pi^i(x^i, y^i)| \times |\mathbb{Y}^i| \times \max_{y^i, x^i, s^i} \left|\bar{Q}^i_{\bar{\mu}^i}(y^i \mid s^i, x^i) - \bar{Q}^i_{\mu^i}(y^i \mid s^i, x^i)\right|
\end{aligned}$$

so by our choice of C, $|U^i(\bar{\mu}^i, s^i, \eta^i, x^i) - U^i(\mu^i, s^i, \eta^i, x^i)| \leq 0.5\varepsilon_t$ for all x^i. Therefore, equation (36) implies equation (35); thus ϕ is asymptotically optimal if $\varepsilon_t \geq 0$ for all t and $\lim_{t\to\infty} \varepsilon_t = 0$.

We now construct a sequence $(\varepsilon_t)_t$ such that $\varepsilon_t \geq 0$ for all t and $\lim_{t\to\infty} \varepsilon_t = 0$. Let $\bar{\phi}^i = (\bar{\phi}^i_t)_t$ be such that $\bar{\phi}^i_t(\mu^i, \cdot, \cdot) = \varphi^i(\bar{\mu}^i, \cdot, \cdot)$ for all μ^i; i.e., $\bar{\phi}^i$ is a stationary policy that maximizes discounted utility under the assumption that the belief is always $\bar{\mu}^i$. Let $\zeta^i(\mu^i) \equiv 2C||\bar{Q}^i_{\mu^i} - \bar{Q}^i_{\bar{\mu}^i}||$ and suppose (the proof is at the end) that

$$\boldsymbol{P}^{\mu_0,\bar{\phi}}(\lim_{t\to\infty} \max_{i\in I} |\zeta^i(\mu_t^i(h))| = 0) = 1 \tag{37}$$

(recall that $\boldsymbol{P}^{\mu_0,\bar{\phi}}$ is the probability measure over $\mathbb{H}$ induced by the policy profile $\bar{\phi}$; by

definition of $\bar{\phi}$, $\boldsymbol{P}^{\mu_0,\bar{\phi}}$ does not depend on μ_0). Then by the 2nd Borel-Cantelli lemma (Billingsley (1995), pages 59-60), for any $\gamma > 0$, $\sum_t \boldsymbol{P}^{\mu_0,\bar{\phi}}(\max_{i\in I} |\zeta^i(\mu_t^i(h))| \geq \gamma) < \infty$. Hence, for any $a > 0$, there exists a sequence $(\tau(j))_j$ such that

$$\sum_{t\geq\tau(j)} \boldsymbol{P}^{\mu_0,\bar{\phi}}\left(\max_{i\in I} |\zeta^i(\mu_t^i(h))| \geq 1/j\right) < \frac{3}{a}4^{-j} \tag{38}$$

and $\lim_{j\to\infty}\tau(j) = \infty$. For all $t \leq \tau(1)$, we set $\varepsilon_t = 3C$, and, for any $t > \tau(1)$, we set $\varepsilon_t \equiv 1/N(t)$, where $N(t) \equiv \sum_{j=1}^{\infty} 1\{\tau(j) \leq t\}$. Observe that, since $\lim_{j\to\infty}\tau(j) = \infty$, $N(t) \to \infty$ as $t \to \infty$ and thus $\varepsilon_t \to 0$.

Next, we show that

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

where $(\sigma_t)_t$ is the sequence of intended strategies given ϕ, i.e.,

$$\sigma_t^i(h)(x^i \mid s^i) = P_{\mathbb{V}^i}\left(\eta^i : \phi_t^i(\mu_t^i(h), s^i, \eta^i) = x^i\right).$$

Observe that, by definition,

$$\sigma^i(x^i|s^i) = P_{\mathbb{V}^i}\left(\eta^i : x^i \in \arg\max_{\hat{x^i}\in\mathbb{X}^i} E_{\bar{Q}_{\bar{\mu}^i}(\cdot|s^i,\hat{x}^i)}\left\{\pi^i(\hat{x}^i, Y^i) + \eta^i(\hat{x}^i)\right\}\right).$$

Since $\varphi^i \in \Phi^i$ and $\delta^i = 0$, it follows that $\sigma^i(x^i|s^i) = P_{\mathbb{V}^i}(\eta^i : \varphi^i(\bar{\mu}^i, s^i, \eta^i) = x^i)$. Let $H \equiv \{h\colon \|\sigma_t(h) - \sigma\| = 0, \text{ for all } t\}$. Note that it is sufficient to show that $\mathbf{P}^{\mu_0,\phi}(H) = 1$. To show this, observe that

$$\begin{aligned}
\mathbf{P}^{\mu_0,\phi}(H) \geq& \mathbf{P}^{\mu_0,\phi}\left(\cap_t\{\max_i \zeta^i(\mu_t) \leq \varepsilon_t\}\right)\\
=& \prod_{t=\tau(1)+1}^{\infty} \mathbf{P}^{\mu_0,\phi}\left(\max_i \zeta^i(\mu_t) \leq \varepsilon_t \mid \cap_{l<t}\{\max_i \zeta^i(\mu_l) \leq \varepsilon_l\}\right)\\
=& \prod_{t=\tau(1)+1}^{\infty} \mathbf{P}^{\mu_0,\phi}\left(\max_i \zeta^i(\mu_t) \leq \varepsilon_t \mid \cap_{l<t}\{\max_i \zeta^i(\mu_l) \leq \varepsilon_l\}\right)\\
=& \mathbf{P}^{\mu_0,\phi}\left(\cap_{t>\tau(1)}\{\max_i \zeta^i(\mu_t) \leq \varepsilon_t\}\right),
\end{aligned}$$

where the second line omits the term $\mathbf{P}^{\mu_0,\phi}(\max_i \zeta^i(\mu_t) < \varepsilon_t \text{ for all } t \leq \tau(1))$ because

it is equal to 1 (since $\varepsilon_t \geq 3C$ for all $t \leq \tau(1)$); the third line follows from the fact that $\phi^i_{t-1} = \bar{\phi}^i_{t-1}$ if $\zeta^i(\mu_{t-1}) \leq \varepsilon_{t-1}$, so the probability measure is equivalently given by $\boldsymbol{P}^{\mu_0,\bar{\phi}}$; and where the last line also uses the fact that $\boldsymbol{P}^{\mu_0,\bar{\phi}}\left(\max_i \zeta^i(\mu_t) < \varepsilon_t \text{ for all } t \leq \tau(1)\right) = 1$. In addition, for all $a > 0$,

$$\begin{aligned}
\boldsymbol{P}^{\mu_0,\bar{\phi}}\left(\cap_{t>\tau(1)}\{\max_i \zeta^i(\mu_t) \leq \varepsilon_t\}\right) =& \boldsymbol{P}^{\mu_0,\bar{\phi}}\left(\cap_{n\in\{1,2,\ldots\}} \cap_{\{t>\tau(1):N(t)=n\}} \{\max_i \zeta^i(\mu_t) \leq n^{-1}\}\right) \\
\geq& 1 - \sum_{n=1}^{\infty} \sum_{\{t:N(t)=n\}} \boldsymbol{P}^{\mu_0,\bar{\phi}}\left(\max_i \zeta^i(\mu_t) \geq n^{-1}\right) \\
\geq& 1 - \sum_{n=1}^{\infty} \frac{3}{a} 4^{-n} = 1 - \frac{1}{a},
\end{aligned}$$

where the last line follows from (38). Thus, we have shown that $\mathbf{P}^{\mu_0,\phi}(H) \geq 1 - 1/a$ for all $a > 0$; hence, $\mathbf{P}^{\mu_0,\phi}(H) = 1$.

We conclude the proof by showing that equation (37) indeed holds. Observe that σ is trivially stable under $\bar{\phi}$. Also, even though μ^i_0 might not have full support, Lemma 4 still holds because $\mu^i_0(\Theta^i(\sigma)) > 0$ and $\Theta^i(\sigma) \neq \Theta^i$—see footnote X. Then, for all $i \in I$ and all open sets $U^i \supseteq \Theta^i(\sigma)$,

$$\lim_{t\to\infty} \mu^i_t\left(U^i\right) = 1 \tag{39}$$

$a.s. - \boldsymbol{P}^{\mu_0,\bar{\phi}}$ (over $\mathbb{H}$). Let $\mathcal{H}$ denote the set of histories such that $x^i_t(h) = x^i$ and $s^i_t(h) = s^i$ implies that $\sigma^i(x^i \mid s^i) > 0$. By definition of $\bar{\phi}$, $\boldsymbol{P}^{\mu_0,\bar{\phi}}(\mathcal{H}) = 1$. Thus, it suffices to show that $\lim_{t\to\infty} \max_{i\in I} |\xi^i(\mu^i_t(h))| = 0$ a.s.-$\boldsymbol{P}^{\mu_0,\bar{\phi}}$ over $\mathcal{H}$. To do this, take any $A \subseteq \Theta$ that is closed. By equation (39), for all $i \in I$, and almost all $h \in \mathcal{H}$,

$$\limsup_{t\to\infty} \int 1_A(\theta)\mu^i_{t+1}(d\theta) = \limsup_{t\to\infty} \int 1_{A\cap\Theta^i(\sigma)}(\theta)\mu^i_{t+1}(d\theta).$$

Moreover,

$$\begin{aligned}
\int 1_{A\cap\Theta^i(\sigma)}(\theta)\mu^i_{t+1}(d\theta) \leq& \int 1_{A\cap\Theta^i(\sigma)}(\theta) \left\{ \frac{\prod_{\tau=1}^t Q^i_\theta(y^i_\tau|s^i_\tau,x^i_\tau)\mu^i_0(d\theta)}{\int_{\Theta^i(\sigma)} \prod_{\tau=1}^t Q^i_\theta(y^i_\tau|s^i_\tau,x^i_\tau)\mu^i_0(d\theta)} \right\} \\
=& \mu^i_0(A|\Theta^i(\sigma)) \\
=& \bar{\mu}^i(A),
\end{aligned}$$

where the first line follows from the fact that $\Theta^i(\sigma) \subseteq \Theta$ and $\prod_{\tau=1}^{t} Q_\theta^i(y_\tau^i|s_\tau^i, x_\tau^i) \geq 0$; the second line follows from the fact that, since $h \in \mathcal{H}$, the fact that the model is identifiable implies that $\prod_{\tau=1}^{t} Q_\theta^i(y_\tau^i|s_\tau^i, x_\tau^i)$ is constant with respect to θ for all $\theta \in \Theta^i(\sigma)$, and the last line follows from our choice of μ_0^i. Therefore, we established that a.s.-$\boldsymbol{P}^{\mu_0,\bar{\phi}}$ over $\mathcal{H}$, $\limsup_{t\to\infty} \mu_{t+1}^i(h)(A) \leq \bar{\mu}^i(A)$ for A closed. By the portmanteau lemma, this implies that, a.s. -$\boldsymbol{P}^{\mu_0,\bar{\phi}}$ over $\mathcal{H}$,

$$\lim_{t\to\infty} \int_\Theta f(\theta)\mu_{t+1}^i(h)(d\theta) = \int_\Theta f(\theta)\bar{\mu}^i(d\theta)$$

for any f real-valued, bounded and continuous. Since, by assumption, $\theta \mapsto Q_\theta^i(y^i|s^i, x^i)$ is bounded and continuous, the previous display applies to $Q_\theta^i(y^i|s^i, x^i)$, and since y, s, x take a finite number of values, this result implies that $\lim_{t\to\infty} ||\bar{Q}^i_{\mu_t^i(h)} - \bar{Q}^i_{\bar{\mu}^i}|| = 0$ for all $i \in I$ a.s. -$\boldsymbol{P}^{\mu_0,\bar{\phi}}$ over $\mathcal{H}$. $\square$