OSIPP Discussion Paper: DP-2023-E-002

Higher-Order Misspecification and Equilibrium Stability

August 2, 2023

Takeshi Murooka*

Osaka School of International Public Policy (OSIPP), Osaka University

Yuichi Yamamoto†

Institute of Economic Research, Hitotsubashi University

Abstract: This paper considers a Bayesian learning problem where strategic players jointly learn an unknown economic state, and show that one's higher-order misspecification (i.e., one's misspecification about the opponent's misspecification) can have a significant impact on the equilibrium outcome. We consider a simple environmental problem where players' production, as well as an unknown state, affects the quality of the environment. Crucially, we assume that one of the players is unrealistically optimistic about the quality of the environment. When this optimism is common knowledge, the equilibrium outcome is continuous in the amount of optimism, and hence small optimism leads to approximately correct learning of the state. In contrast, when the optimism is not common knowledge and each player is unaware of the opponent having a different view about the world, the equilibrium outcome is discontinuous, and even vanishingly small optimism leads to completely incorrect learning. We then analyze a general Bayesian learning model and discuss when such discontinuity arises.

JEL Codes: C73, D83, D90, D91

Keywords: model misspecification, learning, unawareness, convergence, stability, inferential naivety, overconfidence

* Email: murooka@osipp.osaka-u.ac.jp

† Email: yyamamoto@ier.hit-u.ac.jp

Higher-Order Misspecification and Equilibrium Stability*

Takeshi Murooka
Osaka University

Yuichi Yamamoto
Hitotsubashi University

August 2, 2023

Abstract

This paper considers a Bayesian learning problem where strategic players jointly learn an unknown economic state θ, and show that one's higher-order misspecification (i.e., one's misspecification about the opponent's misspecification) can have a significant impact on the equilibrium outcome. We consider a simple environmental problem where players' production, as well as an unknown state θ, affects the quality of the environment. Crucially, we assume that one of the players is unrealistically optimistic about the quality of the environment. When this optimism is common knowledge, the equilibrium outcome is continuous in the amount of optimism, and hence small optimism leads to approximately correct learning of the state θ. In contrast, when the optimism is not common knowledge and each player is unaware of the opponent having a different view about the world, the equilibrium outcome is discontinuous, and even vanishingly small optimism leads to completely incorrect learning. We then analyze a general Bayesian learning model and discuss when such discontinuity arises.

JEL CODE: C73, D83, D90, D91

Keywords: model misspecification, learning, unawareness, convergence, stability, inferential naivety, overconfidence

*This paper subsumes some results of our earlier working paper "Misspecified Bayesian Learning by Strategic Players: First-Order Misspecification and Higher-Order Misspecification." We thank Matthias Fahn, Drew Fudenberg, Kevin He, Paul Heidhues, Botond Kőszegi, Shintaro Miura, Yoko Okuyama, Klaus Schmidt, and seminar and conference audiences for helpful comments. Murooka acknowledges financial support from JSPS KAKENHI (JP16K21740, JP18H03640, JP19K01568, JP20K13451, JP22K13365). Yamamoto acknowledges financial support from JSPS KAKENHI (JP20H00070, JP20H01475).

1 Introduction

Economic agents often take actions based on a misspecified view about the world: A worker may be overconfident about his own capability, a firm may incorrectly assume that the demand function is linear in prices (in reality, the demand is non-linear), an investor may incorrectly believe that the economy is driven by fewer variables, and so on.[1] Recent literature on model misspecification studies how such a bias influences the agents' behavior and payoffs, assuming either a single-agent setup or a multi-agent setup in which the agents' misspecifications are common knowledge (e.g., Esponda and Pouzo, 2016; Heidhues, Kőszegi, and Strack, 2018; Ba and Gindin, 2023). However, this common knowledge assumption leaves out many potential applications, as it does not allow players' *higher-order misspecification*. For example, when a worker is overconfident about his own capability, his colleague may not be aware of it; in this case, this colleague has a misspecified view about the opponent's view about the world.

This paper shows that such higher-order misspecification has a significant impact on players' play. In particular, we find that even a negligible amount of misspecification can drastically change the equilibrium outcome. To illustrate this, we consider a simple model of an environmental problem with two players. There are infinitely many periods, and the players' production, as well as an unknown state θ, influences the quality of the environment each period. We assume that one of the players (say, player 2) is misspecified and is unrealistically optimistic about the quality of the environment. Players are myopic, and actions are unobservable.

In Section 2, we consider a benchmark case in which the players do not have higher-order misspecification, and each player correctly understands the opponent's view about the world. That is, player 1 knows that she is more pessimistic than player 2, while player 2 knows that she is more optimistic than player 1. (Also, these beliefs are common knowledge.) In this case, we find that after a

[1] As experimental and empirical evidence, people exhibit overconfidence in strategic entries (Camerer and Lovallo, 1999), corporate investments (Malmendier and Tate, 2005), and merger decisions (Malmendier and Tate, 2008). See Daniel and Hirshleifer (2015), Malmendier and Tate (2015), and Grubb (2015) for reviews of the literature.

long time, the players' beliefs about the unknown state θ converge to a steady state almost surely, regardless of the initial common prior. This steady state outcome is continuous with respect to the level of player 2's bias (optimism). So when player 2's bias is small, its impact on the long-run outcome is small, i.e., after a long time, the players have approximately correct beliefs about θ.

Then Section 3 considers the model with higher-order misspecification, where the players are unaware of the opponent having a different view about the world. Specifically, player 2 is optimistic about the environment, and on top of that, she naively thinks that the opponent shares the same view with her. (In reality, player 1 isn't optimistic.) Similarly, player 1 has the unbiased view about the environment, and naively thinks that player 2 also has the unbiased view.

As in the benchmark case, when player 2's optimism is small, there is a steady state in which the players learn the true state θ^* almost correctly. However, it turns out that this steady state is unstable, and the players' beliefs converge there with *zero* probability; indeed, the players' beliefs tend to be polarized over time, and they converge to a boundary point almost surely. This result shows that the players' higher-order misspecification has a significant impact on the long-run equilibrium outcome. With small optimism of player 2, the players approximately learn the true state if the optimism is common knowledge, while their beliefs converge to boundary points if the players are unaware of the opponent having a different view about the world.

Our result also shows discontinuity of the equilibrium outcome with respect to the information structure. Indeed, the players' long-run beliefs are concentrated on the true state in the case of no misspecification, but these beliefs jump to boundary points once player 2 has (even vanishingly small) optimism. One may think that this discontinuity contradicts with various continuity results in the literature on incomplete information; e.g., Chen, Di Tillio, Faingold, and Xiong (2017) show that a small change in information structure can have only a small impact on equilibrium in any normal-form game. We will explain how to reconcile this in Section 3.3.

In our model, instability of the steady state is closely related to the inferential naivety arising from the players' higher-order misspecification. Since player 1 is unaware of player 2's optimism, she believes that player 2 will maximize payoffs and update the belief given the unbiased view about the environment (but in reality, player 2 does so given the optimistic view about the world). It turns out that this inferential naivety (about the opponent's action and belief) is *reinforced* through learning in our model; a small gap between one's belief about the opponent's belief (about the state) and the opponent's actual belief can become arbitrarily large after a long time. As we will explain in Section 3.3, this is the source of the instability of the steady state.

Recent work by Frick, Iijima, and Ishii (2020) shows that a small misspecification can lead to a complete breakdown of correct learning, in the context of social learning. In their model, agents observe the opponents' actions every period and learn a payoff-relevant unknown state from it. The agents are misspecified in that they have incorrect views about how the opponents interpret information (and hence they have incorrect views about the opponents' behavior). They show that a steady state is discontinuous in the amount of misspecification, and in particular, even with a vanishingly small amount of misspecification, in the unique steady state, the agents have a point-mass belief on a state which is far away from the true state.

Frick, Iijima, and Ishii (2020) also argue that their result relies on the assumption that the agents have only a limited amount of information about the state, in that the agents observe a noisy signal about the state only once. (So the agents learn mostly from the opponents' actions, and in this sense it is a model of social learning.) Indeed, they show that if the agents observe signals in every period of the infinite-horizon model, then the result is overturned and steady states are continuous in the amount of misspecification. So one may naturally expect that a small misspecification can destroy correct learning in social learning models, but not in models where agents receive feedbacks (signals) repeatedly. Note that repeated feedbacks are common in many economic applications; e.g., if agents observe their own payoffs every period, then it is a model of repeated feedbacks, as payoffs are informative about the state in general.[2]

Our result shows that such a conjecture is not true, and a small misspecification can still have a huge impact on the learning outcome even in a model of repeated feedbacks. In our model, when the agents become slightly misspecified, the probability of the belief converging to the steady state suddenly drops from one to zero. So even though a small misspecification has only a negligible impact on the steady state, it leads to a complete breakdown of correct learning. This is a new mechanism which causes discontinuity of the learning outcome, and in this sense our work complements Frick, Iijima, and Ishii (2020).

In Section 3.4, we extend this non-convergence result to a general setup, and study when a small amount of misspecification leads to a complete breakdown of correct learning. We find that learning tends to be fragile when the state θ and one's belief about the state θ have opposite impacts on the outcome. As we will explain, this condition can be satisfied in a wide range of economic applications, such as team production and Cournot competition.

To prove the fragility of correct learning, we extend the non-convergence result of Pemantle (1990), which shows that if a steady state of a stochastic process is unstable in some sense, then the process converges there with zero probability. His theorem does not apply to our setup directly, for three reasons. First, we assume that players observe *public* signals and update their beliefs, so the stochastic shocks on these beliefs are perfectly correlated. Assumption (iii) in Theorem 1 of Pemantle does not allow such a correlation. Second, we consider a Gaussian noise, which violates the bounded support assumption of Pemantle. Third, the drift term of our stochastic process involves a perturbation term, which is not considered by Pemantle.[3] We show that these features do not cause a serious problem, and the result of Pemantle remains valid in our environment. We believe that this

[2]Frick, Iijima, and Ishii (2020) assume that the agents do not observe payoffs.

[3]Benaïm (1999) also prove a similar non-convergence theorem, but his result does not apply to our model for the same reason. Benaïm and Faure (2012) prove a non-convergence result which allows a Gaussian noise, but they assume that the process is cooperative. Also they make various assumptions on the noise term which are not satisfied in our model (e.g., i.i.d. noise, positive-definite assumption which rules out perfect correlation of a noise).

result can be useful for future works which consider a stochastic process similar to ours (in particular, problems in which stochastic shocks are Gaussian or shocks on multiple variables are perfectly correlated).

2 First-Order Misspecification

2.1 Setup

As a benchmark, we first consider the case in which higher-order misspecification does not exist, i.e., each player correctly understands the opponent's view about the world. There are two players $i = 1,2$ and infinitely many periods $t = 1,2,\cdots$. At the beginning of the game, an unobservable economic state θ^* is drawn from a closed interval $\Theta = [\underline{\theta},\overline{\theta}]$, according to a common prior distribution $\mu \in \triangle\Theta$. We assume that μ has a continuous density μ' with full support. In each period t, each player i has a belief $\mu_i^t \in \triangle\Theta$ about the state θ, and chooses an action x_i from a closed interval $X_i = [0,\overline{x}_i]$. Player i's action x_i is not observable by the opponent $j \neq i$. Given an action profile $x = (x_1,x_2)$, the players observe a noisy public signal $y = Q(x_1,x_2,a,\theta^*)+\varepsilon$, where $a \in \boldsymbol{R}$ is a fixed parameter and ε is a random noise whose distribution is log-concave with mean zero. Player i's stage-game payoff is $u_i(x_i,y)$. We assume that both Q and u_i are twice continuously differentiable.

Crucially, we assume that one of the players (player 2) incorrectly believes that the true parameter is $A \neq a$, while the other player is unbiased and knows the parameter a. These first-order beliefs (about the parameter a) are common knowledge, e.g., player 1 knows that player 2 believes that the true parameter is $A \neq a$. We call it *first-order misspecification*, because player 2 has an incorrect first-order belief about the parameter a.

Player 1's subjective expected stage-game payoff given an action profile x and a state θ is

$$U_1(x,\theta) = E[u_1(x_1,Q(x,a,\theta)+\varepsilon)]$$

and player 2's subjective expected stage-game payoff is

$$U_2(x,A,\theta) = E[u_2(x_2,Q(x,A,\theta)+\varepsilon)],$$

where the expectation is taken with respect to ε. Note that player 2 evaluates payoffs given her subjective signal distribution $Q(x,A,\theta)+\varepsilon$. To economize notation, we will write $U_2(x,\theta)$ instead of $U_2(x,A,\theta)$ when it does not cause a confusion.

We assume that players play a static Nash equilibrium every period. This essentially means that in our model, (i) players are myopic, and (ii) they predict the opponent's play correctly and best-respond to it. Condition (i) shuts down the repeated-game effect, so that a result similar to the folk theorem (which is not of our interest) does not arise.[4] Condition (ii) implies that players recognize that the opponent also learns the state and changes the action as time goes. This setup is different from the one in the literature on learning in games (e.g., Fudenberg and Kreps, 1993; Esponda and Pouzo, 2016), which asks when and why players play equilibria; they assume that players do not know the opponent's strategy and learn it from experience. In our model, players know the opponent's strategy, and learn only the unknown economic state θ.[5]

In period one, players play a Nash equilibrium (x_1^1, x_2^1). Assuming an interior solution, it is an action profile which solves the first-order condition $\frac{\partial E[U_i(x,\theta)|\mu]}{\partial x_i} = 0$ for each i. At the end of period one, players observe a public signal y^1, and update the posterior beliefs using Bayes' rule. Assuming that no one has deviated in period one, each player i's posterior belief μ_1^2, μ_2^2 in period two is

$$\mu_1^2(\theta) = \frac{\mu_1^1(\theta) f(y - Q(x^1,a,\theta))}{\int_\Theta \mu_1^1(\tilde{\theta}) f(y - Q(x^1,a,\tilde{\theta}))d\tilde{\theta}} \quad \text{and} \quad \mu_2^2(\theta) = \frac{\mu_2^1(\theta) f(y - Q(x^1,A,\theta))}{\int_\Theta \mu_2^1(\tilde{\theta}) f(y - Q(x^1,A,\tilde{\theta}))d\tilde{\theta}},$$

where x^1 is the Nash equilibrium played in period one and f is the density function of the noise term ε. Note that player 2's posterior μ_2^2 differs from player 1's posterior μ_1^2, as she incorrectly believes that the mean output is $Q(x^1,A,\theta)$ rather than $Q(x^1,a,\theta)$. Because the players' information structure about the parameter a is common knowledge, these posteriors are common knowledge among

[4] Another way to avoid the repeated-game effect is to use a Markov-perfect equilibrium (where the state is players' beliefs about θ) as a solution concept. With an additional assumption, Appendix A shows that players' long-run behavior is exactly the same as that of myopic players studied in this section. In this sense, our result remains true even for forward-looking players.

[5] Condition (ii) is inessential if the game is dominance solvable. Note that all the examples studied in this paper are actually dominance solvable.

players. So in period two, players play a Nash equilibrium given the belief profile $\mu^2 = (\mu_1^2, \mu_2^2)$, which solves $\frac{\partial E[U_i(x,\theta)|\mu_i^2]}{\partial x_i} = 0$ for each i. Likewise, in any subsequent period t, players play a Nash equilibrium given the belief profile $\mu^t = (\mu_1^t, \mu_2^t)$, where μ^t is computed by Bayes' rule.

A *steady state* in this dynamic learning model is a pair $(x_1^*, x_2^*, \mu_1^*, \mu_2^*)$ of an action profile and a belief profile which satisfies the following four conditions:

$$x_1^* \in \arg\max_{x_1} U_1(x_1, x_2^*, \theta^*), \tag{1}$$

$$x_2^* \in \arg\max_{x_2} U_2(x_1^*, x_2, \theta_2), \tag{2}$$

$$\mu_1^* = 1_{\theta^*}, \tag{3}$$

$$\mu_2^* = 1_{\theta_2} \text{ s.t. } \theta_2 \in \arg\min_{\theta \in \Theta} |Q(x^*, A, \theta) - Q(x^*, a, \theta^*)|. \tag{4}$$

Conditions (1) and (2) are incentive compatibility, which requires that each player maximizes her payoff given some beliefs. The other two conditions require that these beliefs satisfy consistency: Condition (3) asserts that the unbiased player 1 correctly learns the true state θ^* in a steady state. Condition (4) requires that player 2's belief is concentrated on a state θ_2 which best explains the data, in that with this state θ_2, player 2's subjective view about the mean output is closest to the actual mean. This condition must be satisfied in a steady state; otherwise, player 2 is "surprised" by observed signals and her belief will move to the state which better explains the data. In many economic applications (including the environmental problem example in the next subsection), Condition (4) reduces to

$$Q(x^*, A, \theta_2) = Q(x^*, a, \theta^*), \tag{5}$$

i.e., the subjective mean output exactly matches the true mean.

As we will show in Section 2.3, under a mild condition called *identifiability*, players' actions and beliefs converge to this steady state almost surely. So this steady state can be thought of as a "long-run outcome" of the dynamic learning model.

2.2 Application: Optimism in Environmental Problems

To see how the steady state looks like, we will consider a model of environmental problems, which includes air pollution, deforestation, and fishery as special cases.[6] Every period, each player $i = 1,2$ chooses a production level $x_i \in [0,1]$, which has a negative impact on the quality of the environment. As in Chapter 24 of Varian (1992), we assume that the quality of the environment is given by the formula

$$y = Q(x,a,\theta) + \varepsilon = a - \theta(x_1 + x_2) + \varepsilon, \tag{6}$$

where $a \in \boldsymbol{R}$ is a fixed parameter, $\theta \in \Theta = [0.7, 0.9]$ is an unknown fundamental, and ε is a noise term which follows the standard normal distribution $N(0,1)$. Player i's payoff is $y + x_i - c(x_i)$, where x_i is a private benefit from production and $c(x_i) = \frac{1}{2}x^2$ is a production cost. Since we assume $\theta \in (0,1)$, regardless of players' beliefs, the Nash equilibrium in the one-shot game is an interior point. In what follows, we will assume that the true state is $\theta^* = 0.8$.

We assume that one of the players is unrealistically optimistic about the quality of the environment. Such optimism is commonly observed in various environmental problems, as discussed by Dechezleprêtre et al. (2022) and references therein. Formally, we assume that player 2 incorrectly believes that the true parameter is $A > a$.

This example satisfies the sufficient condition for convergence presented in Section 2.3.[7] So regardless of the initial prior, the players' actions and beliefs almost surely converge to the steady state, which is characterized by (1)-(4). We will consider how player 2's bias influences this steady-state outcome.

Let Q_z denote the derivative of Q with respect to a variable z. Since $Q_a > 0$ and $Q_\theta < 0$, Condition (5) implies that player 2's steady-state belief is $\theta_2 > \theta^*$, i.e., the optimistic player overestimates the state in the long run. Intuitively, player 2 is disappointed by observed environmental quality being worse than the antic-

[6] More generally, this is a model of production with negative externalities.

[7] Indeed, since Q is linear in θ and ε follows the normal distribution, the function $K_2(\theta, x)$ defined in Section 2.3 is convex. Hence, $K_2(\theta, \sigma)$ is also convex and has a unique minimizer.

ipation, and becomes pessimistic about the state θ as time goes. This in turn implies that player 2 overestimates the marginal social cost Q_{x_i} of the production.[8] Thus, her steady-state action x_2 is lower than in the correctly-specified case. On the other hand, the unbiased player 1's production is exactly the same as in the correctly-specified case, because player 1's optimal production is independent of the opponent's action. Accordingly, player 2's payoff is lower than in the correctly-specified case, while player 1's payoff is higher than that. So player 2's bias is detrimental for herself, but *improves* the opponent's payoff. Also, in this steady state, player 2's bias improves the social surplus as well.

2.3 Convergence under First-Order Misspecification

Now we will show that players' beliefs indeed converge to a steady state under a mild condition called *identifiability*.

Given an action profile x and a state θ, define the *Kullback-Leibuler divergence* for player 2 as

$$K_2(\theta, x) = E\left[\log \frac{q(y|x,A,\theta)}{q(y|x,a,\theta^*)}\middle| x,a,\theta^*\right] = \int q(y|x,a,\theta^*)\log\frac{q(y|x,A,\theta)}{q(y|x,a,\theta^*)}dy,$$

where $q(y|x,a,\theta)$ is a probability density function of the signal y given (x,a,θ). Intuitively, this KL divergence measures the difference between player 2's subjective signal distribution $q(\cdot|x,A_2,\theta)$ and the true distribution $q(\cdot|x,a,\theta^*)$. For example, when the noise term ε follows the standard normal distribution $N(0,1)$, the above formula reduces to

$$K_2(\theta, x) = \frac{(Q(x,\theta,A) - Q(x,\theta^*,a))^2}{2},$$

which is increasing in the difference between the true mean and subjective mean.

Then for each probability measure $\sigma \in \triangle X$ and state θ, define the *weighted Kullback-Leibuler divergence* as

$$K_2(\theta,\sigma) = \int_X K_2(\theta,x)\sigma(dx).$$

[8] Attari et al. (2010) report that people overestimate certain energy-saving activities (e.g., driving less to save gasoline), but their answers also have substantial variation.

Roughly, this $K_2(\theta,\sigma)$ measures the average difference between the subjective signal distribution and the true distribution in the past, given the past action frequency σ.

The *identifiability* condition requires that for each action frequency $\sigma \in \triangle X$, the Kullback-Leibuler divergence $K_2(\theta,\sigma)$ has a unique minimizer $\theta_2(\sigma)$ and it is an interior point.[9] This KL minimizer $\theta_2(\sigma)$ can be thought of as a state θ which best fits the true distribution (in the sense that the KL divergence is minimized). For the degenerate distribution $\sigma = 1_x$, the KL minimizer $\theta_2(\sigma)$ is simply the state which solves Condition (4).

Esponda, Pouzo, and Yamamoto (2021) show that for a single-agent problem, the identifiability condition above is sufficient for convergence of the agent's belief, i.e., regardless of the initial prior with full support, the agent's belief converges to a steady state almost surely. The following proposition shows that the same result holds for our model of first-order misspecification. Intuitively, in our model, player 1 is unbiased and learns the true state almost surely regardless of players' play; so there is only one player (player 2) whose belief evolves in a nontrivial way, and it is natural to expect that the result for a single-agent problem extends.

Proposition 1. *Suppose that the identifiability condition holds, and that for each state θ, there is a unique pure-strategy Nash equilibrium.*[10] *Then almost surely, player 2's belief converges to the steady state belief, i.e., $\lim_{t\to\infty}\mu_2^t = 1_{\theta_2}$ where θ_2 is a steady state belief.*

[9] Here we do not need to think about player 1's Kullback-Leibuler divergence, because she is unbiased and hence its unique minimizer is θ^* regardless of the action frequency.

[10] Player 2's belief converges even when there are multiple Nash equilibria for some parameter θ. In such a case, however, the limiting belief may be a *mixed-action* steady state (Berk-Nash equilibrium).

3 Higher-Order Misspecification

The benchmark model in the previous section assumed that players correctly understand the opponent's view about the world. Now we consider the case in which players may have *higher-order misspecification*, in that they may have a biased view about the opponent's view about the world (second-order misspecification), a biased view about the opponent's second-order misspecification, and so on.[11]

In what follows, we will focus on a special form of higher-order misspecification: We will assume that each player has a biased view about the world, and on top of that, she naively thinks that the opponent shares the same view about the world (in reality, the opponent has her own view about the world). We call it *double misspecification*, because players have a biased view about the world (first-order misspecification) and a biased view about the opponent's view about the world (second-order misspecification). Of course, we can think of various other forms of higher-order misspecification. In Appendix A, we will present a more general model of higher-order misspecification.[12]

3.1 Setup: Double Misspecification

Our model is the same as the one studied in Section 2.1, except the information structure; now we will assume that each player i (incorrectly) believes that it is common knowledge that the signal y is given by $y = Q(x_1, x_2, A_i, \theta) + \varepsilon$. We allow $A_1 \neq A_2$, so the different players may have different levels of misspecification.

A critical difference from the first-order misspecification is that players have *inferential naivety* and make incorrect predictions about the opponent's play.[13] Indeed, while player i believes that the opponent (player j) maximizes the payoff

[11] As evidence from laboratory experiments, subjects often systematically mispredict other subjects' preferences and actions (e.g., Van Boven, Dunning, and Loewenstein, 2000). Ludwig and Nafziger (2011) report that most subjects in their experiments are not aware of or underestimate overconfidence of other subjects.

[12] In static games with strategic complementarity/substitutability, recent work by McGee (2023) analyzes how certain higher-order misspecification affects equilibrium outcomes.

[13] See Eyster (2019) for a review of the literature.

conditional on the parameter A_i, the opponent maximizes the payoff conditional on the parameter A_j in reality. Accordingly, player i's prediction about the opponent's action does not match the opponent's actual action in general.

To analyze players' behavior in the presence of such inferential naivety, it is useful to consider two *hypothetical players* $i = 1,2$. Hypothetical player i is player j who thinks that it is common knowledge that the true technology is A_j. Intuitively, player j thinks that hypothetical player i is her opponent, and hence each period, player j chooses a Nash equilibrium action against hypothetical player i.

Let $\hat{x}_i$ and $\hat{\mu}_i$ denote hypothetical player i's action and belief, and let $x = (x_1,x_2,\hat{x}_1,\hat{x}_2)$ denote an action profile in the four-player game. Player i's expected stage-game payoff is defined as

$$U_i(x,\theta,A_i) = E[u_i(x_i, Q(x_i,\hat{x}_{-i},A_i,\theta)+\varepsilon)],$$

because she thinks that the parameter is A_i and the opponent is a hypothetical player. Similarly, hypothetical player i's expected stage-game payoff given θ is

$$\hat{U}_i(x,\theta,A_{-i}) = E[u_i(\hat{x}_i, Q(\hat{x}_i,x_{-i},A_{-i},\theta)+\varepsilon)].$$

Using these notations, the equilibrium strategy in the infinite-horizon game is described as follows. In period one, all players have the same belief $\mu_i^1 = \hat{\mu}_i^1 = \mu$. So they play a Nash equilibrium $(x_1^1,x_2^1,\hat{x}_1^1,\hat{x}_2^1)$, which (assuming interior solutions) satisfies the first-order conditions $\frac{\partial E[U_i(x,\theta)|\mu]}{\partial x_i} = 0$ and $\frac{\partial E[\hat{U}_i(x,\theta)|\mu]}{\partial \hat{x}_i} = 0$. At the end of period one, players observe a public signal $y^1 = Q(x_1^1,x_2^1,a,\theta^*)+\varepsilon$, and update the posterior beliefs using Bayes' rule. Their beliefs in period two are given by

$$\mu_i^2(\theta) = \frac{\mu_i^1(\theta)f(y-Q(x_i^1,\hat{x}_{-i}^1,A_i,\theta))}{\int_\Theta \mu_i^1(\tilde{\theta})f(y-Q(x_i^1,\hat{x}_{-i}^1,A_i,\tilde{\theta}))d\tilde{\theta}},$$
$$\hat{\mu}_i^2(\theta) = \frac{\hat{\mu}_i^1(\theta)f(y-Q(\hat{x}_i^1,x_{-i}^1,A_{-i},\theta))}{\int_\Theta \hat{\mu}_i^1(\tilde{\theta})f(y-Q(\hat{x}_i^1,x_{-i}^1,A_{-i},\tilde{\theta}))d\tilde{\theta}}.$$

As is clear from this formula, player i's posterior belief is biased in two ways: She updates the belief conditional on the wrong parameter A_i, and on the wrong prediction $\hat{x}^1_{-i}$ about the opponent's play. Then in period two, players play a Nash equilibrium given this belief profile $\mu^2 = (\mu_1^2, \mu_2^2, \hat{\mu}_1^2, \hat{\mu}_2^2)$.[14] Likewise, in any subsequent period t, players play a Nash equilibrium given the posterior beliefs computed by Bayes' rule.

Given an action profile $x = (x_1, x_2, \hat{x}_1, \hat{x}_2)$, let $\theta_i(x, A_i)$ denote player i's long-run belief when the same action x is chosen every period. That is, let $\theta_i(x, A_i)$ be a state θ which solves

$$\min_{\theta\in\Theta} |Q(x_i, \hat{x}_j, A_i, \theta) - Q(x_1, x_2, a, \theta^*)|,$$

so that player i's subjective model best explains the actual output. Note that when we evaluate the subjective mean, we take into account the fact that player i believes that the opponent's action is $\hat{x}_j$ (rather than x_j). We will assume that $\theta_2(x, A)$ is unique for each x and A_i. As in the case of first-order misspecification, in many examples, the consistency condition above reduces to $Q(x_i, \hat{x}_j, A_i, \theta_i) = Q(x_1, x_2, a, \theta^*)$.

With this notation, a *steady state* under double misspecification is defined as $(x_1^*, x_2^*, \hat{x}_1^*, \hat{x}_2^*, \mu_1^*, \mu_2^*, \hat{\mu}_1^*, \hat{\mu}_2^*)$ which satisfies

$$x_i^* \in \arg\max_{x_i} U_i(x_i, \hat{x}^*_{-i}, A_i, \theta_i) \quad \forall i, \tag{7}$$

$$\hat{x}_i^* \in \arg\max_{\hat{x}_i} \hat{U}_i(\hat{x}_i, x^*_{-i}, A_{-i}, \theta_{-i}) \quad \forall i, \tag{8}$$

$$\mu_1^* = \hat{\mu}_2^* = 1_{\theta_1(x^*, A_1)}, \tag{9}$$

$$\mu_2^* = \hat{\mu}_1^* = 1_{\theta_2(x^*, A_2)}. \tag{10}$$

The first two conditions are the incentive-compatibility conditions, which require that each player maximize her own payoff given some beliefs. The next two conditions require that these beliefs satisfy consistency, in that each (actual and hypothetical) player's belief is concentrated on a state with which her subjective signal distribution coincides with the objective distribution.

[14] Since y is public, player 1 correctly predicts hypothetical player 2's posterior belief $\hat{\mu}_2^2$, and similarly, hypothetical player 2 correctly predicts player 1's posterior belief μ_1^2. So they will indeed play a Nash equilibrium given these beliefs.

Recall that in the case of first-order misspecification, players' actions and beliefs converge to a steady state if the identifiability condition holds. It turns out that under double misspecification, such a result does not hold, and the identifiability condition need not ensure convergence to a steady state. As we show in the next subsection, one of such examples is the environmental problem studied in Section 2.2.

3.2 Environmental Problem under Double Misspecification

Consider the environmental problem in section 2.2, but assume now that players are not aware of the fact that the opponent has a different view about the world. Specifically, consider double misspecification with parameters $A_1 = a$ and $A_2 \geq a$. In this setup, player 1 is unaware of player 2's optimism and naively thinks that player 2 also knows a. To simplify the exposition, assume that the initial prior is a uniform distribution on $\Theta = [0.7, 0.9]$.

The steady state in this setup is characterized by the conditions (7)-(10). For the special case in which $A_1 = A_2 = a$ (i.e., the case with no misspecification), there are three steady states: One of the steady state is an interior point, in which both players learn the true state ($\theta_1 = \theta_2 = \theta^*$), and choose the Nash equilibrium at this state θ^*. The remaining two steady states are boundary points. In these steady states, players' beliefs are $(\theta_1, \theta_2) = (\underline{\theta}, \overline{\theta})$ or $(\theta_1, \theta_2) = (\overline{\theta}, \underline{\theta})$, and they choose a Nash equilibrium given these beliefs.[15] However, these boundary steady states do not arise as a long-run outcome; since there is no misspecification, starting from a common prior μ, players learn the true state with probability one, i.e., the beliefs converge to the interior steady state almost surely when $A_1 = A_2 = a$.

Now, consider the case in which player 2 is slightly optimistic (i.e., A_2 is a bit

[15] To see that $(\theta_1, \theta_2) = (\underline{\theta}, \overline{\theta})$ is a steady state belief, note that $\frac{\partial^2 Q}{\partial x_i \partial \theta} < 0$, so we have $x_1 > \hat{x}_1$ in this steady state. This means that player 2 underestimates the opponent's production, and thus finds that the quality of the environment is worse than the anticipation. This makes player 2 more pessimistic, but her current belief $\overline{\theta}$ already hits the upper bound of the set Θ, so her belief stays there. Similarly, player 1's belief stays at $\underline{\theta}$, which imply that $(\theta_1, \theta_2) = (\underline{\theta}, \overline{\theta})$ is indeed a steady state belief. For the same reason, $(\theta_1, \theta_2) = (\overline{\theta}, \underline{\theta})$ is a steady state belief.

larger than a). By the continuity, there is an interior steady state in which players' beliefs are close to (θ^*, θ^*). Let $m^* = (m_1^*, m_2^*)$ denote this steady state belief. Also, the boundary points $(\theta_1, \theta_2) = (\underline{\theta}, \overline{\theta})$ and $(\theta_1, \theta_2) = (\overline{\theta}, \underline{\theta})$ are still steady states in this case.

One may expect that the beliefs converge to the interior steady state m^*, just as in the case of no misspecification. Proposition 2 shows that such a conjecture is incorrect, and the beliefs converge to the boundary points when $A_2 > a$.

Proposition 2. *(i) Suppose that $A_2 = a$. Then almost surely, players eventually learn the true state θ^*, i.e.,*

$$\Pr\left(\lim_{t\to\infty}(\mu_1^t, \mu_2^t) = (1_{\theta^*}, 1_{\theta^*})\right) = 1.$$

(ii) There is $\overline{A}_2 > a$ such that for any $A_2 \in (a, \overline{A}_2)$, players' posterior beliefs $\mu^t = (\mu_1^t, \mu_2^t)$ converges to the interior steady state 1_{m^*} with zero probability. Indeed, almost surely, the beliefs will be concentrated on boundary points, i.e.,

$$\Pr\left(\lim_{t\to\infty}\mu^t \in \{(1_{\underline{\theta}}, 1_{\overline{\theta}}), (1_{\overline{\theta}}, 1_{\underline{\theta}})\}\right) = 1.$$

Proposition 2 shows that unawareness about the opponent's bias can have a huge impact on the equilibrium outcome. Recall that in the case of first-order misspecification, the beliefs converge to the steady state regardless of the parameter A_2. Since the steady state outcome is continuous in A_2, this means that small optimism of player 2 has only a marginal impact on the long-run outcome, and players approximately learn the true state. In contrast, once players are unaware of the opponent having a different view about the world, the long-run outcome becomes discontinuous at $A_2 = a$, and even vanishingly small optimism completely changes the learning outcome.

In the literature of incomplete-information games, it is well-known that an equilibrium in a normal-form game is continuous with respect to the information structure; Chen, Di Tillio, Faingold, and Xiong (2017) show that a small perturbation of one's belief hierarchy (a belief about an economic state, a belief about the opponent's belief about the state, and so on) has only a marginal impact on the equilibrium. Our Proposition 2 above does not contradict with this result. Indeed, in our model, the *equilibrium strategy* in the infinite-horizon game, which maps one's belief μ_i to an action, is continuous in the parameter A_2, so a small perturbation of one's belief hierarchy has a negligible impact on the equilibrium strategy.[16] In this sense, the main result of Chen, Di Tillio, Faingold, and Xiong (2017) still holds in our model. However, this need not imply that the resulting *equilibrium outcome* is continuous in the parameter A_2, and Proposition 2 shows that our model is one of the cases in which such discontinuity arises.

3.3 Proof Sketch of Proposition 2

Now we will describe an outline of the proof of Proposition 2 (ii). Given an action profile $x = (x_1, x_2, \hat{x}_1, \hat{x}_2)$, player i believes that the signal y is generated by the formula

$$y = A_i - \theta(x_i + \hat{x}_{-i}) + \varepsilon$$

which can be rewritten as

$$\theta - \frac{\varepsilon}{x_i + \hat{x}_{-i}} = \frac{A_i - y}{x_i + \hat{x}_{-i}}.$$

Hence, if she observes a signal y, then the likelihood of the state $\theta \in \Theta$ is the truncated normal distribution on the set Θ, induced by a normal distribution with mean $\frac{A_i - y}{x_i + \hat{x}_{-i}}$ and variance $\left(\frac{1}{x_i + \hat{x}_{-i}}\right)^2$.[17] For shorthand notation, let $I_i(x)$ denote the inverse of this variance, i.e.,

$$I_i(x) = (x_i + \hat{x}_{-i})^2.$$

[16]In our model, the belief hierarchies induced by A_2 and $A_2' \neq A_2$ are close in the uniform-weak topology of Chen, Di Tillio, Faingold, and Xiong (2017) if A_2 and A_2' are close.

[17]The truncated normal distribution is derived from a normally distributed random variable by bounding the random variable from either below or above (or both). For example, when a random variable X follows $N(\mu, \frac{1}{\xi})$, the truncated normal distribution on $\Theta = [\underline{\theta}, \overline{\theta}]$ is obtained by conditioning X on $\underline{\theta} \leq X \leq \overline{\theta}$.

Intuitively, this $I_i(x)$ measures the informativeness of the signal y for player i given an action profile x; high I_i implies low variance, meaning that the signal is more informative.

Since the initial prior is uniform and the likelihood induced by signals are truncated normal distributions, each player's posterior belief is also a truncated normal distribution. Let $\tilde{N}(m,\sigma^2)$ denote the truncated normal distribution induced by the normal distribution $N(m,\sigma^2)$. Then player i's posterior at the beginning of period $t+1$ is the truncated normal distribution $\tilde{N}(m_i^{t+1}, \frac{1}{t\xi_i^{t+1}})$, where the parameters m_i^{t+1} and ξ_i^{t+1} are given by

$$m_i^{t+1} = \frac{\sum_{\tau=1}^{t} I_i(x_i^\tau, \hat{x}_{-i}^\tau)\left(\frac{A_i - y^\tau}{x_i^\tau + \hat{x}_{-i}^\tau}\right)}{\sum_{\tau=1}^{t} I_i(x_i^\tau, \hat{x}_{-i}^\tau)}, \tag{11}$$

$$\xi_i^{t+1} = \frac{1}{t}\sum_{\tau=1}^{t} I_i(x_i^\tau, \hat{x}_{-i}^\tau). \tag{12}$$

In words, the parameter m_i^{t+1} is the weighted average of player i's estimate $\frac{A_i - y^\tau}{x_i^\tau + \hat{x}_{-i}^\tau}$ each period, where the weight is the informativeness I_i. The parameter ξ_i^{t+1} is simply the average of the informativeness I_i of the past signals. Our goal is to show that this posterior belief $\tilde{N}(m_i^{t+1}, \frac{1}{t\xi_i^{t+1}})$ does not converge to the interior steady state.

Step 1: Difference Equation We first show that the motion of the parameters (m_i^t, ξ_i^t) can be described by a system of difference equations. Given an action profile x, let $\theta_i(x)$ be a solution to $Q(x_i, \hat{x}_{-i}, A_i, \theta) = Q(x_1, x_2, a, \theta^*)$, i.e.,

$$\theta_i(x) = \frac{A_i - a + \theta^*(x_1 + x_2)}{x_i + \hat{x}_{-i}}.$$

Intuitively, this $\theta_i(x)$ can be thought of as player i's estimate of θ when the noise is zero (i.e., $\varepsilon = 0$).[18] This suggests that player i's actual estimate $\frac{A_i - y}{x_i + \hat{x}_{-i}}$ appearing

[18]For some x, $\theta_i(x)$ defined above may not be in the state space Θ. So $\theta_i(x)$ should be regarded as a parameter which best explains the data, *when the choice of θ is not restricted on the state space* Θ. In contrast, $\theta_i(x, A_i)$ defined in the previous subsection must be chosen from the state space Θ.

in (11) can be represented as $\theta_i(x)$ plus a noise, and we indeed have[19]

$$\frac{A_i - y}{x_i + \hat{x}_{-i}} = \theta_i(x) - \frac{\varepsilon}{\sqrt{I_i(x_i, \hat{x}_{-i})}}.$$

Plugging the above equation into (11) and arranging it and (12), we obtain the following recursive equations which completely describe the evolution of (m^t, ξ^t).

$$m_i^{t+1} - m_i^t = \frac{1}{t}\left\{\frac{I_i(x_i^t, \hat{x}_{-i}^t)\left(\theta_i(x^t) - m_i^t - \frac{\varepsilon^t}{\sqrt{I_i(x_i^t, \hat{x}_{-i}^t)}}\right)}{\frac{t-1}{t}\xi_i^t + \frac{1}{t}I_i(x_i^t, \hat{x}_{-i}^t)}\right\}, \tag{13}$$

$$\xi_i^{t+1} - \xi_i^t = \frac{1}{t}\left(I_i(x_i^t, \hat{x}_{-i}^t) - \xi_i^t\right). \tag{14}$$

In words, (13) implies that player i updates the mean belief m_i^t depending on how her estimate $\theta_i(x^t) - \frac{\varepsilon^\tau}{\sqrt{I_i(x_i^t, \hat{x}_{-i}^t)}}$ based on the new information today differs from her current mean belief m_i^t. If the new estimate coincides with the current mean belief, she does not update it. Otherwise, the mean belief moves toward the new estimate $\theta_i(x^t) - \frac{\varepsilon^\tau}{\sqrt{I_i(x_i^t, \hat{x}_{-i}^t)}}$, and its magnitude is influenced by the informativeness of the signals; if $I_i(x_i^t, \hat{x}_{-i}^t)$ is relatively larger than $\frac{t-1}{t}\xi_i^t + \frac{1}{t}I_i(x_i^t, \hat{x}_{-i}^t)$, it means that the signal today is more informative relative to the past signals, and hence influences the posterior more.

The second equation (14) has a similar interpretation, and ξ_i^t is updated depending on how the informativeness $I_i(x_i^t, \hat{x}_{-i}^t)$ of the signal today differs from the informativeness ξ_i^t of the past signals.

Note that players' actions $(x_i^t, \hat{x}_{-i}^t)$ in period t is a one-shot Nash equilibrium given the posterior belief $\tilde{N}(m_i^t, \frac{1}{(t-1)\xi_i^t})$, so all terms in the right-hand sides of the above equations are the functions of (m_i^t, ξ_i^t), except the noise term ε^t. Hence the above difference equations are indeed recursive, i.e., once we fix the current value (m_i^t, ξ_i^t) and the noise term ε^t, the next value (m_i^{t+1}, ξ_i^{t+1}) is uniquely determined.

Step 2: Stochastic Approximation Since the difference equations derived in Step 1 involves a stochastic noise term ε, finding its exact solution is a hard prob-

[19]This equation follows from the definition of $\theta_i(x)$ and $y = a - \theta^*(x_1 + x_2) + \varepsilon$.

lem. So instead, we borrow the idea of stochastic approximation and use the fact that the solution to the difference equations can be approximated by much simpler *differential equations*.

Recall that player i's posterior belief in period t is the truncated normal distribution $\tilde{N}(m_i^t, \frac{1}{(t-1)\xi_i^t})$, where m_i^t is the mean and $\frac{1}{(t-1)\xi_i^t}$ is the variance of the (untruncated) normal distribution. When t is sufficiently large, this variance $\frac{1}{(t-1)\xi_i^t}$ approaches zero, and hence the belief $\tilde{N}(m_i^t, \frac{1}{(t-1)\xi_i^t})$ is approximately a degenerate belief. Given m_i, let $\tilde{N}(m,0)$ denote this limiting belief, that is,

$$\tilde{N}(m,0) = \lim_{\sigma^2 \to 0} \tilde{N}(m,\sigma^2) = \begin{cases} 1_m & \text{if } m \in [\underline{\theta}, \overline{\theta}] \\ 1_{\underline{\theta}} & \text{if } m < \underline{\theta} \\ 1_{\overline{\theta}} & \text{if } m > \overline{\theta} \end{cases}$$

where $\underline{\theta} = 0.7$ and $\overline{\theta} = 0.9$ denote the boundary points of the state space $\Theta = [0.7, 0.9]$.

Then players' actions $(x_i^t, \hat{x}_{-i}^t)$ in period t can be approximated by the one-shot Nash equilibrium given this limiting belief $\tilde{N}(m_i^t, 0)$. Let $\theta_i(m_1^t, m_2^t)$ and $I_i(m_i^t)$ denote player i's estimate and the signal informativeness when players play this Nash equilibrium. That is, $\theta_i(m_1, m_2) = \theta_i(x)$ and $I_i(m_i) = I_i(x_i, \hat{x}_{-i})$, where $(\hat{x}_j, \hat{x}_{-j})$ is the Nash equilibrium for the belief $\tilde{N}(m_j, 0)$ for each j.

This in turn implies that the drift terms of the difference equations (13) and (14) are approximated by much simpler terms $\frac{I_i(m_i^t)(\theta_i(m^t)-m_i^t)}{\xi_i^t}$ and $I_i(m_i^t) - \xi_i^t$, in the sense that there is $K > 0$ such that for any t and for $\alpha = 0.5$,

$$\left| \frac{I_i(x_i^t, \hat{x}_{-i}^t)(\theta_i(x^t) - m_i^t)}{\frac{t-1}{t}\xi_i^t + \frac{1}{t} I_i(x_i^t, \hat{x}_{-i}^t)} - \frac{I_i(m_i^t)(\theta_i(m^t) - m_i^t)}{\xi_i^t} \right| < \frac{K}{t^\alpha}, \tag{15}$$

$$\left| \left(I_i(x_i^t, \hat{x}_{-i}^t) - \xi_i^t\right) - \left(I_i(m_i^t) - \xi_i^t\right) \right| < \frac{K}{t^\alpha}. \tag{16}$$

Then it follows from the theory of stochastic approximation (e.g., Theorem 2.1 of Kushner and Yin (2003)) that the asymptotic behavior of the process (13) and (14) is approximated by the ordinal differential equations (ODE)

$$\frac{dm_i(t)}{dt} = \frac{I_i(m_i(t))(\theta_i(m(t)) - m_i(t))}{\xi_i(t)}, \tag{17}$$

$$\frac{d\xi_i(t)}{dt} = I_i(m_i(t)) - \xi_i(t), \tag{18}$$

which do not involve a noise term.

Step 3: Instability of the Interior Steady State Figure 1 is the phase portrait which describes the solution to the ODE (17), when there is no misspecification (i.e., $A_2 = a$) and the variable $\xi(t)$ is fixed at $\xi_1(t) = \xi_2(t) = 1$ for all t.[20] The origin is the point in which both players learn the true state (i.e., $(m_1, m_2) = (\theta^*, \theta^*)$ where $\theta^* = 0.8$), which is the interior steady state in this special case. There are only two paths converging to this steady state m^*, one from the top-right and the one from bottom-left. These paths are the *basin of attraction* of the steady state. If the initial value is not on this basin of attraction, the solution to the ODE does not converge to m^*, and it moves toward the boundary points. (This is so even if the initial value is in a neighborhood of the origin.) In this sense, the origin is an *unstable* steady state.

Figure 2 is the phase portrait when player 2 is optimistic (precisely, when $A_2 - a = 0.03$). Due to the misspecification, the origin is not a steady state; now the steady state m^* moves toward the top-right corner. Other than that, the motion of the mean belief $m(t)$ is very similar to that for the case with no misspecification. In particular, the steady state m^* is still unstable, in that there are only two paths converging to this point.

The instability of the steady state m^* here is deeply related to the inferential naivety. To see this, note first that in the steady state m^*, each player's subjective output exactly matches the objective output. Suppose now that the mean belief is

[20] In reality, the parameter ξ_i is *not* fixed, and evolves according to (18). However, this does not influence the motion of the mean belief $m(t)$ much, because the variable ξ does not influence the sign of $\frac{dm_i(t)}{dt}$ in the ODE (17), i.e., it does not influence whether the mean belief $m_i(t)$ increases or decreases at the next instant. Accordingly, the motion of $m(t)$ is very similar to the one described in Figure 1 even for the case in which $\xi(t)$ changes over time.

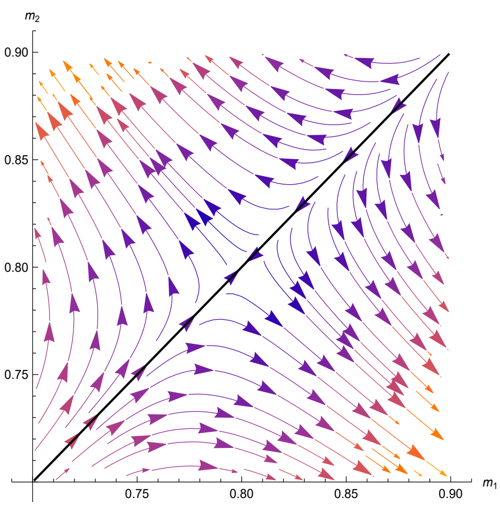

Figure 1: Motion of $m(t)$ when $A_2 = a$

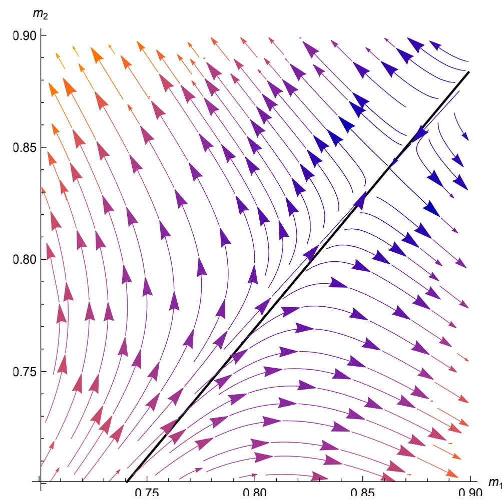

Figure 2: Motion of $m(t)$ when $A_2 > a$

perturbed toward the bottom-right direction, and we have $m(t) = (m_1(t), m_2(t)) = (m_1^* + \eta, m_2^* - \eta)$ for small $\eta > 0$. That is, we consider the case in which player 1 is more pessimistic about the state θ than in the steady state, while player 2 is more optimistic about the state. With this belief profile, player 1 reduces the production than in the steady state, while player 2 increases the production. However, due to the inferential naivety, player 1 misestimates the opponent's production; player 1 incorrectly believes that the opponent is similarly pessimistic and reduces the production (in reality, the opponent increases the production). This means that player 1 observes the environmental quality worse than her anticipation, and becomes even more pessimistic about the state θ. Similarly, player 2 becomes even more optimistic. Hence the gap between players' beliefs become larger, and the belief profile $m(t)$ moves toward the bottom-right corner, rather than moving back to the origin. This process continues over and over; so even if the initial gap η is small, it becomes arbitrarily large, and the belief converges to the boundary point.

Remark 1. For the argument in the last paragraph of this step to work, it is critical that the state θ and one's belief about θ have conflicting effects on the outcome y, in the sense that Q_θ (which measures the effect of θ) and $Q_{x_i}\frac{\partial x_i}{\partial m_i}$ (which measures the effect of player i's belief about θ through her action) have opposite signs. To see this, suppose that the current mean belief is $m(t) = (m_1(t), m_2(t)) = (m_1^* + \eta, m_2^* - \eta)$ as in the discussion above, and suppose now that $Q_\theta > 0$ and $Q_{x_i}\frac{\partial x_i}{\partial m_i} > 0$.[21] The latter inequality implies that player 1's inferential naivety leads her to overestimate the output, i.e., she finds that the actual output is worse than her anticipation. Then because we have $Q_\theta > 0$, player 1's mean belief m_1 *goes down and approaches to the steady state belief* m_1^*. In Section 3.4, we will consider a general setup, and show that this property is indeed necessary for the instability. See Proposition 4.

Step 4: Non-Convergence to Unstable Steady State In the previous step, we have seen that the motion of the solution to the ODE is very sensitive to the initial value: If the initial value is on the basin of attraction of the interior steady state, the solution eventually converges to there. However, once the initial value is perturbed, the solution moves toward a different direction and converges to the boundary points. It turns out that this property is the key to obtain our discontinuity result.

To begin with, consider the case with $A_2 > a$, and suppose that the current mean belief m^t is at the steady state (or in its neighborhood). The key observation is that due to the stochastic shock ε, the mean belief m^t cannot stay at the basin of attraction of the steady state forever. Indeed, as can be seen from (13), the shock ε pushes m^t toward the direction

$$b = \left(\frac{\sqrt{I_1}}{(t-1)\xi_1 + I_1}, \frac{\sqrt{I_2}}{(t-1)\xi_2 + I_2} \right),$$

which is simplified to $b = (\frac{1}{t\sqrt{I_1}}, \frac{1}{t\sqrt{I_2}})$ in the steady state.[22] This vector b is represented by the black thick line in Figure 2, and it does not coincide with the basin of attraction of the steady state. Hence at some point, the mean belief m^t will be "kicked out" from the basin due to the shock, and then the mean belief should move toward the boundary points, rather than reverting to the steady state.[23]

[21] Note that $Q_\theta < 0$ and $Q_{x_i}\frac{\partial x_i}{\partial m_i} > 0$ in the above example.

[22] Here we use the fact that $\xi_i = I_i$ at the steady state.

[23] More precisely, in the proof of the non-convergence theorem, we show that the mean belief

In contrast, when there is no misspecification (i.e., $A_2 = a$), the mean belief m^t remains on the basin of attraction, even after the shock ε. Indeed, if the current mean belief is at the steady state, the shock pushes the mean belief toward the direction

$$b = \left(\frac{1}{t\sqrt{I_1}}, \frac{1}{t\sqrt{I_2}} \right),$$

which is proportional to $(1,1)$ in this special case. This means that the mean belief m^t remains on the 45-degree line, even after the shock. As can be seen from Figure 1, this 45-degree line is precisely the basin of attraction, and hence the solution to the ODE starting from the current mean belief reverts to the steady state. This suggests that the beliefs converge to the interior steady state in this case. Intuitively, in the case of no misspecification, player 1's posterior belief about θ coincides with player 2's belief after every history, and hence the amplifying effect is *never* triggered; accordingly the mean belief does not move toward the boundary points.

In the proof of Proposition 2, we formalize the idea above by borrowing a technique developed by Pemantle (1990), who establish a non-convergence theorem for a class of stochastic processes. Pemantle's theorem does not apply to our model directly, as our stochastic process does not satisfy some technical assumptions of Pemantle. So we show that Pemantle's result remains valid even in our setup; see Appendix B for details.

Remark 2. Heidhues, Kőszegi, and Strack (2021) consider a single-agent learning problem and show that the agent's belief does not converge to an unstable steady state. Unfortunately, their proof has a flaw: In the proof, they show that the mean belief visits the basin of attraction of stable steady states infinitely often, but this need not imply non-convergence to unstable equilibria. (As shown in Theorem 6.10 of Benaïm (1999), we need to show that the mean belief visits a *compact subset of* the basin of stable steady states infinitely often.) To fix it, one can use our non-convergence theorem; the non-convergence theorem of Heidhues,

leaves $\frac{c}{\sqrt{t}}$-neighborhood of the basin of attraction infinitely often, due to the shock ε. Once the mean belief leaves this neighborhood, the drift term (which represents the "amplifying effect" we explained in the last step) is bounded away from zero and dominates the impact of the stochastic shock ε, and hence the mean belief moves toward the boundary points as described in Figure 2.

Kőszegi, and Strack (2021) is correct, and it directly follows from our Proposition 14 in Appendix B.

3.4 Non-Convergence in a General Setup

We have seen that in the environmental problem under double misspecification, players' beliefs do not converge to the interior steady state. Now we will consider a general model and provide a condition under which discontinuity similar to that in the environmental problem occurs.

Consider a general model of double misspecification with a compact state space $\Theta = [\underline{\theta}, \overline{\theta}]$. Consider an initial prior μ with a continuous density. Assume that for any state θ, a Nash equilibrium $(x_i, \hat{x}_{-i})$ given (A_i, θ) is unique. We also assume that the noise term ε follows the standard normal distribution $N(0,1)$.

Assume that the output function Q is linear in θ, in that $Q = R(x,a)\theta + S(x,a)$, and define $I_i(x_i, \hat{x}_{-i}) = (R(x_i, \hat{x}_{-i}, A_i))^2$. This linearity assumption is a bit restrictive, but it ensures that the likelihood induced by any signal sequence $(y^1, \cdots, y^t)$ is the truncated normal distribution $\tilde{N}(m_i^t, \frac{1}{(t-1)\xi_i^t})$, where the parameters m_i^t and ξ_i^t are determined by (11) and (12) with the term $\frac{A_i - y^\tau}{x_i^\tau + \hat{x}_{-i}^\tau}$ in (11) being replaced by $\frac{y^\tau - S(x_i^\tau, \hat{x}_{-i}^\tau, A_i)}{R(x_i, \hat{x}_{-i}, A_i)}$. We assume that $R(x_i, \hat{x}_{-i}, A_i) \neq 0$ for all on-path actions $(x_i, \hat{x}_{-i})$; this implies that $\xi_i^t > 0$, and hence the distribution $\tilde{N}(m_i^t, \frac{1}{(t-1)\xi_i^t})$ is well-defined. To keep our notation as simple as possible, in what follows, we strengthen this assumption and focus on the case in which $R(x_i, \hat{x}_{-i}, A_i) < 0$ for all on-path actions $(x_i, \hat{x}_{-i})$.[24] Then just as in the environmental problem, the evolution of the parameters m_i^t and ξ_i^t is governed by the difference equations (13) and (14), where $\theta_i(x)$ is a solution to $Q(x_i, \hat{x}_{-i}, A_i, \theta) = Q(x_1, x_2, a, \theta^*)$.

Assume that the drift terms of these difference equations are approximated by the drift terms of the ODE. That is, assume that there is $K > 0$ and $\alpha > 0$ such that (15) and (16) hold. Then it follows from the theory of stochastic approximation

[24]When $R(x_i, \hat{x}_{-i}, A_i) > 0$ for some actions $(x_i, \hat{x}_{-i})$, the term $-\varepsilon^t / \sqrt{I_i(m_i^t, \frac{1}{(t-1)\xi_i^t})}$ in (13) should be replaced with $+\varepsilon^t / \sqrt{I_i(m_i^t, \frac{1}{(t-1)\xi_i^t})}$ for such actions, and it may influence the specification of the vector b. All the remaining arguments are not affected.

that the evolution of the parameters (m^t, ξ^t) is asymptotically approximated by the ODE (17) and (18). Here, the functions $I_i(m_i^t)$ and $\theta_i(m^t)$ in these equations are defined as in the environmental problem (i.e., these are the informativeness $I_i(x_i, \hat{x}_{-i})$ and the estimate $\theta_i(x)$ when players have degenerate beliefs and play a Nash equilibrium given these beliefs).

A steady state of the ODE is a point $p = (m_1, m_2, \xi_1, \xi_2)$ where $\frac{dm_i(t)}{dt} = \frac{d\xi_i(t)}{dt} = 0$. A steady state p is *linearly unstable* if the Jacobian J of the ODE at the point p has at least one eigenvalue with positive real part. When p is linearly unstable, the basin of attraction of p is locally approximated by $p + H$, where H is the space spanned by the eigenvectors associated with eigenvalues with negative real parts. So if the initial value is not in this set $p + H$, the solution to the ODE eventually leaves a neighborhood of p. The interior steady state of the environmental problem in Section 3.2 is an example of linearly unstable steady states; as described in Figure 1, when $A_1 = A_2 = a$, the solution to the ODE leaves a neighborhood of the origin unless the initial value is on the 45-degree line.

In our general model, the set H, which determines the basin of attraction of the stead state, can be computed as follows: Note that $\theta_i(m) - m_i = 0$ and $\xi_i = I_i$ in any steady state. Hence the Jacobian J of the ODE at a steady state p can be written as

$$J = \begin{pmatrix} \frac{\partial \theta_1}{\partial m_1} - 1 & \frac{\partial \theta_1}{\partial m_2} & 0 & 0 \\ \frac{\partial \theta_2}{\partial m_1} & \frac{\partial \theta_2}{\partial m_2} - 1 & 0 & 0 \\ \frac{\partial I_1}{\partial m_1} & 0 & -1 & 0 \\ 0 & \frac{\partial I_2}{\partial m_2} & 0 & -1 \end{pmatrix}.$$

Obviously this Jacobian J has an eigenvalue $\lambda = -1$ (multiplicity 2), and the corresponding eigenspace is

$$\{(0, 0, \xi_1, \xi_2) | \forall\ \xi_1, \xi_2 \in \boldsymbol{R}\}. \tag{19}$$

The remaining two eigenvalues of J are the ones for the submatrix

$$J' = \begin{pmatrix} \frac{\partial \theta_1}{\partial m_1} - 1 & \frac{\partial \theta_1}{\partial m_2} \\ \frac{\partial \theta_2}{\partial m_1} & \frac{\partial \theta_2}{\partial m_2} - 1 \end{pmatrix}.$$

So a steady state p is linearly unstable if and only if this submatrix J' has an eigenvalue with positive real part.

If the two eigenvalues of J' have positive real part, then the set H is simply the set described by (19). So the solution to the ODE leaves a neighborhood of p, unless the mean belief (m_1,m_2) of the initial value exactly matches that of the steady state belief. If the matrix J' has one eigenvalue with positive real part and one eigenvalue with negative real part, then the set H is

$$(ch_1,ch_2,\xi_1,\xi_2)|\forall c,\xi_1,\xi_2 \in \boldsymbol{R}\}$$

where $h=(h_1,h_2)$ is an eigenvector associated with the eigenvalue with negative real part. In this case, if the initial value of the mean belief (m_1,m_2) is perturbed toward a direction other than h, then it leaves the basin $p+H$, so that the solution to the ODE must leave a neighborhood of p.

The following proposition shows that the process does not converge to a linearly unstable steady state, if the noise kicks out the mean belief from its basin of attraction in the above sense. This result is a direct consequence of the general non-convergence theorem in Appendix B (Proposition 14), and hence we omit the proof. Let b denote the coefficient vector on the noise term ε in the difference equations (13) and (14) at the steady state p, i.e.,

$$b=\left(\frac{\sqrt{I}_1}{t\xi_1+I_1},\frac{\sqrt{I}_2}{t\xi_2+I_2},0,0\right)=\frac{1}{t+1}\left(\frac{1}{\sqrt{I_1}},\frac{1}{\sqrt{I_2}},0,0\right).$$

Proposition 3. *Pick the parameters (A_1,A_2) arbitrarily, and pick an initial prior μ with a continuous density. Let p be a linearly unstable steady state of the ODE (17) and (18). Assume that the following properties hold.*

(i) For each i and θ, a Nash equilibrium $(x_i,\hat{x}_{-i})$ given (A_i,θ) is unique.

(ii) The noise term ε follows the standard normal distribution $N(0,1)$,

(iii) $Q=R(x,a)\theta+S(x,a)$ and $R<0$ for all on-path actions.

(iv) Given any on-path action profile x, there is a unique $\theta\in\boldsymbol{R}$ which solves $Q(x_i,\hat{x}_{-i},A_i,\theta)=Q(x_1,x_2,a,\theta^*)$. (Hence $\theta_i(x)$ is well-defined.)

(v) The functions $I_i(m_i)$ and $\theta_i(m)$ are Lipschitz-continuous.

(vi) There is $K > 0$ such that (15) and (16) hold for $\alpha = 1$ at the steady state p.

(vii) $b \notin H$.

Then the probability of $\lim_{t\to\infty}(m^t, \xi^t) = p$ *is zero.*

The critical assumption in Proposition 3 is (vii), which implies that the process (m^t, ξ^t) cannot stay in the basin of attraction of p due to the stochastic shock ε. Proposition 3 asserts that if this assumption (as well as other standard assumptions (i)-(vi)) holds, then the process converges to the unstable steady state p with zero probability.

We view assumption (vii) as a mild restriction, because it is satisfied for generic choice of parameters. For example, in the environmental problem studied in the previous subsections, the assumption (vii) is satisfied for any value $A_2 \neq a$ in a neighborhood of a. In this sense, linear instability of p "almost always" implies non-convergence to p.

So it is important to understand when an interior steady state p is linearly unstable, and the next proposition characterizes it. We use the following notation. For each i and θ_{-i}, let $f_i^*(\theta_{-i})$ denote the set of all θ_i such that $\theta_i = \theta_i(x, A_i)$ for some $(x_1, x_2, \hat{x}_1, \hat{x}_2)$ such that $(x_i, \hat{x}_{-i})$ is a Nash equilibrium given (A_i, θ_i) for each i. Intuitively, this $f_i^*(\theta_{-i})$ is the set of steady states in player i's *single-agent learning problem*, where the opponent $-i$'s belief is fixed at θ_{-i} (and hence she chooses the same Nash equilibrium action for θ_{-i} every period), while player i incorrectly believes that the opponent $-i$'s belief changes over time.

Figure 3 describes the graph of f_i^* for the environmental problem. The blue line is the graph of $f_1^*(\theta_2)$. It shows that when θ_2 is fixed at a low value, we have $f_1^*(\theta_2) = \{\overline{\theta}_1\}$, i.e., the highest state $\overline{\theta}_1$ is the unique steady state for player 1's learning problem. Indeed, as can be seen from Figure 3, if θ_2 is low, the arrow points toward the east, so θ_1 goes up. Similarly, when θ_2 is fixed at a high value, we have $f_1^*(\theta_2) = \{\underline{\theta}_1\}$. When θ_2 takes an intermediate value, both boundary points $\underline{\theta}_1$ and $\overline{\theta}_1$ are still steady states, and on top of that, there is an interior steady state which is described by the upward-sloping blue curve in the figure. Note that at any point on this blue curve, the solution of the ODE moves toward the vertical direction only; i.e., player 1's belief m_1 does not change at the next instant. This means that these are indeed steady states for player 1's learning problem. The orange line in the figure is the graph of $f_2^*(\theta_1)$, and it can be interpreted in the same way. Note that a steady state of the joint learning problem is the intersections of the blue and orange lines.

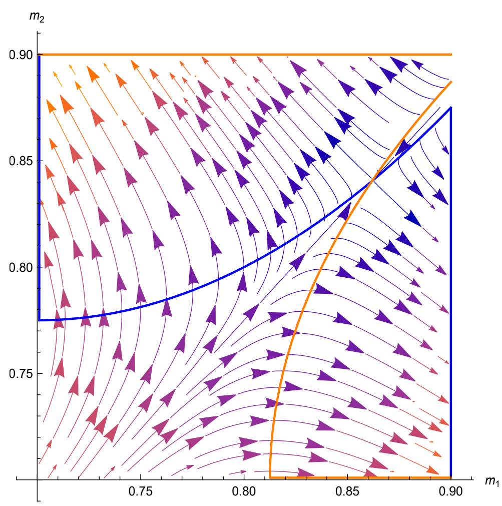

Figure 3: Example of $\frac{\partial \theta_i(m)}{\partial m_i} - 1 > 0$

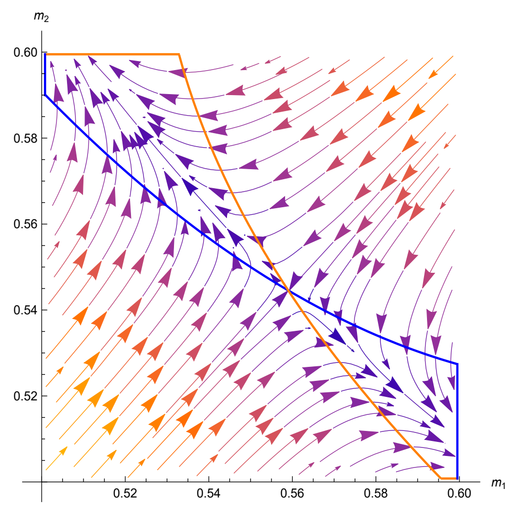

Figure 4: Example of $\frac{\partial \theta_i(m)}{\partial m_i} - 1 < 0$

Proposition 4. *Let $p = (m_1^*, m_2^*, \xi_1^*, \xi_2^*)$ be a steady state such that $Q(x_i, \hat{x}_{-i}, A_i, m_i^*) = Q(x_1, x_2, a, \theta^*)$ where $(x_1, x_2, \hat{x}_1, \hat{x}_2)$ denotes steady-state actions. Suppose also that for each i, there is an open interval U_i containing m_{-i}^* such that there is a unique continuous function $f_i : U_i \to \boldsymbol{R}$ with $f_i(m_{-i}^*) = m_i^*$ and $f_i(\theta_{-i}) \in f_i^*(\theta_{-i})$ for all $\theta_{-i} \in U_i$. Assume that f_i is differentiable. Then the following properties hold:*

(i) Suppose that $\frac{\partial \theta_i(m)}{\partial m_i} - 1 > 0$ at p for each i. Then p is linearly unstable.

(ii) Suppose that $\frac{\partial \theta_i(m)}{\partial m_i} - 1 < 0$ at p for each i. Then p is linearly unstable if $f_1'(\theta_2) f_2'(\theta_1) > 1$ at p.

The function f_i defined in Proposition 4 maps the opponent's belief θ_{-i} to player i's interior steady-state belief. For example, in Figure 3, f_1 is the blue flatter upward-sloping curve, and f_2 is the orange steeper upward-sloping curve.

Proposition 4 (i) assumes $\frac{\partial \theta_i(m)}{\partial m_i} - 1 > 0$. To interpret this assumption, consider player i's single-agent learning problem in which the opponent's belief is fixed at m^*_{-i}. In this problem, player i's posterior belief is represented by the parameters (m_i^t, ξ_i^t) just as in the original problem, and the motion of (m_i^t, ξ_i^t) is approximated by the ODE (17) and (18). The Jacobian of this system of the ODE at the steady state belief is

$$J_i = \begin{pmatrix} \frac{\partial \theta_i}{\partial m_i} - 1 & 0 \\ \frac{\partial I_i}{\partial m_i} & -1 \end{pmatrix},$$

so its eigenvalues are $\lambda = -1, \frac{\partial \theta_i}{\partial m_i} - 1$. The assumption $\frac{\partial \theta_i(m)}{\partial m_i} - 1 > 0$ ensures that the latter eigenvalue is positive, which means that the steady state is linearly unstable in this single-agent problem. Proposition 4 (i) shows that in such a case, the steady state is similarly unstable even when players jointly learn the state.

Proposition 4 (ii) assumes $\frac{\partial \theta_i(m)}{\partial m_i} - 1 < 0$, in which case the Jacobian J_i has two negative eigenvalues, $\lambda = -1, \frac{\partial \theta_i}{\partial m_i} - 1$. This means that in the single-agent learning problem, the steady state p is *asymptotically stable*; i.e., if the opponent's belief is fixed at the steady state value, any solution to the ODE starting from a neighborhood of p converges to p. Proposition 4 (ii) shows even in such a case, the steady state p can be unstable when players jointly learn the state. Specifically, if players' beliefs have strong complementarity/substitutability in that $f_1' f_2' > 1$, then the steady state is unstable. On the other hand, if $f_1' f_2' < 1$, then it is not difficult to show that the steady state is asymptotically stable, in that all eigenvalues of the Jacobian have negative real part. In this case, if the initial value is in a neighborhood of p, any solution to the ODE (17) and (18) converges to p.

The environmental problem studied in the previous subsections is one of the examples which satisfies the assumption stated in Proposition 4 (i). Indeed, as described in Figure 3, when $m_2(t)$ is fixed at the steady state value and $m(t)$ moves on the horizontal axis only, the solution leaves a neighborhood of p. Hence the steady state p is unstable in the single-agent learning problem.

We can also construct an example which satisfies the assumption stated in Proposition 4 (ii), by replacing the parameters of the environmental problem with $\Theta = [0.5, 0.6]$ and $\theta^* = 0.55$. Figure 4 describes the solution of the ODE for this modified environmental problem with small optimism (specifically, $A_2 - a = 0.003$). Now the interior steady state is asymptotically stable in the single-agent learning problem; indeed, when $m_2(t)$ is fixed at the steady state value and $m(t)$ moves on the horizontal axis only, the solution converges to p. Nonetheless this steady state is unstable in the joint learning problem, because the slope of the curve f_i is steep at the origin and $f_1' f_2' > 1$.

Proposition 4 is applicable to other economic applications as well. Here are two examples:

1. *Team production.* Suppose that two players work on a joint project. Each period, each player i chooses an effort level $x_i \in [0,1]$ and observes an output

$$y = a - \theta \left(\frac{1}{x_1 + x_2} - \frac{1}{2} \right) + \varepsilon, \tag{20}$$

   where a is the capability of the team, θ is an unknown state, and ε is a noise term which follows the standard normal distribution. Assume that the true state is $\theta^* = 0.5$. Player i's payoff is $y - c(x_i)$, where $c(x_i) = \frac{1}{8}x^2$ is a production cost. Consider the double misspecification model with $A_1 = a$ and $A_2 > a$, where player 2 is overconfident about the capability and player 1 is unaware of it. When $A_2 = a$, simple algebra shows that $f_1' f_2' \approx 17.7$ at the interior steady state $\theta^* = 0.5$. By the continuity, this implies that $f_1' f_2' > 1$ for any A_2 close to a. So from Proposition 4, the interior steady state is linearly unstable whenever player 2 has small overconfidence.

2. *Cournot duopoly with linear demand.* Suppose that each period, each firm $i = 1, 2$ chooses its quantity $x_i \in [0, \bar{x}]$, and a publicly observable market price is given by

$$y = a - \theta (x_1 + x_2) + \varepsilon,$$

   where θ is an unknown state and ε is a noise term which follows the standard normal distribution. Firm i's payoff is $yx_i - c(x_i)$, where yx_i is firm i's revenue and $c(x_i)$ is firm i's production cost. Consider the double misspecification model with $A_1 = a$ and $A_2 > a$, where firm 2 is overconfident about the demand and firm 1 is unaware of it. When $A_2 = a$, simple algebra shows that $f_1' f_2' > 1$ at the interior steady state $\theta^* = 0.5$ if and only if $c''(x_i) < 0$. By the continuity, the same is true for any $A_2 > a$ close to a. This means that if the cost function is strictly concave at the steady state action, then player 2's small overconfidence leads to instability of the steady state.

As noted in Remark 1, for the steady state to be unstable in the environmental problem, it is critical that the state θ and one's belief about θ have conflicting effects on the output y. Proposition 4 implies that the same is true for a general setup, i.e., for a steady state to be unstable, it is necessary that $\frac{\partial Q}{\partial \theta}$ and $Q_{x_i} \frac{\partial x_i}{\partial m_i}$ have opposite signs, at least for symmetric games with small misspecification. Indeed, Proposition 4 asserts that in symmetric games with $A_1 = A_2 = a$, a steady state is unstable if $\frac{\partial \theta_i}{\partial m_i} > 1$ or $|f_i'(\theta_{-i})| > 1$.[25] By the implicit function theorem, we have[26]

$$\frac{\partial \theta_i}{\partial m_i} = -\frac{Q_{x_i} \frac{\partial x_i}{\partial m_i}}{Q_\theta} \quad \text{and} \quad f_i'(\theta_{-i}) = \frac{Q_{x_i} \frac{\partial x_i}{\partial m_i}}{Q_{x_i} \frac{\partial x_i}{\partial m_i} + Q_\theta}.$$

[25] Although it is not stated in Proposition 4, it is straightforward to show that these conditions are necessary and sufficient for instability.

[26] The formal derivation is as follows. Note that $\theta_i(m_i)$ is a solution to $Q(x_i(m_i), x_{-i}, \theta^*, a) = Q(x_i(m_i), \hat{x}_{-i}(m_i), \theta_i, A)$. Let Q denote the left-hand side (the true mean) and $\hat{Q}$ denote the right-hand side (the subjective mean). Then by the implicit function theorem, we have

$$\frac{\partial \theta_i}{\partial m_i} = -\frac{\hat{Q}_{x_i} \frac{\partial x_i}{\partial m_i} + \hat{Q}_{x_{-i}} \frac{\partial \hat{x}_{-i}}{\partial m_i} - Q_{x_i} \frac{\partial x_i}{\partial m_i}}{\hat{Q}_\theta} = -\frac{Q_{x_i} \frac{\partial x_i}{\partial m_i}}{Q_\theta}.$$

Here the second equality follows from the fact that we assume $A_1 = A_2 = a$ (and hence $Q = \hat{Q}$) and symmetry. Similarly, $f_i(\theta_{-i})$ is a solution to $Q(x_i(f_i), x_{-i}(\theta_{-i}), \theta^*, a) = Q(x_i(f_i), \hat{x}_{-i}(f_i), f_i, A)$. With an abuse of notation, let Q denote the left-hand side and $\hat{Q}$ denote the right-hand side. Then

$$f_i'(\theta_{-i}) = \frac{Q_{x_{-i}} \frac{\partial x_{-i}}{\partial \theta_{-i}}}{\hat{Q}_{x_i} \frac{\partial x_i}{\partial m_i} + \hat{Q}_{x_{-i}} \frac{\partial \hat{x}_{-i}}{\partial m_i} + \hat{Q}_\theta - Q_{x_i} \frac{\partial x_i}{\partial m_i}} = \frac{Q_{x_i} \frac{\partial x_i}{\partial m_i}}{Q_{x_i} \frac{\partial x_i}{\partial m_i} + Q_\theta}.$$

Again, the second equality uses $A_1 = A_2 = a$ and symmetry.

Hence the condition for instability can be rewritten as

$$-\frac{Q_{x_i}\frac{\partial x_i}{\partial m_i}}{Q_\theta} > 1 \quad \text{or} \quad \left| \frac{Q_{x_i}\frac{\partial x_i}{\partial m_i}}{Q_{x_i}\frac{\partial x_i}{\partial m_i} + Q_\theta} \right| > 1. \tag{21}$$

It is obvious that this condition requires Q_θ and $Q_{x_i}\frac{\partial x_i}{\partial m_i}$ to have different signs, as claimed above.

3.5 Sufficient Conditions for Convergence

So far we have seen that one's unawareness about the opponent's misspecification can influence a learning dynamic and cause non-convergence to an interior steady state. However, this does not imply that one's unawareness always cause non-convergence, and indeed, there are many cases in which the beliefs *do* converge to an interior steady state.

For example, consider the team production problem discussed in the last subsection, but assume now that the output function (20) is replaced with

$$y = a + (1-\theta)(x_1 + x_2) + \varepsilon. \tag{22}$$

This new output function is similar to the previous one in that the output y is increasing in the capability a and the total effort $x_1 + x_2$, and is decreasing in the state θ. However, there is one critical difference: with this new output function, the state θ has a *negative impact* on the marginal productivity (i.e., $\frac{\partial^2 Q}{\partial x_i \partial \theta} < 0$), and and hence has a *negative impact* on the effort level (i.e., $\frac{\partial x_i}{\partial m_i} < 0$). Accordingly, Q_θ and $Q_{x_i}\frac{\partial x_i}{\partial m_i}$ have the same sign, which means that the condition for instability (21) is not satisfied.

Actually, in this example, players' beliefs converge to the interior steady state almost surely, regardless of the initial prior. Figure 6 is the phase portrait of the solution to the ODE (17) for the double-misspecification model with $A_1 = A_2 = a$, $\theta \in \Theta = [0.3, 0.7]$, $\theta^* = 0.5$, and fixed $\xi_1 = \xi_2$. As illustrated in the figure, the mean belief (m_1, m_2) converges to the steady state (the origin) for all initial values. It is not difficult to show that the same result holds even when ξ_i is not fixed, and the next proposition ensures almost sure convergence in such a case. Given a vector $x \in \boldsymbol{R}^4$ and a compact set $A \subset \boldsymbol{R}^4$, let $d(x,A) = \min_{y \in A} |x-y|$ denote the distance from x to the set A.

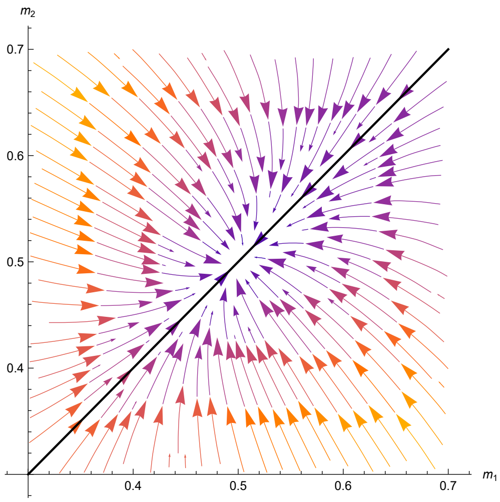

Figure 5: Motion of $m = (m_1, m_2)$

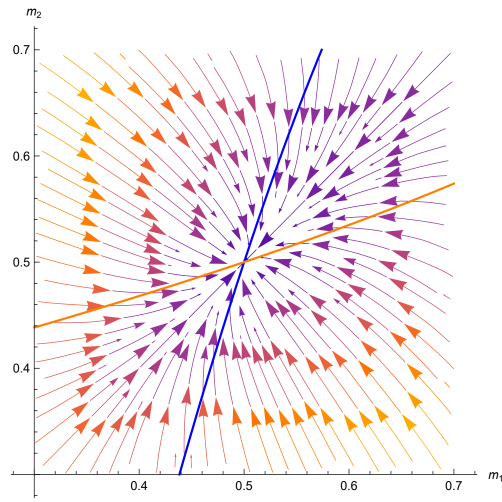

Figure 6: Graphs of f_1 and f_2

Definition 1. A set $A \subset \boldsymbol{R}^4$ is *attracting* if there is a set W such that $A \subseteq \text{int}W$ and such that for any $\varepsilon > 0$, there is $T > 0$ such that $d((m(t), \xi(t)), A) < \varepsilon$ for any initial value $(m, \xi) \in W$ and for any $t > T$. A set A is *globally attracting* if it is attracting and the stochastic process (m^t, ξ^t) moves within the basin W almost surely.

Proposition 5. *Suppose that Assumption (i)-(v) in Proposition 3 hold. Assume also that* $\liminf_{t\to\infty} m_i^t > -\infty$ *and* $\limsup_{t\to\infty} m_i^t < \infty$ *for each i with probability one, and that there is* $K > 0$ *and* $\alpha > 0$ *such that (15) and (16) hold for all t and* (m^t, ξ^t)*. If a set A is globally attracting, then the process approaches this set A almost surely, i.e., the probability of* $\lim_{t\to\infty} d((m^t, \xi^t), A) = 0$ *is one regardless of the initial prior. In particular, if A is a singleton, then the process converges to this point almost surely.*

The convergence result above relies on the assumption that the noise term ε is normally distributed and the output function Q is linear in θ. Without these assumptions, the likelihood induced by a signal sequence $(y^1, y^2, \cdots)$ need not be a truncated normal distribution, and thus it is difficult to compute the exact shape of the posterior belief μ_i^t after a long time t. Nonetheless, we can still partially characterize the limiting outcome in such a case; the next proposition establishes global convergence, with a few technical assumptions.

As in the case of first-order misspecification, we assume identifiability, which is formally defined as follows. Given a state θ and an action profile $x = (x_1, x_2, \hat{x}_1, \hat{x}_2)$, the Kullback-Leibuler divergence for player i is

$$K_i(\theta, x) = E\left[\log \frac{q(y|x_i, \hat{x}_{-i}, A_i, \theta)}{q(y|x, a, \theta^*)} \middle| x, a, \theta^*\right] = \int q(y|x, a, \theta^*) \log \frac{q(y|x_i, \hat{x}_{-i}, A_i, \theta)}{q(y|x, a, \theta^*)} dy,$$

and the Kullback-Leibuler divergence given an action frequency $\sigma \in \triangle(X_1 \times X_2 \times X_1 \times X_2)$ is

$$K_i(\theta, \sigma) = \int_X K_i(\theta, x) \sigma(dx).$$

These formulas are a bit different from those under first-order misspecification, due to the inferential naivety; player i thinks that players play $(x_i, \hat{x}_{-i})$, but the actual actions are (x_i, x_{-i}). The *identifiability* requires that given any i and σ, the Kullback-Leibulaer divergence $K_i(\theta, x)$ has a unique minimizer $\theta_i(\sigma)$ and it is an interior point.[27]

Proposition 6. *Suppose that there is a unique steady state* $(x_1^*, x_2^*, \theta_1, \theta_2)$ *and that* $Q(x_i, \hat{x}_j, A_i, \theta_i) = Q(x_1, x_2, a, \theta^*)$ *for each i in this steady state. Suppose also that the identifiability condition holds, and that for each parameter* (θ_1, θ_2)*, there is a unique Nash equilibrium* $(x_1, x_2, \hat{x}_1, \hat{x}_2)$*. In addition, assume that*

(i) For each i, $f_i^*(\theta_{-i})$ *is a function (rather than a correspondence), and is continuously differentiable in* θ_{-i}*.*

(ii) $\max_{\theta_1} |\frac{\partial f_2^*(\theta_1)}{\partial \theta_1}| \max_{\theta_2} |\frac{\partial f_1^*(\theta_2)}{\partial \theta_2}| < 1$.

Then players' beliefs converge to the steady state almost surely, regardless of the initial prior.

[27] Here we do not consider the Kullback-Leibuler divergence of hypothetical player i, because it coincides with that of actual player $j \neq i$.

Assumption (i) implies (and actually it is stronger than) asymptotic stability of the steady state p in the single-agent learning problem. Indeed, when p is unstable in the single-agent learning problem, $f_i(\theta^*_{-i})$ consists of multiple points as in Figure 3.

Assumption (ii) requires that each player's steady-state belief f_i is not too sensitive to the opponent's belief; this means that one's learning is not influenced by the the opponent's learning by much, at least asymptotically. Proposition 6 shows that the beliefs converge under these conditions. This convergence theorem applies to many economic applications.

The team production problem with the new output function (22) indeed satisfies these two assumptions: The curves in Figure 6 are the graphs of f_1 (the blue steeper curve) and f_2 (the orange flatter curve). One can easily see that the slope of the orange curve is less than one (i.e., $|\frac{\partial f_2^*(\theta_1)}{\partial \theta_1}| < 1$) and that of the blue curve is larger than one (i.e., $|\frac{\partial f_1^*(\theta_2)}{\partial \theta_2}| < 1$).

In the proof of Proposition 6, we extend Esponda, Pouzo, and Yamamoto (2021) and show that the asymptotic motion of the KL minimizer $\theta_i^t = \theta_i(\sigma^t)$ is approximated by a differential inclusion, which is somewhat similar to the differential equation (17). This result is weaker than what we have seen in the case of Gaussian noise (i.e., approximation by (17) and (18)), because given an initial value, our differential inclusion has multiple solutions, and our theorem does not tell us which solution approximates the actual evolution of θ_i^t. In other words, our differential inclusion is a "loose" approximation of the motion of the KL minimizer.

However, it turns out that this result is good enough to establish global convergence. As can be seen in the proof, if the assumptions stated in Proposition 6 hold, we can show that the steady state is globally attracting: *All* solutions to the differential inclusion converge to the steady state, regardless of the initial value. This immediately implies global convergence to the steady state, as claimed in Proposition 6.

4 Related Literature and Concluding Remarks

There is a rapidly growing literature on Bayesian learning with model misspecification. Nyarko (1991) presents a model in which the agent's action does not converge. Fudenberg, Romanyuk, and Strack (2017) consider a general two-state model and characterize the agent's asymptotic actions and behavior. Ba and Gindin (2023), He (2022), and Heidhues, Kőszegi, and Strack (2018, 2021) study a continuous-state setup, and they show that the agent's action and belief converge to a Berk-Nash equilibrium of Esponda and Pouzo (2016), under some assumptions on payoffs and information structure. Esponda, Pouzo, and Yamamoto (2021) characterize the agent's asymptotic behavior in a general single-agent model. Fudenberg, Lanzani, and Strack (2021) discuss stability of steady states. All these papers look at a single-agent problem or a multi-agent setup in which each player's bias (first-order misspecification) is common knowledge.

Higher-order misspecification has been studied in the literature on social learning (e.g., DeMarzo, Vayanos, and Zwiebel, 2003; Eyster and Rabin, 2010; Gagnon-Bartsch and Rabin, 2016; Bohren and Hauser, 2021). Most of these papers do not discuss discontinuity of the equilibrium outcome, and indeed, one of the main result of Bohren and Hauser (2021) is that the long-run outcome is robust to a small perturbation of the information structure. An exception is Frick, Iijima, and Ishii (2020), who show that the equilibrium outcome is discontinuous in the information structure in a model of information aggregation. As explained in Introduction, a key assumption is that the agents observe a noise signal about the state only once, which leads to discontinuity of the steady state. In contrast, in our model, the agents have repeated feedbacks about the state, and accordingly the steady states are continuous in the information structure. Nonetheless the equilibrium outcome is discontinuous, because a small misspecification influences the entire learning dynamics and the convergence probability suddenly drops to zero.

In this paper, we have focused on a particular form of higher-order misspecification, where both players are unaware of the opponent having a different view about the world. Of course, there are many other forms of higher-order misspecification which are prevalent in the real world. For example, in some markets, a fraction of consumers is overconfident; cellular phone customers tend to underestimate their usage next month, gym members tend to overestimate how often they will visit the gym, and so on.[28] In these cases, misspecification can be on *one side*, i.e., the consumers are misspecified while the companies are rational and understand the consumers' overconfidence. It turns out that this type of *one-sided double misspecification* also leads to the discontinuity of the equilibrium outcome as in our model, i.e., small overconfidence leads to a complete breakdown of correct learning.[29] This shows that our model is just an example in which the equilibrium outcome is sensitive to small misspecification. For future research, it may be interesting to study whether other forms of misspecification lead to this kind of sensitivity.

References

Attari, S.Z., M.L. DeKay, C.I. Davidson, and W.B. de Bruin (2010): "Public Perceptions of Energy Consumption and Savings" *Proceedings of the National Academy of Sciences*, 107 (37), 16054-16059.

Ba, C. and A. Gindin (2023): "A Multi-Agent Model of Misspecified Learning with Overconfidence," Working Paper.

Benaïm, M. (1999): "Dynamics of Stochastic Approximation Algorithms," Séminaire de Probabilités XXXIII, Lecture Notes in Math. 1709, Springer.

[28] See Grubb (2015) for a review of the literature.

[29] A rough idea is as follows. Suppose that player 2 has the same information structure as in our main model (i.e., she believes that the true parameter is $A_2 > a$, she believes that the opponent believes A_2, and so on), while player 1 is fully rational in that she correctly understands the parameter a and player 2's belief hierarchy. When $A_2 = a$, there is no misspecification, so that both players learn the true state θ^*. On the other hand, when $A_2 > a$, rational player 1 learns the true state but player 2 does not. In particular, the evolution of player 2's belief is asymptotically approximated by the belief evolution in the single-agent learning problem appearing in Section 3.4, where player 1's belief is fixed at θ^*. So if the interior steady state of this single-agent learning problem is unstable (Proposition 4 (i) gives the precise condition under which the steady state is indeed unstable), then player 2's belief converges there with zero probability under one-sided double misspecification.

Benaïm, M. and M. Faure (2012): “Stochastic Approximation, Cooperative Dynamics and Supermodular Games,” *Annals of Applied Probability*, 22, 2133-2164.

Benaïm, M., J. Hofbauer, and S. Sorin (2005): “Stochastic Approximations and Differential Inclusions,” *SIAM Journal on Control and Optimization*, 44, 328-348.

Billingsley, P. (1999): *Convergence of Probability Measures*, Wiley.

Bohren, J.A. and D.N. Hauser (2021): “Learning with Heterogeneous Misspecified Models: Characterization and Robustness,” *Econometrica*, 89 (6), 3025-3077.

Camerer, C. and D. Lovallo (1999): “Overconfidence and Excess Entry: An Experimental Approach,” *American Economic Review*, 89 (1), 306-318.

Chen, Y-C., A. Di Tillio, E. Faingold, and S. Xiong (2017): “Characterizing the Strategic Impact of Misspecified Beliefs,” *Review of Economic Studies*, 84, 1424-1471.

Daniel, K. and D. Hirshleifer (2015): “Overconfident Investors, Predictable Returns, and Excessive Trading,” *Journal of Economic Perspectives*, 29 (4), 61-88.

Dechezleprêtre, A., A. Fabre, T. Kruse, B. Planterose, A.S. Chico, and S. Stantcheva (2022): “Fighting Climate Change: International Attitudes Toward Climate Policies,” Working Paper.

Deimling, K. (1992): *Multivalued Differential Equations*, Walter de Gruyter.

DeMarzo, P.M., D. Vayanos, and J. Zwiebel (2015): “Persuasion Bias, Social Influence, and Unidimensional Opinions,” *Quarterly Journal of Economics*, 118 (3), 909-968.

Dudley, R. (1966): “Convergence of Baire Measures,” *Studia Mathematica*, 27, 251-268.

Esponda, I. and D. Pouzo (2016): “Berk-Nash Equilibrium: A Framework for Modeling Agents with Misspecified Models,” *Econometrica*, 84, 1093-1130.

Esponda, I., D. Pouzo, and Y. Yamamoto (2021): “Asymptotic Behavior of Bayesian Learners with Misspecified Models,” *Journal of Economic Theory*, 195, 105260.

Esponda, I., D. Pouzo, and Y. Yamamoto (2022): "Corrigendum to "Asymptotic Behavior of Bayesian Learners with Misspecified Models" [J. Econ. Theory 195 (2021) 105260]," *Journal of Economic Theory*, 204, 105513.

Eyster, E. (2019): "Errors in Strategic Reasoning," In D.B. Bernheim, S. DellaVigna & D. Laibson, (Eds.), *Handbook of Behavioral Economics: Foundations and Applications 2*, 187-259, North Holland.

Eyster, E. and M. Rabin (2010): "Naïve Herding in Rich-Information Settings," *American Economic Journal: Microeconomics*, 2 (4), 221-43.

Frick, M., R. Iijima, and Y. Ishii (2020): "Misinterpreting Others and the Fragility of Social Learning," *Econometrica*, 88, 2281-2328.

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

Fudenberg, D., G. Lanzani, and P. Strack (2021): "Limit Points of Endogenous Misspecified Learning," *Econometrica*, 89, 1065-1098.

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

Gagnon-Bartsch, T. and M. Rabin (2016): "Naive Social Learning, Mislearning, and Unlearning," Working Paper.

Grubb, M. (2015): "Overconfident Consumers in the Marketplace," *Journal of Economic Perspectives*, 29 (4), 9-36.

He, K. (2022): "Mislearning from Censored Data: The Gambler's Fallacy and Other Correlational Mistakes in Optimal-Stopping Problems," *Theoretical Economics*, 17, 1269-1312.

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

Heidhues, P., B. Kőszegi, and P. Strack (2021): "Convergence in Models of Misspecified Learning," *Theoretical Economics*, 16, 73-99.

Hofbauer, J., J. Oechssler, and F. Riedel (2009): "Brown-von Neumann-Nash Dynamics: the Continuous Strategy Case," *Games and Economic Behavior*, 65, 406-429.

Kushner, H.J. and G.G. Yin (2003): *Stochastic Approximation and Recursive Algorithms and Applications*, Springer.

Ludwig, S. and J. Nafziger (2011): " Beliefs about Overconfidence," *Theory and Decision*, 70, 475-500.

Malmendier, U. and G. Tate (2005): "CEO Overconfidence and Corporate Investment," *Journal of Finance*, 60 (6), 2661-2700.

Malmendier, U. and G. Tate (2008): "Who Makes Acquisitions? CEO Overconfidence and the Market's Reaction," *Journal of Financial Economics*, 89 (1), 20-43.

Malmendier, U. and G. Tate (2015): "Behavioral CEOs: The Role of Managerial Overconfidence," *Journal of Economic Perspectives*, 29 (4), 37-60.

McGee, D. (2023): "Stereotypes and Strategic Discrimination," Working Paper.

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

Pemantle, R. (1990): "Nonconvergence to Unstable Points in Urn Models and Stochastic Approximations," *Annals of Probability*, 18, 698-712.

Perkins, S. and D.S. Leslie (2014): "Stochastic Fictitious Play with Continuous Action Sets," *Journal of Economic Theory*, 152, 179-213.

Van Boven, L., D. Dunning, and G. Loewenstein (2000): "Egocentric empathy gaps between owners and buyers: Misperceptions of the endowment effect," *Journal of Personality and Social Psychology*, 79 (1), 66-76.

Varian, H.R. (1992): *Microeconomic Analysis* (3rd ed.), Norton.

Appendices

A Asymptotic Behavior of Misspecified Players

In this appendix, we will characterize the asymptotic behavior of players' beliefs, which is useful to establish the convergence theorems presented in the main text. It takes two steps. As a first step, we will develop a general model which encompasses both first-order misspecification and double misspecification as special cases, and show that the motion of players' *action frequency* is asymptotically approximated by a solution to a differential inclusion. This result can be seen as a generalization of the main theorem of Esponda, Pouzo, and Yamamoto (2021) to the case with multiple players and continuous actions. When actions are continuous, the action frequency becomes an infinite-dimensional vector, so we have to carefully choose a norm for the set of action frequencies; indeed, the meaning of "approximation" can be very different for different choices of the norm. We find that the result similar to Esponda, Pouzo, and Yamamoto (2021) holds if we use the *dual bounded-Lipschitz norm*.

This result is useful because it characterizes players' asymptotic behavior using a *deterministic* dynamic process (a differential inclusion) which does not involve a stochastic component. However, in our continuous-action setup, this differential inclusion becomes an infinite-dimensional problem, and in practice, it is impossible to solve such a differential inclusion, So as a second step, we show that under some technical assumption, the asymptotic motion of players' *beliefs* can be approximated by a more tractable, *finite-dimensional* differential inclusion. This result is new to the literature, and as will be seen, it plays a key role in the proof of our convergence theorems (Propositions 1 and 6).

A.1 General Setup

For each compact set $A \subset \boldsymbol{R}^n$ (or more generally, separable metric space A), let $\triangle A$ denote the set of probability measures over the set A. We consider the *dual bounded-Lipschitz norm* on $\triangle A$, that is, for each $\mu \in \triangle A$, let

$$\|\mu\| = \sup_{f \in BL(A)} \int_A f d\mu$$

where $BL(A)$ is the set of bounded Lipschitz continuous functions f on A with $\sup_{x\in A} |f(x)| + \sup_{x\neq y} \frac{|f(x)-f(y)|}{|x-y|} \leq 1$. This norm has two nice properties. First, it metrizes the weak topology, that is, the topology induced by the dual bounded-Lipschitz norm coincides with the weak topology on $\triangle A$. Second, with this norm, $\triangle A$ is a compact subset of a Banach space, i.e., the set of finite signed measures on A is a Banach space when paired with the dual bounded-Lipschitz norm, and $\triangle A$ is a compact subset in it. See Dudley (1966) and Billingsley (1999) for references. The first property is needed to obtain our Proposition 9. The second property is crucial in order to use a stochastic approximation technique in the proof of Proposition 10. The dual bounded-Lipschitz norm is used in Hofbauer, Oechssler, and Riedel (2009) and Perkins and Leslie (2014), who study learning dynamics in games with continuous actions.

A.1.1 Objective World

There are two players $i = 1,2$ and infinitely many periods $t = 1,2,\cdots$. In each period t, each player i chooses an action x_i from a compact set $X_i \subset \boldsymbol{R}$. These actions are not observable. Then they observe a noisy public output $y \in Y$ which is distributed according to a probability measure $Q(\cdot|x) \in \triangle Y$, where $x = (x_1,x_2)$ denotes the chosen action profile. Each player i's payoff is $u_i(x_i,y)$.

In the infinite-horizon game, each player i's t-period history is $h_i^t = (x_i^\tau, y^\tau)_{\tau=1}^t$, where (x_i^t, y^t) is player i's action and the public outcome in period t. Let H_i^t denote the set of all t-period history, and let $H_i^0 = \{\emptyset\}$. Player i's *pure strategy* in the infinite-horizon game is a mapping $s_i : \bigcup_{t=0}^\infty H_i^t \to X_i$. Let S_i denote the set of player i's pure strategies. Let $h_Y^t = (y^\tau)_{\tau=1}^t$ denote the t-period public history. A strategy is *public* if it depends only on public histories.

A.1.2 Subjective World and Model Hierarchy

We assume that the output distribution Q is not common knowledge among players. Instead, each player i has a set $\Theta_{i,1}$ of subjective models, and in each model $\theta_{i,1} \in \Theta_{i,1}$, the output distribution given an action profile x is $Q_{\theta_{i,1}}(\cdot|x)$. Player i thinks that the true world is described by one of these models, and her initial prior about the model is $\mu_{i,1} \in \triangle\Theta_{i,1}$. Player i's models are *correctly specified* if there is $\theta_{i,1}$ such that $Q(\cdot|x) = Q_{\theta_{i,1}}(\cdot|x)$ for all x. Otherwise her models are *misspecified*. Player i also has models about the opponent j's model, that is, player i believes that the opponent j has an initial prior $\mu_{i,2}$ over a model set $\Theta_{i,2}$, where each model $\theta_{i,2}$ induces the output distribution $Q_{\theta_{i,2}}(\cdot|x)$ for each action profile x. This triplet $M_{i,2} = (\mu_{i,2}, \Theta_{i,2}, (Q_{\theta_{i,2}}(\cdot|x))_{(x,\theta_{i,2})})$ is player i's *second-order model* in that it is her subjective view about player j's subjective model. More generally, we assume that each player i has a *model hierarchy* $M_i = (M_{i,1}, M_{i,2}, \cdots)$ where each

$M_{i,k} = (\mu_{i,k}, \Theta_{i,k}, (Q_{\theta_{i,k}}(\cdot|x))_{(x,\theta_{i,k})})$ is player i's *kth-order model*. That is, player i believes that player j believes that player i's model is $M_{i,3}$, player i believes that player j believes that player i believes that player j's model is $M_{i,4}$, and so on.

This framework is flexible and allows us to study a variety of information structures. For example, we obtain the model of first-order misspecification studied in Section 2 when $M_{1,1} = M_{2,2} = M_{1,3} = M_{2,4} = M_{1,5} = \cdots$, $M_{2,1} = M_{1,2} = M_{2,3} = M_{1,4} = M_{2,5} = \cdots$, and $M_{1,1}$ is correctly specified; here the first condition implies that player 1's model $M_{1,1}$ is common knowledge, and the second condition implies that player 2's model $M_{2,1}$ is common knowledge. Similarly, we obtain the model of double misspecification studied in Section 3 when $M_{i,1} = M_{i,2} = M_{i,3} = \cdots$ for each i.

In what follows, we will maintain the following technical assumptions.

Assumption 1. The following conditions hold:

(i) Y and Θ are Borel subsets of the Euclidean space, and Θ is compact.

(ii) There is a Borel probability measure $\nu \in \triangle Y$ such that $Q(\cdot|x)$ and $Q_{\theta_{i,k}}(\cdot|x)$ are absolutely continuous with respect to ν for all x and i, k, and $\theta_{i,k}$. (An implication is that there are densities $q(\cdot|x)$ and $q_{\theta_{i,k}}(\cdot|x)$ such that $\int_A q(y|x)\nu(dy) = Q(A|x)$ and $\int_A q_{\theta_{i,k}}(y|x)\nu(dy) = Q_{\theta_{i,k}}(A|x)$ for any $A \subseteq Y$ Borel.)

(iii) $q(\cdot|x)$ and $q_{\theta_{i,k}}(\cdot|x)$ are continuous in θ and x.

(iv) There is a function $g : X \times Y \to \boldsymbol{R}$ such that (a) for each y, $g(x,y)$ is continuous in x, (b) $g(x,\cdot) \in L^2(Y, Q(\cdot|x))$ for each x, and (c) for all x, $\hat{x}$ i, k, and $\theta_{i,k}$, $\log \frac{q(\cdot|x)}{q_{\theta_{i,k}}(\cdot|\hat{x})} \leq g(x,\cdot)$ $Q(\cdot|x)$-a.s. .

The parts (i)-(iii) are fairly standard. The part (iv) implies that every outcome y is generated by each player i's model, which is useful to establish a uniform version of the law of large numbers. The assumption above is similar to Assumptions 1 and 2 of Esponda, Pouzo, and Yamamoto (2021), but there are two differences. First, we allow the action set X_i to be continuous, in which case we require continuity of q, as described in parts (iii) and (iv-a). Second, we allow inferential naivety, so when we consider the log-likelihood $\log \frac{q(\cdot|x)}{q_{\theta_{i,k}}(\cdot|\hat{x})}$ of the true output probability and the subjective probability, we distinguish the actual action profile x from the inferred action profile $\hat{x}$.

Recall that in the cases of first-order misspecification and double misspecification, each player i believes that (i) her view $M_{i,1}$ about the world is common knowledge (i.e., $M_{i,1} = M_{i,3} = M_{i,5} = \cdots$) and that (ii) her view $M_{i,2}$ about the opponent's view about the world is common knowledge (i.e., $M_{i,2} = M_{i,4} = M_{i,6} = \cdots$). This ensures that player i's decision making problem is equivalent to solving a game played by this player i and a hypothetical player.[30] In the general model here, we will impose a (similar but) weaker assumption:

Assumption 2. Player i believes that the models (M_{i,k_i}, M_{i,k_i+1}) are common knowledge after level $k_i < \infty$, that is, for each i, there is $k_i < \infty$ such that $(M_{i,k_i}, M_{i,k_i+1}) = (M_{i,k_i+2n}, M_{i,k_i+1+2n})$ for each $n = 1, 2, \cdots$.

For the special case in which $k_i = 1$, this assumption implies that player i believes that the models $(M_{i,1}, M_{i,2})$ are common knowledge, just as in the case of first-order misspecification and double misspecification. The assumption above is more general than that, because it allows $k_i > 1$; in such a case, the assumption implies that player i believes that models are common knowledge at higher levels, i.e., she believes that the opponent believes that $\cdots$ that the models (M_{i,k_i}, M_{i,k_i+1}) are common knowledge. Note that this assumption is about whether *player i thinks that* the models are common knowledge, and *not* about whether the models are common knowledge in the objective sense. We believe that Assumption 2 is satisfied in most applications.[31]

Pick k_i as stated in Assumption 2. Then player i's problem is strategically equivalent to solving the following hypothetical game with $k_i + 1$ agents:

- Each period, each agent $k = 1, 2, \cdots, k_i + 1$ chooses an action $\hat{x}_{i,k}$ from a set $\hat{X}_{i,k}$, where $\hat{X}_{i,k} = X_i$ for odd k, and $\hat{X}_{i,k} = X_j$ for even k.

- Agent 1 is player i herself. She has the model $M_{i,1}$, and thinks that her opponent is agent 2. That is, she thinks that the distribution of the public outcome is $Q_{\theta_{i,1}}(\hat{x}_{i,1}, \hat{x}_{i,2})$ for some $\theta_{i,1}$, where $(\hat{x}_{i,1}, \hat{x}_{i,2})$ is the action chosen by agents 1 and 2.

- Other agents are hypothetical players appearing in player i's reasoning. Each agent $k = 2, 3, \cdots, k_i + 1$ has the model $M_{i,k}$, and thinks that her opponent is agent $k+1$. That is, she thinks that the distribution of the public

[30] In the case of first-order misspecification, this hypothetical player is redundant in that her action coincides with the actual player's action. So such a hypothetical player does not appear in our analysis in Section 2.

[31] This assumption is needed to establish Propositions 9 and 10. Indeed, if this assumption is not satisfied, then we need infinite agents to describe player i's reasoning, so the set $\hat{X}$ becomes the product of infinitely many X_1 and X_2. This set $\hat{X}$ is not separable (it is well-known that the l^∞-space is not separable), so the dual bounded-Lipschitz norm on $\triangle\hat{X}$ may not coincide with the topology of weak convergence.

outcome is $Q_{\theta_{i,k}}(\hat{x}_{i,k}, \hat{x}_{i,k+1})$ for some $\theta_{i,k}$. Here, agent $k_i + 2$ refers to agent k_i, so agents k_i and $k_i + 1$ play the game with each other.

- All the information structure above is common knowledge among the agents.

Intuitively, agent 1's action $\hat{x}_{i,1}$ in this hypothetical game is player i's actual action, agent 2's action $\hat{x}_{i,2}$ is player i's prediction about the opponent j's action, agent 3's action $\hat{x}_{i,3}$ is player i's prediction about j's prediction about i's action, and so on. So the action profile $\hat{x}_i = (\hat{x}_{i,k})_{k=1}^{k_i+1}$ in this hypothetical game is essentially player i's *prediction hierarchy*. Let $\hat{X}_i = \times_{k=1}^{k_i+1} X_{i,k}$ denote the set of all these action profiles.

In what follows, each agent k in this hypothetical game is labelled as (i,k), because these agents describe player i's reasoning. The opponent j has a different model hierarchy $M_j \neq M_i$, and hence her reasoning is represented by a different set of agents labelled as (j,k).

Let $\hat{s}_{i,k}$ denote agent (i,k)'s strategy in the infinite-horizon hypothetical game, and let $\hat{s}_i = (\hat{s}_{i,k})_{k=1}^{k_i+1}$ denote a strategy profile. This profile $\hat{s}_i$ is also interpreted as player i's *prediction hierarchy* about strategies in the infinite-horizon game. That is, $\hat{s}_{i,1}$ is player i's actual strategy, $\hat{s}_{i,2}$ is player i's prediction about player j's strategy, and so on. So $\hat{s}_{i,k} \in S_i$ for odd k, and $\hat{s}_{i,k} \in S_j$ for even k. We assume that each $\hat{s}_{i,k}$ is pure and public.

Given a pure strategy profile $\hat{s}_i = (\hat{s}_{i,k})$ in the hypothetical game, each agent k's posterior belief $\hat{\mu}_{i,k}^{t+1} \in \triangle\Theta_{i,k}$ can be computed using Bayes' rule, after every public history h_Y^t. Formally, for each t and k, we have

$$\hat{\mu}_{i,k}^{t+1}(\theta_{i,k}) = \frac{\hat{\mu}_{i,k}^{t}(\theta_{i,k}) Q_{\theta_{i,k}}(y^t | \hat{s}_{i,k}(h_Y^{t-1}), \hat{s}_{i,k+1}(h_Y^{t-1}))}{\int_{\Theta_{i,k}} \hat{\mu}_{i,k}^{t}(\theta_{i,k}) Q_{\theta_{i,k}}(y^t | \hat{s}_{i,k}(h_Y^{t-1}), \hat{s}_{i,k+1}(h_Y^{t-1})) d\theta_{i,k}}$$

where $\hat{s}_{i,k_i+2} = \hat{s}_{i,k_i}$. Here we use the fact that agent k thinks that the signal y^t in period t is drawn given the action profile $(\hat{s}_{i,k}(h_Y^{t-1}), \hat{s}_{i,k+1}(h_Y^{t-1}))$, where $\hat{s}_{i,k}(h_Y^{t-1})$ is her own action, and $\hat{s}_{i,k+1}(h_Y^{t-1})$ is the opponent $k+1$'s action. The above formula is valid only if no one deviates from the profile $\hat{s}_i$; if some agent k deviates, then her posterior belief must be computed using a different formula. A strategy profile $\hat{s}_i$ is *Markov* if each agent's strategy depends only on the belief hierarchy $\hat{\mu}_i^t$, i.e., for each k and t, $\hat{s}_{i,k}(h_Y^t)$ depends on h_Y^t only through $\hat{\mu}_i^{t+1}$.

Example 1. (Myopically optimal agents) Suppose that the agents are myopic and maximize their expected stage-game payoffs each period. In such a case, they play a one-shot equilibrium given a belief-hierarchy $\hat{\mu}^t$ in each period t. Recall that each agent (i,k) thinks that her opponent is agent $(i,k+1)$, so her subjective expected stage-game payoff given a model $\theta_{i,k}$ is

$$U_{\theta_{i,k}}(\hat{x}_{i,k},\hat{x}_{i,k+1}) = \int_Y u_{i,k}(\hat{x}_{i,k},y)Q_{\theta_{i,k}}(dy|\hat{x}_{i,k},\hat{x}_{i,k+1})$$

where $u_{i,k}=u_1$ when $i+k$ is even, and $u_{i,k}=u_2$ when $i+k$ is odd. So the strategy profile $\hat{s}_i$ must satisfy the following equilibrium condition:

$$\hat{s}_{i,k}(\hat{\mu}_i) \in \arg\max_{\hat{x}_{i,k}\in\hat{X}_{i,k}} \int_{\Theta_{i,k}} U_{\theta_{i,k}}(\hat{x}_{i,k},\hat{s}_{i,k+1}(\hat{\mu}_i))\hat{\mu}_{i,k}(d\theta_{i,k}) \quad \forall k\forall\hat{\mu}_i. \tag{23}$$

It is obvious that this strategy profile $\hat{s}_i$ is Markov.

Example 2. (Dynamically optimal agents) Now consider dynamically optimal agents, who maximize the expectation of the discounted sum of the stage-game payoffs, $\sum_{t=1}^{\infty}\delta^{t-1}u_{i,k}(\hat{x}_{i,k},y)$. Many applied papers use Markov perfect equilibria as a solution concept. In our context, $\hat{s}_i$ is a Markov perfect equilibrium if given any belief hierarchy $\hat{\mu}_i$, the continuation strategy profile $\hat{s}_i|_{\hat{\mu}_i}$ satisfies

$$\hat{s}_{i,k}|_{\hat{\mu}_i} \in \arg\max_{\hat{s}_{i,k}} \int_{\Theta_{i,k}} \sum_{t=1}^{\infty} \delta^{t-1}E[U_{\theta_{i,k}}(\hat{x}^t_{i,k},x^t_{i,k+1})|\hat{s}_{i,k},\hat{s}_{i,k+1}|_{\hat{\mu}_i}]\hat{\mu}_{i,k}(d\theta_{i,k})$$

for each k, where the expectation is taken over $(\hat{x}^t_{i,k},x^t_{i,k+1})$.

Let $h=(x^t,y^t)_{t=1}^{\infty}$ denote a sample path (a history in the infinite-horizon game). Also, let $\hat{X}=\hat{X}_1\times\hat{X}_2$ be the product of the sets of all action profiles of the two hypothetical games. Given a sample path h and given strategy profiles $\hat{s}=(\hat{s}_1,\hat{s}_2)$ of the two hypothetical games (for players 1 and 2), let $\sigma^t(h)\in\triangle\hat{X}$ denote the action frequency up to period t, that is,

$$\sigma^t(h)[(\hat{x}_1,\hat{x}_2)] = \frac{1}{t}\sum_{\tau=1}^{t} 1_{\{\hat{s}_{i,k}(h_Y^{\tau-1})=\hat{x}_{i,k}\ \ \forall i\forall k\}}.$$

Intuitively, $\sigma^t(h)[(\hat{x}_1,\hat{x}_2)]$ describes how often the action profile $\hat{x}_i$ was chosen in each hypothetical game. (In other words, it describes how often each player i made a prediction hierarchy $\hat{x}_i$.) Note that we cannot directly observe the actions $\hat{x}_{i,k}$ of the higher-level agents (i,k) with $k\geq 2$, as they are hypothetical agents. However, since each agent uses a public strategy $\hat{s}_{i,k}$, we can back it up from the past public history; given a history $h_Y^{\tau-1}$, the hypothetical agent k's action in period τ must be $\hat{s}_{i,k}(h_Y^{\tau-1})$. This allows us to define the action frequency in the hypothetical game as a function of the observed history h.

A.2 Posterior Beliefs and Kullback-Leibler Divergence

We first show that after a long time t, the posterior belief is concentrated on the models which best explain the data. Specifically, we show that the belief is concentrated on the models which minimize the Kullback-Leibler divergence, which is defined as follows. Let $\sigma \in \triangle \hat{X}$ be a probability measure over $\hat{X}$. For each σ, the *Kullback-Leibler divergence* of model $\theta_{i,k}$ for agent k is defined as

$$K_{i,k}(\theta_{i,k}, \sigma) = \int_{\hat{X}} \int_{Y} \log \frac{q(y|\hat{x}_{1,1}, \hat{x}_{2,1})}{q_{\theta_{i,k}}(y|\hat{x}_{i,k}, \hat{x}_{i,k+1})} Q(dy|\hat{x}_{1,1}, \hat{x}_{2,1}) \sigma(d\hat{x}). \tag{24}$$

Intuitively, $K_{i,k}(\theta_{i,k}, \sigma)$ measures the distance between the true output distribution and the subjective distribution induced by agent k's model $\theta_{i,k}$. To see this, think about the special case in which σ is a degenerate distribution $1_{\hat{x}_1, \hat{x}_2}$. Then the Kullback-Leibler divergence of model $\theta_{i,k}$ can be rewritten as

$$\int_{Y} \log \frac{q(y|\hat{x}_{1,1}, \hat{x}_{2,1})}{q_{\theta_{i,k}}(y|\hat{x}_{i,k}, \hat{x}_{i,k+1})} Q(dy|\hat{x}_{1,1}, \hat{x}_{2,1}).$$

This measures the distance between the true distribution $q(\cdot|\hat{x}_{1,1}, \hat{x}_{2,1})$ and the subjective distribution $q_{\theta_{i,k}}(\cdot|\hat{x}_{i,k}, \hat{x}_{i,k+1})$ induced by the model $\theta_{i,k}$. Indeed, this value is always non-negative, and equals zero if and only if the true and subjective distributions are the same. When σ is not a degenerate distribution, we take a weighted sum of the Kullback-Leibler divergence over $\hat{x} = (\hat{x}_1, \hat{x}_2)$, which leads to the definition of $K_{i,k}(\theta_{i,k}, \sigma)$ above.

As is clear from this formula, agent k's subjective signal distribution $q_{\theta_{i,k}}(y|\hat{x}_{i,k}, \hat{x}_{i,k+1})$ is potentially different from the true distribution $q(y|\hat{x}_{1,1}, \hat{x}_{2,1})$ in two ways. First, agent k's model $\theta_{i,k}$ can be *misspecified* in that the distribution $q_{\theta_{i,k}}$ as a function of the chosen action can be different from the true distribution q. Second, agent k can have an *inferential naivety*. That is, while the true distribution is determined by the actual actions chosen by players 1 and 2 (which is denoted by $(\hat{x}_{1,1}, \hat{x}_{2,1})$ in our setup), agent k thinks that the output distribution is determined by the actions chosen by agents k and $k+1$.

For each measure $\sigma \in \triangle \hat{X}$, let $\Theta_{i,k}(\sigma)$ denote the minimizers of the Kullback-Leibler divergence, that is,

$$\Theta_{i,k}(\sigma) = \arg \min_{\theta_{i,k} \in \Theta_{i,k}} K_{i,k}(\theta_{i,k}, \sigma).$$

Intuitively, this is the set of models which best explains the data when the past action frequency was σ. The minimized Kullback-Leibler divergence is $K^*_{i,k}(\sigma) =$

$\min_{\theta_{i,k}\in\Theta_{i,k}} K_{i,k}(\theta_{i,k},\sigma)$. We first show that these minimizers have useful properties:

Lemma 1. *For each i and k, (i) $K_{i,k}(\theta_{i,k},\sigma)-K^*_{i,k}(\sigma)$ is continuous in $(\theta_{i,k},\sigma)$, and (ii) $\Theta_{i,k}(\sigma)$ is upper hemi-continuous, non-empty, and compact-valued.*

The following proposition shows that after a long time t, the posterior is concentrated on the best models $\Theta_{i,k}(\sigma^t)$. This extends Theorem 1 of Esponda, Pouzo, and Yamamoto (2021) to the case with continuous action set X_i and with multiple players. Let H denote the set of all sample paths $h=(x^t,y^t)_{t=1}^{\infty}$. Given strategy profiles $\hat{s}$, let $P^{\hat{s}}\in\triangle X$ denote the probability distribution of the sample path h. Given a sample path h, let $\hat{\mu}_i^t(h)$ denote the belief hierarchy in period t.

Proposition 7. *Given any i, k, and $\hat{s}$, $P^{\hat{s}}$-almost surely, we have*

$$\lim_{t\to\infty}\int_{\Theta_{i,k}}(K_{i,k}(\theta_{i,k},\sigma^t(h))-K^*_{i,k}(\sigma^t(h)))\hat{\mu}^{t+1}_{i,k}(h)[d\theta_{i,k}]=0. \tag{25}$$

Let $\mathscr{H}$ denote the set of sample paths h which satisfy (25). By Proposition 7, $P^{\hat{s}}(\mathscr{H})=1$.

A.3 Asymptotic Motion of Action Frequency

A.3.1 Stochastic Approximation and Differential Inclusion

Now we will show that given any Markov strategy $\hat{s}$, the asymptotic motion of the action frequency σ^t is approximated by a solution to a differential inclusion. Pick a Markov strategy $\hat{s}$, and pick a sample path $h\in\mathscr{H}$. By the definition, the action frequency in each period is written as

$$\sigma^{t+1}(h)=\frac{t}{t+1}\sigma^t(h)+\frac{1}{t+1}1_{\hat{s}(\hat{\mu}^{t+1}(h))}.$$

That is, the action frequency in period $t+1$ is a weighted average of the past action frequency σ^t and today's action $1_{\hat{s}(\hat{\mu}^{t+1}(h))}$. In what follows, we will show that this second term $1_{\hat{s}(\hat{\mu}^{t+1}(h))}$ can be written as a function of σ^t, so that σ^{t+1} is determined recursively.

Pick an arbitrary small $\varepsilon>0$. Then let $B_\varepsilon:\triangle\hat{X}\to\prod_{i=1}^{2}\prod_{k=1}^{k_i+1}\triangle\Theta_{i,k}$ be the *ε-perturbed belief correspondence* defined as

$$B_\varepsilon(\sigma)=\left\{\hat{\mu}\,\middle|\,\forall i\forall k\int_{\Theta_{i,k}}(K_{i,k}(\theta_{i,k},\sigma)-K^*_{i,k}(\sigma))\hat{\mu}_{i,k}(d\theta_{i,k})\le\varepsilon\right\}.$$

Roughly, $B_\varepsilon(\sigma)$ is the set of all belief hierarchies $\hat{\mu}$ such that each $\hat{\mu}_{i,k}$ is concentrated on the best models $\Theta_{i,k}(\sigma)$ in the sense of (25), given the mixture σ.

Since $h \in \mathscr{H}$, there is T such that for all $t > T$, $\hat{\mu}^{t+1}(h) \in B_\varepsilon(\sigma^t)$. This in turn implies that the action $\hat{s}(\hat{\mu}^{t+1})$ in period $t+1$ must be chosen from the ε-*enlarged policy correspondence* $S_\varepsilon(\sigma^t)$, which is defined as

$$S_\varepsilon(\sigma) = \{\hat{s}(\hat{\mu}) | \forall \hat{\mu} \in B_\varepsilon(\sigma)\}$$

for each σ. This immediately implies the following result:

Proposition 8. *Pick a Markov strategy $\hat{s}$. Then given any $h \in \mathscr{H}$, there is a decreasing sequence $\{\varepsilon^t\}_{t=1}^\infty$ with $\lim_{t\to\infty} \varepsilon^t = 0$ such that*

$$\sigma^{t+1}(h) \in \frac{t}{t+1}\sigma^t(h) + \frac{1}{t+1} S_{\varepsilon^t}(\sigma^t(h)).$$

This proposition implies that in a later period t, the action chosen in that period is selected from the set $S_\varepsilon(\sigma^t)$ for small ε. Now we ask how this set looks like in the limit as $\varepsilon \to 0$. Given a Markov strategy $\hat{s}$, let

$$\hat{S}(\mu) = \left\{ \hat{x} \middle| \hat{x} = \lim_{n\to\infty} \hat{s}(\hat{\mu}^n) \text{ for some } (\hat{\mu}^n)_{n=1}^\infty \text{ with } \lim_{n\to\infty} (\hat{\mu}^n) = \hat{\mu} \right\}$$

for each μ. This $\hat{S}$ is an *upper hemi-continuous policy correspondence induced by $\hat{s}$.* It is obvious that $\hat{s}(\hat{\mu}) \in \hat{S}(\hat{\mu})$ for each $\hat{\mu}$. Also a standard argument shows that $\hat{S}$ is indeed upper hemi-continuous with respect to $\hat{\mu}$. Note that $\hat{S} = \hat{s}$ if $\hat{s}$ is continuous. Then define

$$S_0(\sigma) = \{\hat{x} \in \hat{S}(\hat{\mu}) | \forall \hat{\mu} \in B_0(\sigma)\}$$

where

$$B_0(\sigma) = \{\hat{\mu} | \hat{\mu}_{i,k} \in \triangle\Theta_{i,k}(\sigma) \ \ \forall i \forall k\}.$$

The following proposition shows that when $\varepsilon \to 0$, the set $S_\varepsilon(\sigma)$ which appears in the previous proposition is approximated by $S_0(\sigma)$.

Proposition 9. *$S_\varepsilon(\sigma)$ is upper hemi-continuous in (ε, σ) at $\varepsilon = 0$. So with the dual bounded-Lipschitz norm, $\triangle S_\varepsilon(\sigma)$ is upper hemi-continuous at $\varepsilon = 0$.*

Propositions 8 and 9 suggest that after a long time, the motion of the action frequency is approximated by

$$\sigma^{t+1}(h) \in \frac{t}{t+1}\sigma^t(h) + \frac{1}{t+1} S_0(\sigma^t(h)),$$

which is equivalent to

$$\sigma^{t+1}(h) - \sigma^t(h) \in \frac{t}{t+1}(S_0(\sigma^t(h)) - \sigma^t(h))$$

That is, the drift of the action frequency, $\sigma^{t+1}(h) - \sigma^t(h)$, should be proportional to the difference between today's action chosen from $S_0(\sigma^t(h))$ and the current action frequency $\sigma^t(h)$. The next proposition formalizes this idea using the stochastic approximation technique developed by Benaïm, Hofbauer, and Sorin (2005): It shows that the asymptotic motion of the action frequency is described by the differential inclusion

$$\dot{\boldsymbol{\sigma}}(t) \in \triangle S_0(\boldsymbol{\sigma}(t)) - \boldsymbol{\sigma}(t). \tag{26}$$

In this differential inclusion, the drift of the action frequency is $\triangle S_0(\boldsymbol{\sigma}(t)) - \boldsymbol{\sigma}(t)$, rather than $S_0(\boldsymbol{\sigma}(t)) - \boldsymbol{\sigma}(t)$. The reason is as follows. As will be shown in Proposition 10 below, the differential inclusion (26) approximates the motion of the action frequency in the limit as the period length in the discrete-time model shrinks to zero. This means that a small time interval $[t, t+\varepsilon]$ in the continuous-time model should be interpreted as a collection of arbitrarily many periods in the discrete-time model. Suppose now that players' beliefs are in a neighborhood of μ during this time interval $[t, t+\varepsilon]$. In all periods included in this interval, players choose an action profile from the set $S_0(\mu)$, and in particular, if $S_0(\mu)$ contains two or more action profiles, then different action profiles can be chosen in different periods. Accordingly, the action frequency during this interval can take any value in $\triangle S_0(\mu)$, as described by the differential inclusion (26).[32]

To state the result formally, we use the following terminologies, which are standard in the literature on stochastic approximation. Let $\tau_0 = 0$ and $\tau_t = \sum_{n=1}^{t} \frac{1}{n}$ for each $t = 1, 2, \cdots$. Then given a sample path h, the *continuous-time interpolation* of the action frequency σ^t is a mapping $\boldsymbol{w}(h) : [0, \infty) \to \triangle \hat{X}$ such that

$$\boldsymbol{w}(h)[\tau_t + s] = \sigma^t(h) + \frac{\tau}{\tau_{t+1} - \tau_t}(\sigma^{t+1}(h) - \sigma^t(h))$$

for all $t = 0, 1, \cdots$ and $\tau \in [0, \frac{1}{t+1})$. Intuitively, $\boldsymbol{w}$ represents the motion of the action frequency as a piecewise linear path with re-indexed time. A mapping

[32]There is also a technical reason: In the proof of Proposition 10, we apply the stochastic approximation method of Benaïm, Hofbauer, and Sorin (2005), which requires that the drift term be a convex-valued (and upper hemi-continuous) correspondence. So we need to convexify the drift term by taking $\triangle S_0(\boldsymbol{\sigma}(t))$, rather than $S_0(\boldsymbol{\sigma}(t))$.

$\boldsymbol{\sigma} : [0,\infty) \to \triangle\hat{X}$ is a *solution to the differential inclusion (26) with an initial value* $\sigma \in \triangle\hat{X}$ if it is absolutely continuous in all compact intervals, $\boldsymbol{\sigma}(0) = \sigma$, and (26) is satisfied for almost all t. Since $\triangle S_0(\sigma)$ is upper hemi-continuous with closed convex values, given any initial value $\sigma \in \triangle\hat{X}$, the differential inclusion (26) has a solution. (See Theorem 9 of Deimling (1992) on page 117.) Let $Z(\sigma)$ denote the set of all these solutions given an initial value σ, and let $Z = \bigcup_\sigma Z(\sigma)$ denote the set of all solutions.

Proposition 10. *Pick a Markov strategy $\hat{s}$. Then for any $T > 0$ and any sample path $h \in \mathscr{H}$,*

$$\lim_{t\to\infty} \inf_{\boldsymbol{\sigma}\in Z} \sup_{\tau\in[0,T]} \|\boldsymbol{w}(h)[t+\tau] - \boldsymbol{\sigma}(\tau)\| = 0.$$

When $S_0(\sigma)$ is a singleton for all σ (which means that the differential inclusion is actually a differential equation), this reduces to

$$\lim_{t\to\infty} \inf_{\boldsymbol{\sigma}\in Z(\boldsymbol{w}(t))} \sup_{\tau\in[0,T]} \|\boldsymbol{w}(h)[t+\tau] - \boldsymbol{\sigma}(\tau)\| = 0.$$

A.3.2 Steady State and Generalized Berk-Nash Equilibrium

$\sigma \in \triangle\hat{X}$ is a *steady state* of the differential inclusion (26) if $\sigma \in \triangle S_0(\sigma)$. The following proposition shows that if the action frequency σ^t converges, then its limit point must be a steady state. The proof is exactly the same as Proposition 1 of EPY, and hence we omit it.

Proposition 11. *Pick a Markov strategy s. Then for each sample path $h \in \mathscr{H}$, if the action frequency $\sigma^t(h)$ converges, then its limit point $\lim_{t\to\infty}\sigma^t(h)$ is a steady state of (26).*

In all the examples in this paper, we assume that the agents are myopic so that the strategy profile $\hat{s}$ satisfies (23). In this special case, steady states of our differential inclusion are *generalized Berk-Nash equilibria* in the following sense:

Definition 2. A probability measure $\sigma \in \triangle\hat{X}$ is a *generalized Berk-Nash equilibrium* (*GBNE*) if for each pure action profile $\hat{x} = (\hat{x}_1, \hat{x}_2)$ in the support of σ, for each i and for each k, there is a belief $\hat{\mu}_{i,k} \in \triangle\Theta_{i,k}(\sigma)$ such that

$$\hat{x}_{i,k} \in \arg\max_{\hat{x}'_{i,k}} \int_{\Theta_{i,k}} U_{\theta_{i,k}}(\hat{x}'_{i,k}, \hat{x}_{i,k+1})\hat{\mu}_{i,k}(d\theta_{i,k}).$$

A generalized Berk-Nash equilibrium is *degenerate* if it is a point mass on some pure action profile $\hat{x}$.

In words, in a generalized Berk-Nash equilibrium σ, each action profile $\hat{x}$ which has a positive weight in σ is a one-shot equilibrium for some belief $\hat{\mu}$, and this belief $\hat{\mu}$ is concentrated on the models $\Theta_{i,k}(\sigma)$ which minimize the Kullback-Leibular divergence. In a non-degenerate GBNE which assign positive weights on multiple action profiles $\hat{x}$, different action profiles $\hat{x}$ may be supported by different beliefs $\hat{\mu}$.

Proposition 12. *Suppose that the strategy profile $\hat{s}$ satisfies (23). Then any steady state of our differential inclusion (26) is a generalized Berk-Nash equilibrium. So for each sample path $h \in \mathscr{H}$, if the action frequency $\sigma^t(h)$ converges, then its limit point $\lim_{t\to\infty} \sigma^t(h)$ is a generalized Berk-Nash equilibrium.*

Note that the action frequency may converge to non-degenerate equilibrium σ, which assigns positive probability to multiple action profiles $\hat{x}$. An intuition is as follows. If the action frequency σ^t converges to some σ, then from Proposition 7, the posterior belief $\hat{\mu}^t$ will be concentrated on $\triangle\Theta(\sigma)$ after a long time, that is, $\hat{\mu}^t$ is in a neighborhood of $\triangle\Theta(\sigma)$ for large t. If all the beliefs in this neighborhood induce the same equilibrium action $\hat{x}$ (i.e., $\hat{s}(\hat{\mu}) = \hat{x}$ for all beliefs $\hat{\mu}$ in a neighborhood of $\triangle\Theta(\sigma)$), then the action frequency will eventually converge to a point mass on $\hat{x}$. But in general, this need not be the case; different beliefs $\hat{\mu}$ and $\hat{\mu}'$ in this neighborhood may induce different equilibrium actions $\hat{x}$ and $\hat{x}'$. In such a case, both $\hat{x}$ and $\hat{x}'$ can be chosen infinitely often on the path, and hence have positive weights in the limiting action frequency σ.

Note, however, that in many applications, all GBNE are degenerate. Indeed, if (i) there is a unique equilibrium $\hat{x}$ for each belief $\hat{\mu}$ and (ii) identifiability holds in that there is a unique minimizer $\theta_{i,k}$ of the Kullback-Leibular divergence for each action frequency σ, then obviously any GBNE is degenerate. All our examples in the paper satisfy these assumptions.

Proposition 12 above implies that when agents are myopic, a limiting action frequency must be a GBNE. It turns out that the same result holds for dynamically optimal agents, provided that identifiability holds and agents play a Markov perfect equilibrium. This follows from the fact that under identifiability, the differential inclusion (26) for myopic agents is exactly the same as that for dynamically optimal agents who play a Markov perfect equilibrium. So all the results presented i the main text of the paper area valid for dynamically optimal agents, as long as identifiability holds.

A.4 Motion of the KL Minimizer

A.4.1 Identifiability and Differential Inclusion

Our Proposition 10 shows that the asymptotic motion of the action frequency σ^t is described by the differential inclusion (26). However, solving the differential inclusion (26) is not easy in general. For example, in many applications (including the ones in this paper), there are continuous actions, in which case the action frequency σ^t is a probability distribution over an infinite-dimensional (continuous) space, and thus the differential inclusion becomes an infinite-dimensional problem. In this section, we show that this dimensionality problem can be avoided if we look at the asymptotic motion of the belief, rather than that of the action frequency.

We will impose the following *identifiability* assumption, which requires that there be a unique KL minimizer $\theta_{i,k}(\sigma)$ for each measure $\sigma \in \triangle \hat{X}$. This assumption is satisfied in many applications, see Esponda and Pouzo (2016) for more detailed discussions on this assumption.

Assumption 3. For each i, k, and σ, there is a unique minimizer $\theta_{i,k}(\sigma) \in \Theta_{i,k}$ of the Kullback-Leibular divergence $K_{i,k}(\theta_{i,k}, \sigma)$.

Since $\Theta_{i,k}(\sigma)$ is upper hemi-continuous in σ, under the identifiability assumption, each KL minimizer $\theta_{i,k}(\sigma)$ is continuous in σ. The next lemma shows that $\theta(\sigma) = (\theta_{i,k}(\sigma))_{i,k}$ is Lipschitz continuous if some additional assumptions hold. With an abuse of notation, let $K_{i,k}(\theta_{i,k}, \hat{x}) = K_{i,k}(\theta_{i,k}, \sigma)$ for $\sigma = 1_{\hat{x}}$.

Assumption 4. The following conditions hold:

(i) For each i, k, and m, $\frac{\partial K_{i,k}(\theta_{i,k},\hat{x})}{\partial \theta_{i,k,m}} < \infty$, where $\theta_{i,k,m}$ denotes the m-th component of $\theta_{i,k}$. Also for each $\hat{x}$, $K_{i,k}(\theta_{i,k}, \hat{x})$ is twice-continuously differentiable with respect to $\theta_{i,k}$, that is, $\frac{\partial^2 K_{i,k}(\theta_{i,k},\hat{x})}{\partial \theta_{i,k,m} \partial \theta_{i,k,n}}$ is continuous in $\theta_{i,k}$.

(ii) $\frac{\partial K_{i,k}(\theta_{i,k},\hat{x})}{\partial \theta_{i,k,m}}$ is equi-Lipschitz continuous, that is, there is $L > 0$ such that $\left|\frac{\partial K_{i,k}(\theta_{i,k},\hat{x})}{\partial \theta_{i,k,m}} - \frac{\partial K_{i,k}(\theta_{i,k},\hat{x}')}{\partial \theta_{i,k,m}}\right| < L|\hat{x} - \hat{x}'|$ for all i, k, m, $\theta_{i,k}$, $\hat{x}$, and $\hat{x}'$.

(iii) The KL minimizer $\theta(\sigma)$ satisfies both the first-order and second-order conditions for each σ. (An implication is that the inverse of the Hessian matrix exists.)

Lemma 2. *$\theta(\sigma)$ is Lipschitz continuous in σ. That is, there is $L > 0$ such that $|\theta(\sigma) - \theta(\tilde{\sigma})| \leq L\|\sigma - \tilde{\sigma}\|$.*

Now we consider the motion of the KL minimizer $\theta^t = (\theta^t_{i,k})_{i,k}$. Let $\boldsymbol{w}_\theta$ denote the continuous-time interpolation of θ^t. Let $\nabla K_{i,k}(\theta_{i,k}, x) = (\frac{\partial K_{i,k}(\theta_{i,k},x)}{\partial \theta_{i,k,m}})_m$, and $\nabla K(\theta, x) = (\frac{\partial K_{i,k}(\theta_{i,k},x)}{\partial \theta_{i,k,m}})_{i,k,m}$. Also let $\nabla^2 K_{i,k}(\theta_{i,k}, \sigma)$ denote the Hessian matrix of $K_{i,k}(\theta_{i,k}, \sigma)$ with respect to $\theta_{i,k}$, that is, each component of $\nabla^2 K_{i,k}(\theta_{i,k}, \sigma)$ is $\frac{\partial^2 K_{i,k}(\theta_{i,k},\sigma)}{\partial \theta_{i,k,m} \partial \theta_{i,k,n}}$. Let $\nabla^2 K(\theta, \sigma)$ denote a block diagonal matrix whose main diagonal blocks are $\nabla^2 K_{i,k}(\theta_{i,k}, \sigma)$, that is,

$$\nabla^2 K(\theta, \sigma) = \begin{pmatrix} \nabla^2 K_{1,1}(\theta_{1,1}, \sigma) & & 0 \\ & \nabla^2 K_{1,2}(\theta_{1,2}, \sigma) & \\ 0 & & \ddots \end{pmatrix}.$$

With an abuse of notation, let $S_0(\theta)$ denote $S_0(\sigma)$ for σ with $\theta(\sigma) = \theta$. The following proposition shows that the asymptotic motion of the KL minimizer is described by the differential inclusion

$$\dot{\boldsymbol{\theta}}(t) \in \bigcup_{\sigma:\theta(\sigma)=\boldsymbol{\theta}(t)} \bigcup_{\sigma' \in \triangle S_0(\boldsymbol{\theta}(t))} -(\nabla^2 K(\boldsymbol{\theta}(t), \sigma))^{-1} \left(\nabla K(\boldsymbol{\theta}(t)), \sigma'\right). \tag{27}$$

Let $Z_\theta(\boldsymbol{\theta}(0))$ be the set of solutions to the differential inclusion (27) with the initial value $\boldsymbol{\theta}(0)$. Also let $Z_\theta(\boldsymbol{\theta}(0)) = \bigcup_{\boldsymbol{\theta}(0)} Z_\theta(\boldsymbol{\theta}(0))$ denote the set of all solutions.

Proposition 13. *Suppose that Assumptions 3 and 4 hold. Then for any $T > 0$ and any sample path $h \in \mathscr{H}$,*

$$\lim_{t\to\infty} \inf_{\boldsymbol{\theta} \in Z_\theta} \sup_{\tau \in [0,T]} |\boldsymbol{w}_\theta(h)[t+\tau] - \boldsymbol{\theta}(\tau)| = 0.$$

In particular, when $S_0(\sigma)$ is a singleton for all σ, we have

$$\lim_{t\to\infty} \inf_{\boldsymbol{\theta} \in Z_\theta(\boldsymbol{w}_\theta(h)[t])} \sup_{\tau \in [0,T]} |\boldsymbol{w}_\theta(h)[t+\tau] - \boldsymbol{\theta}(\tau)| = 0.$$

To interpret the differential inclusion (27), consider the special case in which $\Theta_{i,k} \subset \boldsymbol{R}$, i.e., assume that agent k's model $\theta_{i,k}$ is one-dimensional. Then from (26), we have

$$\dot{\boldsymbol{\theta}}_{i,k}(t) \in \bigcup_{\sigma:\theta(\sigma)=\boldsymbol{\theta}(t)} \bigcup_{\sigma' \in \triangle S_0(\boldsymbol{\theta}(t))} -\frac{K'_{i,k}(\boldsymbol{\theta}_{i,k}(t), \sigma')}{K''_{i,k}(\boldsymbol{\theta}_{i,k}(t), \sigma)} \tag{28}$$

for each i and k, where $K'_{i,k}(\theta,\sigma) = \frac{\partial K_{i,k}(\theta,\sigma)}{\partial \theta}$ and $K''_{i,k}(\theta,\sigma) = \frac{\partial^2 K_{i,k}(\theta,\sigma)}{\partial \theta^2}$.

The denominator $K''_{i,k}(\boldsymbol{\theta}_{i,k}(t),\sigma)$ measures the curvature of the Kullback-Leibular divergence. Note that this term is always positive, because the second-order condition must be satisfied (Assumption 4(iii)). So this term influences the absolute value of $\boldsymbol{\theta}(t)$, but not the sign of $\dot{\boldsymbol{\theta}}_{i,k}(t)$; this in turn implies that this denominator influences the speed of $\boldsymbol{\theta}_{i,k}(t)$, but not the direction. Intuitively, when the curve is flatter (i.e., $K''_{i,k}$ is close to zero), all models in a neighborhood of $\boldsymbol{\theta}(t)$ almost equally fit the past data. Hence the KL minimizer $\boldsymbol{\theta}(t)$ is more sensitive to the new data generated by today's action, and it changes quickly.

The numerator $-K'_{i,k}(\boldsymbol{\theta}_{i,k}(t),\sigma')$ measures how much an increase in $\theta_{i,k}$ improves fitness to the new data generated by today's action σ'. This term influences the sign of $\dot{\boldsymbol{\theta}}_{i,k}(t)$, so it determines whether $\boldsymbol{\theta}_{i,k}(t)$ moves up or down. Intuitively, when this numerator is positive, (at least in a neighborhood of $\boldsymbol{\theta}(t)$) higher θ better explains the new data generated by today's action, so $\boldsymbol{\theta}(t)$ moves up. On the other hand, when this numerator is negative, lower θ better explains the new data, so $\boldsymbol{\theta}(t)$ moves down.

When we consider the dynamic of $\theta^t = \theta(\sigma^t)$, the drift of θ^t cannot be uniquely determined, for two reasons. First, the KL minimizer θ^t may not uniquely determine the agents' actions today, in the sense that $S_0(\theta^t)$ may not be a singleton. (As pointed out by Esponda, Pouzo, and Yamamoto (2021), in the single-agent setup, this happens when the agent is indifferent over multiple actions at a model $\theta = \theta^t$.) In our differential inclusion (28), this multiplicity is captured by taking the union over $\sigma' \in \triangle S_0(\boldsymbol{\theta}(t))$. Note that the same multiplicity problem appears in the differential inclusion (26).

Second, the KL minimizer θ^t may not uniquely determine the past action frequency, in the sense that there may be more than one σ such that $\theta(\sigma) = \theta^t$. Note that even if two action frequencies σ and $\tilde{\sigma}$ yield the same KL minimizer (i.e., $\theta(\sigma) = \theta(\tilde{\sigma})$), they may yield different curvatures of the KL divergence, so they influence the speed of $\boldsymbol{\theta}_{i,k}(t)$ differently. In our differential inclusion, this multiplicity is captured by taking the union over σ with $\theta(\sigma) = \boldsymbol{\theta}(t)$.

B Non-Convergence Theorem

In this appendix, we will extend the non-convergence theorem of Pemantle (1990) and show that the same non-convergence result holds in our setup. This result is used in the proofs of the various non-convergence results in the main text.

Consider a stochastic difference equation

$$v(t+1)-v(t)=\frac{1}{t+1}\left(F(v(t))+b(t,v(t))\varepsilon\right) \tag{29}$$

where $v(t)\in \boldsymbol{R}^n$, $F:\boldsymbol{R}^n\to\boldsymbol{R}^n$, $b(t,v(t))\in\boldsymbol{R}^n$, and $\varepsilon\sim N(0,1)$. We assume that F is Lipschitz-continuous, and that there is $\overline{b}$ such that $|b_i(t,v)|<\overline{b}$ for all i, t, and $v\in\boldsymbol{R}^n$, where $b_i(t,v)$ is the ith component of the vector $b(t,v)$. This second assumption essentially means that the variance of the noise is bounded.

A stochastic process $\{v(t)\}_{t=1}^{\infty}$ is a *perturbed solution to* (29) if it solves

$$v(t+1)-v(t)=\frac{1}{t+1}\left(\tilde{F}(t,v(t))+b(t,v(t))\varepsilon\right)$$

for some $\tilde{F}$ such that there is $K>0$ and $\alpha>0$ such that for all t and v,

$$|F(v)-\tilde{F}(t,v)|<\frac{K}{t^{\alpha}}.$$

It follows from the theory of stochastic approximation (e.g, Theorem 2.1 of Kushner and Yin (2003)) that if a stochastic process $\{v(t)\}$ is a perturbed solution to (29), and if this process $\{v(t)\}_{t=1}^{\infty}$ is bounded with probability one, i.e.,

$$\Pr\left(\limsup_{t\to\infty}|v(t)|<\infty\right)=1.$$

then the asymptotic motion $\{v(t)\}$ is approximated by the ODE

$$\frac{dw(t)}{dt}=F(v(t)). \tag{30}$$

Here the meaning of “approximation” is the same as that in Proposition 10.

A point $p\in\boldsymbol{R}^n$ is a *steady state* of the ODE if $F(p)=0$. A steady state p is *linearly unstable* if the Jacobian of F at p has at least one eigenvalue with a positive real part. Pemantle (1990) shows that there is zero probability of the stochastic process converging to linearly unstable steady states, i.e., $\Pr(\lim_{t\to\infty}v(t)=p)=0$ for any linearly unstable steady state p, under a few technical assumptions. His result does not apply to our setup, because (i) we consider a perturbed solution to (29), (ii) the noise term ε has an unbounded support, and (iii) the noise term ε is common for all variables, F_1, $\cdots$, F_n. An important consequence of the second assumption (ii) is that the step size $v(t+1)-v(t)$ is bounded by $\frac{\tilde{c}}{t+1}$, which is frequently used in Pemantle’s proof.

The following proposition shows that these features are not essential and Pemantle’s result still holds in our model.

Proposition 14. *Let p be a linearly unstable steady state of the ODE (30). Let H be the affine space spanned by the eigenvectors associated with the eigenvalues with negative real parts, and let H^* be the set of all unit vectors orthogonal to H. Assume that there is $\kappa > 0$, $t^* > 0$, and a neighborhood U of p such that $|b(t,v) \cdot h| \geq \kappa$ for all $h \in H^*$, $t \geq t^*$, and $v \in U$. If there is K and a neighborhood U' of p such that $|F(v) - \tilde{F}(t,v)| < \frac{K}{t}$ for all $v \in U'$ and t, then* $\Pr(\lim_{t\to\infty} v(t) = p) = 0$.

C Proofs

C.1 Proof of Proposition 1

In this proof, we will use the tools developed in Section A. Let $h = (x^t, y^t)_{t=1}^{\infty}$ denote a sample path of the infinite horizon game. Given a sample path h, let $\sigma^t(h) \in \triangle X$ denote the action frequency up to period t, i.e.,

$$\sigma^t(h)[x] = \frac{|\{\tau \leq t | x^t = x\}|}{t}$$

for each action profile x. Proposition 7 shows that almost surely, each player i's belief in a later period t will be concentrated on the minimizer of the KL divergence (the surprise function) with weight σ^{t-1}. More formally, there is a set $\mathscr{H}$ of sample paths such that a sample path h must be in this set $\mathscr{H}$ with probability one, and such that for any sample path $h \in \mathscr{H}$, each player i's belief in period t is approximately $1_{\theta_i(\sigma^{t-1}(h))}$ for large t. This result immediately implies that player 1 correctly learns the true state θ^*, as her KL minimizer is constant and $\theta_1(\sigma) = \theta^*$ for any frequency $\sigma \in \triangle X$.

We will show that player 2's belief also converges to the steady-state belief almost surely. For this, it suffices to show that for every sample path $h \in \mathscr{H}$, her KL minimizer $\theta_2(\sigma^t(h))$ converges to the steady state. In what follows, we will prove a bit stronger result; we allow multiple steady states, and show that for each sample path $h \in \mathscr{H}$, $\lim_{t\to\infty} d(\sigma^t(h), E_2) = 0$ where E_2 is the set of all steady-state beliefs θ of player 2. This implies that player 2's belief converges even when the steady state is not unique.

So pick an arbitrary sample path $h \in \mathscr{H}$. To think about a dynamic of the KL minimizer $\theta_2^t(h) = \theta_2(\sigma^t(h))$, Proposition 13 is useful; it shows that the motion of θ_2^t is asymptotically approximated by the differential inclusion (28), which reduces to the one-dimensional differential inclusion

$$\dot{\boldsymbol{\theta}}_2(t) \in \bigcup_{\sigma:\theta(\sigma)=\boldsymbol{\theta}(t)} -\frac{K_2'(\boldsymbol{\theta}_2(t), s(1_{\theta^*}, 1_{\boldsymbol{\theta}_2(t)}))}{K_2''(\boldsymbol{\theta}_2(t), \sigma)} \tag{31}$$

where $K_2' = \frac{\partial K_2}{\partial \theta}$ and $K_2'' = \frac{\partial^2 K_2}{\partial \theta^2}$. Here we ignore the dynamic of player 1's KL minimizer θ_1, as it is constant and $\theta_1(\sigma) = \theta^*$ for all σ. With an abuse of notation, let $Z_\theta(\theta)$ denote the set of solutions to the differential inclusion above with an initial value $\theta \in \Theta$.

We will consider the following two cases separately.

C.1.1 Case 1: $\liminf_{t\to\infty} \theta_2^t(h) \neq \limsup_{t\to\infty} \theta_2^t(h)$.

We will show that $[\liminf_{t\to\infty} \theta_2^t(h), \limsup_{t\to\infty} \theta_2^t(h)] \subseteq E_2$.

Suppose not, so that there is a model $\theta' \in [\liminf_{t\to\infty} \theta_2^t(h), \limsup_{t\to\infty} \theta_2^t(h)]$ such that $\theta' \notin E_2$. Then $K_2'(\theta', s(1_{\theta^*}, 1_{\theta'})) \neq 0$, meaning that (i) $K_2'(\theta', s(1_{\theta^*}, 1_{\theta'})) > 0$ or (ii) $K_2'(\theta', s(1_{\theta^*}, 1_{\theta'})) < 0$. In what follows, we will focus on the case (i). The proof for the case (ii) is symmetric.

Since $K_2'(\theta, \sigma)$ is continuous in (θ, σ) and $s(1_{\theta^*}, 1_\theta)$ is continuous in θ, there is $\varepsilon > 0$ such that $K_2'(\theta, s(1_{\theta^*}, 1_\theta)) > 0$ for any θ with $|\theta - \theta'| \leq \varepsilon$. Pick such $\varepsilon > 0$. Then the right-hand side of (31) is positive for any $\boldsymbol{\theta(t)}$ in the ε-neighborhood of θ', which means that $\boldsymbol{\theta}(t)$ increases as time goes in this neighborhood.[33] Hence there is $T > 0$ such that

$$\boldsymbol{\theta}_2(t) \geq \theta' + \varepsilon \tag{32}$$

for any $t \geq T$ and for any solution $\boldsymbol{\theta}_2 \in Z_\theta(\theta)$ to the differential inclusion with any initial value θ with $\theta \geq \theta_2' - \varepsilon$. Pick such T.

With an abuse of notation, let $\boldsymbol{w}_\theta(t)$ denote the continuous-time interpolation of the KL minimizer $(\theta_2^t(h))_{t=1}^\infty$. From Proposition 13, there is t^* such that for any $t > t^*$, $\boldsymbol{\theta}_2 \in Z_\theta(\boldsymbol{w}_\theta(t))$, and $s \in [0, 2T]$,

$$|\boldsymbol{w}_\theta(t+s) - \boldsymbol{\theta}_2(s)| < \frac{\varepsilon}{2}. \tag{33}$$

Pick such t^*. Since $\theta' \leq \limsup_{t\to\infty} \theta_2^t(h)$, there is $t^{**} > t^*$ such that $\boldsymbol{w}_\theta(t^{**}) \geq \theta' - \varepsilon$. Pick such t^{**}. Then from (32), we have

$$\boldsymbol{\theta}_2(s) \geq \theta' + \varepsilon$$

[33] Note that $K'' < 0$ because K is convex.

for any $s \geq T$ and for any solution $\boldsymbol{\theta} \in Z_\theta(\boldsymbol{w}_\theta(t^{**}))$. This inequality and (33) implies

$$\boldsymbol{w}_\theta(t^{**}+s) \geq \theta' + \frac{\varepsilon}{2} \quad \forall s \in [T, 2T].$$

Likewise, since $\boldsymbol{w}_\theta(t^{**}+T) \geq \theta' + \frac{\varepsilon}{2}$, it follows from (32) that

$$\boldsymbol{\theta}_2(s) \geq \theta' + \varepsilon$$

for any $s \geq T$ and for any solution $\boldsymbol{\theta}_2 \in Z_\theta(\boldsymbol{w}_\theta(t^{**}+T))$. This inequality and (33) implies

$$\boldsymbol{w}_\theta(t^{**}+s) \geq \theta' + \frac{\varepsilon}{2} \quad \forall s \in [2T, 3T].$$

Iterating this argument, we can show that

$$\boldsymbol{w}_\theta(t^{**}+s) \geq \theta' + \frac{\varepsilon}{2} \quad \forall s \in [T, \infty).$$

But this means that $\liminf_{t\to\infty} \theta_2^t(h) \geq \theta' + \frac{\varepsilon}{2}$, which is a contradiction.

C.1.2 Case 2: $\liminf_{t\to\infty} \theta_{i,k}^t(h) = \limsup_{t\to\infty} \theta_{i,k}^t(h)$.

In this case, $\lim_{t\to\infty} \theta_{i,k}^t(h)$ exists. Let $\theta_{i,k}^* = \lim_{t\to\infty} \theta_{i,k}^t(h)$. We will show that $\theta_{i,k}^* \in E$.

Suppose not so that $\theta^* \notin E$. Then as in the previous case, (i) $K'_{i,k}(\theta_{i,k}^*, \sigma') > 0$ for all $\sigma' \in \triangle S_0(\theta(\theta_{i,k}^*))$, or (ii) $K'_{i,k}(\theta_{i,k}^*, \sigma') < 0$ for all $\sigma' \in \triangle S_0(\theta(\theta_{i,k}^*))$. We will focus on the case (i).

As in the previous case, there is $\varepsilon > 0$ such that $K'_{i,k}(\theta_{i,k}, \sigma') > 0$ for any $\theta_{i,k}$ with $|\theta_{i,k} - \theta_{i,k}^*| \leq \varepsilon$ and any $\sigma' \in \triangle S_0(\theta(\theta_{i,k}))$. Pick such $\varepsilon > 0$. Then pick T such that (32) holds for any $t \geq T$ and for any solution $\boldsymbol{\theta} \in Z_\theta(\theta(\theta_{i,k}))$ with any $\theta_{i,k}$ with $\theta_{i,k} \geq \theta_{i,k}^* - \varepsilon$.

From Proposition 13, there is t^* such that (33) holds for any $t > t^*$, $\boldsymbol{\theta} \in Z'_\theta(\boldsymbol{w}_\theta(t))$, and $s \in [0, 2T]$. Pick such t^*. Since $\theta_{i,k}^* = \lim_{t\to\infty} \theta^t(h)$, there is $t^{**} > t^*$ such that $\boldsymbol{w}_{\theta,i,k}(t^{**}) \geq \theta_{i,k}^* - \varepsilon$. Pick such t^{**}. Then as in the previous case, we can show that

$$\boldsymbol{w}_{\theta,i,k}(t^{**}+s) \geq \theta_{i,k}^* + \frac{\varepsilon}{2} \quad \forall s \in [T, \infty).$$

But this means that $\lim_{t\to\infty} \theta_{i,k}^t(h) \geq \theta_{i,k}^* + \frac{\varepsilon}{2}$, which is a contradiction. *Q.E.D.*

C.2 Proof of Proposition 2

For part (i), note that when $A_2 = a(= A_1)$, players have the same view about the world and hence have the same posterior belief $\mu_1^t = \mu_2^t$ every period. Hence the problem reduces to the classical individual learning problem with no misspecification, and the result follows from a standard argument (a version of the law of large numbers).

So in what follows, we will prove part (ii) of the proposition. We will first show that the process converges to the interior steady state with zero probability. Then we will show that the process converges to the boundary steady states.

Part 1: Non-convergence to p First, we will show that there is zero probability of the process converging to the interior steady state p. For this, it suffices to show that assumptions (i)-(vii) stated in Proposition 3 hold and that the interior steady state p is linearly unstable.

From the first-order condition, a Nash equilibrium $(x_i, \hat{x}_{-i})$ given θ_i is unique and $x_i = \hat{x}_{-i} = 1 - \theta_i$. Hence assumption (i) holds.

Assumptions (ii)-(v) are obviously satisfied. To check assumptions (vi), let $\tilde{m}(m, \xi)$ denote the mean of the truncated normal $\tilde{N}(m, \frac{1}{\xi})$. Since each player's payoffs are linear in θ, given a posterior belief $\mu_i^t = \tilde{N}\left(m_i^t, \frac{1}{(t-1)\xi_i^t}\right)$, player i and hypothetical player j chooses a Nash equilibrium for a state $\theta = \tilde{m}(m_i^t, (t-1)\xi_i^t)$. So from the Lipschitz-continuity of θ_i and I_i, assumption (vi) follows from part (iv) of the next lemma. (Parts (i)-(iii) of this lemma are not used here, but we will use them when we prove convergence to the boundary steady states.)

Lemma 3. *There is* $k > 0$ *and* $\bar{t} > 0$ *such that for all* $t > \bar{t}$ *and all* ξ *which arises on the equilibrium path,*

(i) $|\tilde{m}(m, t\xi) - m| < \frac{k}{\sqrt{t}}$ *for all* $m \in \Theta$*,*

(ii) $|\tilde{m}(m, t\xi) - \underline{\theta}| < \frac{k}{\sqrt{t}}$ *for all* $m < \underline{\theta}$*,*

(iii) $|\tilde{m}(m, t\xi) - \overline{\theta}| < \frac{k}{\sqrt{t}}$ *for all* $m > \overline{\theta}$*.*

Also, for any interior point $\theta^* \in \Theta$*, there is a neighborhood* U *of* θ^**,* $k > 0$*, and* $\bar{t}' > 0$ *such that for all* $t > \bar{t}'$ *and* $m \in U$*,*

(iv) $|\tilde{m}(m, t\xi) - m| < \frac{k}{t}$*.*

Proof. Let ξ be the minimum of $I_i(x_i, \hat{x}_{-i})$ over all actions $(x_i, \hat{x}_{-i})$ which can be chosen on the equilibrium path, and we will show that (i)-(iv) hold for this particular ξ. Then it is straightforward to see that (i)-(iv) holds for all other ξ.

Let ϕ denote the pdf of the standard normal $N(0,1)$, and let Φ denote its cdf. Pick some truncated normal distribution $\tilde{N}(m, \frac{1}{t\xi})$. It is well-known that that the mean of this truncated normal distribution is

$$\tilde{m}(m, t\xi) = m + \frac{1}{\sqrt{t\xi}} \cdot \frac{\phi(\sqrt{t\xi}(\overline{\theta} - m)) - \phi(\sqrt{t\xi}(\underline{\theta} - m))}{\Phi(\sqrt{t\xi}(\overline{\theta} - m)) - \Phi(\sqrt{t\xi}(\underline{\theta} - m))} \tag{34}$$

Since $\phi(x) = \frac{1}{\sqrt{2\pi}} \exp(-0.5x^2) \leq \frac{1}{\sqrt{2\pi}}$,

$$|\phi(\sqrt{t\xi}(\overline{\theta} - m)) - \phi(\sqrt{t\xi}(\underline{\theta} - m))| < \frac{1}{\sqrt{2\pi}}.$$

Also there is $\bar{t} > 0$ such that for all $m \in \Theta$ and $t > \bar{t}$,

$$\Phi(\sqrt{t\xi}(\overline{\theta} - m)) - \Phi(\sqrt{t\xi}(\underline{\theta} - m)) > \frac{1}{3}. \tag{35}$$

Plugging these into (34), we have

$$|\tilde{m}(m, t\xi) - m| < \frac{1}{\sqrt{t\xi}} \cdot \frac{3}{\sqrt{2\pi}}$$

for all $m \in \Theta$ and $t > \bar{t}$, which implies (i).

Next, we will prove (iv). Note that

$$|\phi(\sqrt{t\xi}(\overline{\theta} - m)) - \phi(\sqrt{t\xi}(\underline{\theta} - m))| = \frac{1}{\sqrt{2\pi}} \left| \left(\frac{1}{(\sqrt{e})^{(\overline{\theta} - m)^2}} \right)^{t\xi} - \left(\frac{1}{(\sqrt{e})^{(\underline{\theta} - m)^2}} \right)^{t\xi} \right|.$$

Pick $\theta^* \in (\underline{\theta}, \overline{\theta})$. Then there is a neighborhood U of θ^* such that we have $\frac{1}{(\sqrt{e})^{(\overline{\theta} - m)^2}} < 1$ and $\frac{1}{(\sqrt{e})^{(\underline{\theta} - m)^2}} < 1$ for all $m \in U$. Then there is $\bar{t}'$ such that

$$|\phi(\sqrt{t\xi}(\overline{\theta} - m)) - \phi(\sqrt{t\xi}(\underline{\theta} - m))| < \frac{1}{t}$$

for all $m \in U$ and $t > \bar{t}'$. Plugging this and (35) into (34), we have (iv).

Finally, we will prove (ii) and (iii). Let $\tilde{\phi}(m, \frac{1}{\xi})$ denote the pdf of the truncated normal $\tilde{N}(m, \frac{1}{\xi})$. Then for any $x > 0$ and $m < \underline{\theta}$,

$$\frac{\tilde{\phi}\left(\underline{\theta}, \frac{1}{\xi}\right)[\underline{\theta} + x]}{\tilde{\phi}\left(\underline{\theta}, \frac{1}{\xi}\right)[\underline{\theta}]} = \frac{\phi(\sqrt{\xi}x)}{\phi(0)} > \frac{\phi(\sqrt{\xi}(\underline{\theta} - m + x))}{\phi(\sqrt{\xi}(\underline{\theta} - m))} = \frac{\tilde{\phi}\left(m, \frac{1}{\xi}\right)[\underline{\theta} + x]}{\tilde{\phi}\left(m, \frac{1}{\xi}\right)[\underline{\theta}]}.$$

This means that the truncated normal $\tilde{N}(\underline{\theta}, \frac{1}{\xi})$ first-order stochastically dominates $\tilde{N}(m, \frac{1}{\xi})$ for all $m < \underline{\theta}$. Hence

$$\underline{\theta} < \tilde{m}(m, \xi) < \tilde{m}(\underline{\theta}, \xi)$$

for all $m < \underline{\theta}$. Together with part (i) of the lemma, we obtain (ii). The proof of (iii) is similar and hence omitted. *Q.E.D.*

Next, we will check (vii). We need to show that $(\frac{1}{\sqrt{I_1}}, \frac{1}{\sqrt{I_2}})$ is not an eigenvector of J', i.e.,

$$\begin{pmatrix} \frac{\partial \theta_1}{\partial m_1} - 1 & \frac{\partial \theta_1}{\partial m_2} \\ \frac{\partial \theta_2}{\partial m_1} & \frac{\partial \theta_2}{\partial m_2} - 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{I_1(p)}} \\ \frac{1}{\sqrt{I_2(p)}} \end{pmatrix} \neq \lambda \cdot \begin{pmatrix} \frac{1}{\sqrt{I_1(p)}} \\ \frac{1}{\sqrt{I_2(p)}} \end{pmatrix}$$

for any $\lambda \in \boldsymbol{R}$. This is equivalent to show that

$$\left(\frac{\partial \theta_1}{\partial m_1} - 1\right) + \frac{\partial \theta_1}{\partial m_2} \frac{\sqrt{I_1(p)}}{\sqrt{I_2(p)}} \neq \frac{\partial \theta_2}{\partial m_1} \frac{\sqrt{I_2(p)}}{\sqrt{I_1(p)}} + \left(\frac{\partial \theta_2}{\partial m_2} - 1\right). \tag{36}$$

Note that $\theta_i(m)$ solves $Q(x_i(m_i), \hat{x}_{-i}(m_i), A_i, \theta_i) = Q(x_i(m_i), x_{-i}(m_{-i}), a, \theta^*)$. By the implicit function theorem, we have

$$\frac{\partial \theta_i}{\partial m_i} = -\frac{\frac{\partial Q_i}{\partial x_i}\frac{\partial x_i}{\partial m_i} + \frac{\partial Q_i}{\partial \hat{x}_{-i}}\frac{\partial \hat{x}_{-i}}{\partial m_i} - \frac{\partial Q^*}{\partial x_i}\frac{\partial x_i}{\partial m_i}}{\frac{\partial Q_i}{\partial \theta_i}} = -\frac{2\frac{\partial Q_i}{\partial x_i}\frac{\partial x_i}{\partial m_i} - \frac{\partial Q^*}{\partial x_i}\frac{\partial x_i}{\partial m_i}}{\sqrt{I_i}},$$

$$\frac{\partial \theta_i}{\partial m_j} = \frac{\frac{\partial Q^*}{\partial x_j}\frac{\partial x_j}{\partial m_j}}{\frac{\partial Q_i}{\partial \theta_i}} = \frac{\frac{\partial Q^*}{\partial x_j}\frac{\partial x_j}{\partial m_j}}{\sqrt{I_i}},$$

where $Q_i = Q(x_i(m_i), \hat{x}_{-i}(m_i), A_i, \theta_i)$ denotes player i's subjective expectation of the output and $Q^* = Q(x_i(m_i), x_{-i}(m_{-i}), a, \theta^*)$ is the true mean. Combining these two equalities, we have

$$\frac{\partial \theta_i}{\partial m_i} - 1 + \frac{\sqrt{I_i}}{\sqrt{I_j}} \frac{\partial \theta_i}{\partial m_j} = \frac{-2\frac{\partial Q_i}{\partial x_i}\frac{\partial x_i}{\partial m_i} + \frac{\partial Q^*}{\partial x_i}\frac{\partial x_i}{\partial m_i}}{\sqrt{I_i}} - 1 + \frac{\frac{\partial Q^*}{\partial x_{-i}}\frac{\partial x_{-i}}{\partial m_{-i}}}{\sqrt{I_{-i}}}.$$

Using this equation, (36) can be rewritten as

$$\frac{-2\frac{\partial Q_1}{\partial x_1}\frac{\partial x_1}{\partial m_1}}{\sqrt{I_1}} \neq \frac{-2\frac{\partial Q_2}{\partial x_2}\frac{\partial x_2}{\partial m_2}}{\sqrt{I_2}},$$

which is further simplified to

$$-\frac{m_1}{1-m_1} \neq -\frac{m_2}{1-m_2}$$

because $x_i = \hat{x}_{-1} = 1 - \theta_i$. This inequality indeed holds (and hence assumption (vii) is satisfied), because $\frac{m_i}{1-m_i}$ is increasing in m_i on the set Θ, and the consistency condition implies that $m_1^* \neq m_2^*$ in any interior steady state with $A_2 \neq A_1 = a$.

To conclude the proof, we will show that the interior steady state p is linearly unstable. From Proposition 4 (i), it suffices to show that $\frac{\partial \theta_i(m)}{\partial m_i} > 1$ for each i. For the special case with $A_1 = A_2 = a$, we have $\frac{\partial \theta_i(m)}{\partial m_i} = 2$. Then by the continuity, for any A_2 close to a, we still have $\frac{\partial \theta_i(m)}{\partial m_i} > 1$.

Part 2: Convergence to boundary beliefs We will first show that the stochastic process (m^t, ξ^t) is bounded with probability one. Recall that regardless of the parameter A_i, a Nash equilibrium given a state θ_i is $x_i = \hat{x}_{-i} = 1 - \theta_i$. Hence on the equilibrium path, each player's production is at least $\underline{x} = 1 - \overline{\theta}$ but does not exceed $\overline{x} = 1 - \underline{\theta}$.

Let $\underline{m}_i$ be such that

$$A_i - \underline{m}_i(\overline{x} + \overline{x}) = a - \theta^*(\overline{x} + \underline{x}).$$

In words, $\underline{m}_i \in \boldsymbol{R}$ denotes a state with which player i's subjective expectation about the output matches the true mean, when player i thinks that the opponent chooses the maximal effort $\overline{x}$ but in reality she chooses the minimal effort $\underline{x}$. Note that this $\underline{m}_i$ is the minimum of $\theta_i(m)$ over all m, and that $\underline{m}_i$ need not be in the state space Θ. Similarly, let $\overline{m}_i$ be such that

$$A_i - \overline{m}_i(\underline{x} + \underline{x}) = a - \theta^*(\underline{x} + \overline{x}).$$

This is a state with which player i's subjective expectation about the output matches the true mean, when player i thinks that the opponent chooses the minimal effort $\underline{x}$ but in reality she chooses the maximal effort $\overline{x}$ (which yields the most optimistic belief $\overline{m}_i$). Note that this $\overline{m}_i$ is the maximum of $\theta_i(m)$ over all m.

The following lemma shows that almost surely, m_i^t is in a neighborhood of $[\underline{m}_i, \overline{m}_i]$ after a long time. This immediately implies that the process (m^t, ξ^t) is bounded almost surely; indeed, $I_i(x^\tau)$ has the minimal value $\underline{I} = I_i(\overline{x}, \overline{x})$ and the maximal value $\overline{I} = I_i(\underline{x}, \underline{x})$, so it is obvious that ξ_i^t is always in the bounded interval $[\underline{I}, \overline{I}]$.

Lemma 4. *Given any A_2, almost surely, $\underline{m}_i \leq \liminf_{t\to\infty} m_i^t \leq \limsup_{t\to\infty} m_i^t \leq \overline{m}_i$ for each i.*

From Lemma 3, it is obvious that there is $K > 0$ such that (15) and (16) hold for $\alpha = 0.5$. Then since the process is bounded with probability one, Theorem 2.1 of Kushner and Yin (2003) implies the following lemma: Given a realized infinite-horizon outcome $(m^t, \xi^t)_{t=1^\infty}$, define the *continuous-time interpolation* as a mapping $\boldsymbol{w} : [0,\infty) \to \boldsymbol{R}^4$ such that

$$\boldsymbol{w}[\tau_t + s] = (m^t, \xi^t) + \frac{\tau}{\tau_{t+1} - \tau_t}((m^{t+1}, \xi^{t+1}) - (m^t, \xi^t))$$

for all $t = 0, 1, \cdots$ and $\tau \in [0, \frac{1}{t+1})$. This $\boldsymbol{w}$ is an *asymptotic pseudotrajectory* of the ODE if for any $T > 0$,

$$\lim_{t\to\infty} \sup_{\tau\in[0,T]} |\boldsymbol{w}(t+\tau) - s(\boldsymbol{w}(t))[\tau]| = 0 \tag{37}$$

where $s(m, \xi) : \boldsymbol{R}_+ \to \boldsymbol{R}^4$ is a solution to the ODE (17) and (18) given the initial value (m, ξ).

Lemma 5. *With probability one, **w** is an asymptotic pseudotrajectory of the ODE (17) and (18).*

This lemma implies that after a long time, the path $\boldsymbol{w}$ of the stochastic process is approximated by the solution s to the ODE (17) and (18). So in order to know the long-run outcome of the stochastic process, it suffices to investigate the ODE.

The next lemma characterizes the behavior of the solution to the ODE when player 2's overconfidence is small.

Lemma 6. *Pick a arbitrarily. There is $\overline{A} > a$ such that for any $A_2 \in [a, \overline{A})$ and i, there are values θ'_{-i} and θ''_{-i} with $\underline{\theta} < \theta'_{-i} < \theta''_{-i} < \overline{\theta}$ and differentiable functions $f_i : [\theta'_{-i}, \theta''_{-i}] \to \Theta$, $\tilde{f}_i : [\underline{\theta}, \theta''_{-i}] \to [\overline{\theta}, \overline{m}_i]$, and $\hat{f}_i : [\theta'_{-i}, \overline{\theta}] \to [\underline{m}_i, \underline{\theta}]$ such that the following properties hold:*

(i) $f_i'(m_{-i}) > 1$ *for all* m_{-i}, $f_i(\theta'_{-i}) = \underline{\theta}$, $f_i(\theta''_{-i}) = \overline{\theta}$, $\tilde{f}_i'(m_{-i}) < 0$ *for all* m_{-i}, $\tilde{f}_i(\underline{\theta}) = \overline{m}_i$, $\tilde{f}_i(\theta''_{-i}) = \overline{\theta}$, $\hat{f}_i'(m_{-i}) < 0$ *for all* m_{-i}, $\hat{f}_i(\theta'_{-i}) = \underline{\theta}$, $\hat{f}_i(\overline{\theta}) = \underline{m}_i$,

(ii) For any $m_{-i} < \underline{\theta}$, $\theta_i(m) - m_i$ *is positive if* $m_i < \overline{m}_i$, *is zero if* $m_i = \overline{m}_i$, *and is negative if* $m_i > \overline{m}_i$.

(iii) For any $m_{-i} \in [\underline{\theta}, \theta'_{-i})$, $\theta_i(m) - m_i$ *is positive if* $m_i < \tilde{f}_i(m_{-i})$, *is zero if* $m_i = \tilde{f}_i(m_{-i})$, *and is negative if* $m_i > \tilde{f}_i(m_{-i})$,

(iv) For any $m_{-i} \in [\theta'_{-i}, \theta''_{-i}]$, $\theta_i(m) - m_i$ *is positive if* $m_i < \hat{f}_i(m_{-i})$, *is zero if* $m_i = \hat{f}_i(m_{-i})$, *is negative if* $m_i \in (\hat{f}_i(m_{-i}), f_i(m_{-i}))$, *is zero if* $m_i = f_i(m_{-i})$, *is positive if* $m_i \in (f_i(m_{-i}), \tilde{f}_i(m_{-i}))$, *is zero if* $m_i = \tilde{f}_i(m_{-i})$, *and is negative if* $m_i > \tilde{f}_i(m_{-i})$.

v) For any $m_{-i} \in (\theta''_{-i}, \overline{\theta}]$, $\theta_i(m) - m_i$ *is positive if* $m_i < \hat{f}_i(m_{-i})$, *is zero if* $m_i = \hat{f}_i(m_{-i})$, *and and is negative if* $m_i > \hat{f}_i(m_{-i})$.

(vi) For any $m_{-i} > \overline{\theta}$, $\theta_i(m) - m_i$ *is positive if* $m_i < \underline{m}_i$, *is zero if* $m_i = \underline{m}_i$, *and is negative if* $m_i > \underline{m}_i$.

Proof. We will first explain how to choose θ'_{-i}, θ''_{-i}, f_i, $\tilde{f}_i$, and $\hat{f}_i$. Let θ'_{-i} be a state θ which solves

$$A_i - \underline{\theta}(x_i(\underline{\theta}) + x_{-i}(\underline{\theta})) = a - \theta^*(x_i(\underline{\theta}) + x_{-i}(\theta)).$$

When $A_i = a$, the right-hand side $(a - \theta^*(2 - \underline{\theta} - \theta))$ is less than the left-hand side $(a - \underline{\theta}(2 - 2\underline{\theta}))$ at $\theta = \underline{\theta}$, and is greater than that at $\theta = \theta^*$. Also the right-hand side is increasing in θ. Hence θ'_{-i} which solves the equality above is unique and $\underline{\theta} < \theta'_{-i} < \theta^*$. Then by the continuity, the same result holds as long as A_2 is close to a.

Similarly, let θ''_{-i} be a state θ which solves

$$A_i - \overline{\theta}(x_i(\overline{\theta}) + x_{-i}(\overline{\theta})) = a - \theta^*(x_i(\overline{\theta}) + x_{-i}(\theta)).$$

Then again, for A_i close to a, θ''_{-i} is uniquely determined and $\theta^* < \theta''_{-i} < \overline{\theta}$. Hence we have $\underline{\theta} < \theta'_{-i} < \theta''_{-i} < \overline{\theta}$ as stated in the lemma.

Then for each $m_{-i} \in [\underline{\theta}, \theta''_{-i}]$, define $\tilde{f}_i(m_{-i})$ as a value m_i which solves

$$A_i - m_i(x_i(\overline{\theta}) + x_{-i}(\overline{\theta})) = a - \theta^*(x_i(\overline{\theta}) + x_{-i}(m_{-i})),$$

i.e., with this belief m_i, player i's subjective expectation about the output matches the true mean when she believes that the Nash equilibrium for $\overline{\theta}$ will be chosen but in reality the opponent chooses the Nash equilibrium action for m_{-i}. Note that the above equation is linear in m_i, and hence indeed has a unique solution. By the definition, $\tilde{f}_i(\underline{\theta}) = \overline{m}_i$ and $\tilde{f}_i(\theta''_{-i}) = \overline{\theta}$. Also by the implicit function theorem, $\tilde{f}_i(m_{-i})$ is decreasing in m_{-i}, as stated in the lemma.

Similarly, for each $m_{-i} \in [\theta'_{-i}, \overline{\theta}]$, define $\hat{f}_i(m_{-i})$ as a value m_i which solves

$$A_i - m_i(x_i(\underline{\theta}) + x_{-i}(\underline{\theta})) = a - \theta^*(x_i(\underline{\theta}) + x_{-i}(m_{-i})).$$

Again this equation is linear in m_i, and hence has a unique solution. Also it is easy to check that $\hat{f}_i(\theta'_{-i}) = \underline{\theta}$, $\hat{f}_i(\overline{\theta}) = \underline{m}_i$, and $\hat{f}_i(m_{-i})$ is decreasing in m_{-i}.

Also for each $m_{-i} \in [\theta'_{-i}, \theta'']$, define $f_i(m_{-i})$ as a value $m_i \in \Theta$ which solves

$$A_i - m_i(x_i(m_i) + x_{-i}(m_i)) = a - \theta^*(x_i(m_i) + x_{-i}(m_{-i})).$$

To see that this equation has a solution, let

$$g(m_i, m_{-i}) = A_i - m_i(x_i(m_i) + x_{-i}(m_i)) - a + \theta^*(x_i(m_i) + x_{-i}(m_{-i})).$$

By the definition of θ''_{-i}, $g(\overline{\theta}, \theta''_{-i}) = 0$. Then since g is decreasing in m_{-i}, we have $g(\overline{\theta}, m_{-i}) \geq 0$ for all $m_{-i} \in [\theta'_{-i}, \theta'']$. Likewise, since $g(\underline{\theta}, \theta'_{-i}) = 0$. we have $g(\underline{\theta}, m_{-i}) \leq 0$ for all $m_{-i} \in [\theta'_{-i}, \theta'']$. Taken together, given any $m_{-i} \in [\theta'_{-i}, \theta'']$, we have $g(\underline{\theta}, m_{-i}) \leq 0 \leq g(\overline{\theta}, m_{-i})$, so there is at least one $m_i \in \Theta$ which solves $g(m_i, m_{-i}) = 0$. Also this solution is unique, because given any $m_{-i} \in [\theta'_{-i}, \theta'']$, g is strictly increasing in m_i when $m_i \in \Theta$. (Note that g is a quadratic function of m_i.)

By the definition of θ'_{-i} and θ''_{-i}, we have $f_i(\theta'_{-i}) = \underline{\theta}$ and $f_i(\theta''_{-i}) = \overline{\theta}$. Also, by the implicit function theorem,

$$f'_i(m_{-i}) = -\frac{\frac{\partial g}{\partial m_{-i}}}{\frac{\partial g}{\partial m_i}} = \frac{\theta^*}{-2 + 4m_i - \theta^*}.$$

We have $f'_i(m_{-i}) = 2$ at $m_i = m_{-i} = \theta^* = 0.8$ and $f'_i(m_{-i}) > 1$ for any $m_i, m_{-i} \in \Theta$. So all the properties stated in part (i) holds.

Next, we will prove part (iv). Pick $m_{-i} \in (\underline{\theta'}_{-i}, \underline{\theta''}_{-i})$ arbitrarily. By the definition of $\hat{f}_i$, we have $\theta_i(m) = \hat{f}_i(m_{-i})$ for any $m_i \leq \underline{\theta}$. Hence $\theta_i(m) - m_i$ is positive for $m_i < \hat{f}_i(m_{-i})$, is zero for $m_i = \hat{f}_i(m_{-i})$, and is negative for $m_i \in (\hat{f}_i(m_{-i}), \underline{\theta}]$, as stated in the lemma.

For $m_i \in (\underline{\theta}, f_i(m_{-i})$, we claim that $\theta_i(m) - m_i$ is negative. Suppose not so that $\theta_i(m) - m_i \geq 0$. If $\theta_i(m) - m_i = 0$, then by the definition of f_i, we must have $m_i = f_i(m_{-i})$, which contradicts with $m_i < f_i(m_{-i})$. If $\theta_i(m) - m_i > 0$, then there must be $m_i' \in (\underline{\theta}, m_i)$ such that $\theta_i(m_i', m_{-i}) - m_i' = 0$. (This is so because $\theta_i(\underline{\theta}, m_{-i}) - \underline{\theta} < 0$.) But then we must have $m_i' = f_i(m_{-i})$, which is a contradiction. Hence $\theta_i(m) - m_i$ is negative in this case.

By the symmetry, for $m_i > f_i(m_{-i})$, all the properties stated in part (iv) of the lemma are satisfied. Also, by the definition of f_i, we have $\theta_i(m) - m_i = 0$ for $m_i = f_i(m_{-i})$. Hence part (iv) follows.

The proofs of the other parts of the lemma are very similar, and hence omitted.

Q.E.D.

Figure 7 highlights what is shown in the lemma above. Here the the horizontal axis represents m_{-i} and the vertical axis represents m_i. The origin is the interior steady state belief. The large dotted square is $\times_{i=1,2}[\underline{m}_i, \overline{m}_i]$, and recall that after a long time, (m_1^t, m_2^t) is in a neighborhood of this square almost surely. The small dotted square is the state space $\times_{i=1,2}\Theta$. The thick polygonal line is the set of points at which $\frac{d\theta_i(t)}{dt} = \theta_i(m(t)) - m_i(t) = 0$; the downward-sloping line at the top is the graph of the function $\tilde{f}_i(m_{-i})$ defined in the lemma above, the upward-sloping line in the middle is the graph of f_i, and the downward-sloping line at the bottom is the graph of $\hat{f}_i$. On the left side of this thick line, $\frac{d\theta_i(t)}{dt} = \theta_i(m(t)) - m_i(t) > 0$, which means that the solution $\theta_i(t)$ to the ODE increases over time. In contrast, on the right side of the line, $\frac{d\theta_i(t)}{dt} = \theta_i(m(t)) - m_i(t) < 0$, and hence $\theta_i(t)$ decreases over time. See the thick arrows in the figure.

Figure 8 describes how the solution to the ODE behaves when both $m_1(t)$ and $m_2(t)$ change over time. The horizontal axis represents m_1 and the vertical axis represents m_2. The two thick polygonal lines are the set of points at which $\frac{dm_i(t)}{dt} = 0$. If the current value $m(t)$ is on the polygonal line with $\frac{dm_1(t)}{dt} = 0$, only $m_2(t)$ changes at the next instant, so $m(t)$ moves vertically, as shown by the arrows in the figure. Similarly, If the current value is on the polygonal line with $\frac{dm_2(t)}{dt} = 0$, only $m_1(t)$ changes at the next instant, so $m(t)$ moves horizontally. For all other points, both m_1 and m_2 move simultaneously. We cannot pin down the exact motion of $m(t)$ in this case (hence we have fork arrows in the picture) because it depends on the current value of $\xi(t)$, which is not specified here; in general, when ξ_1 is relatively larger than ξ_2, m_1 moves faster than m_2, and hence the arrow becomes flatter.

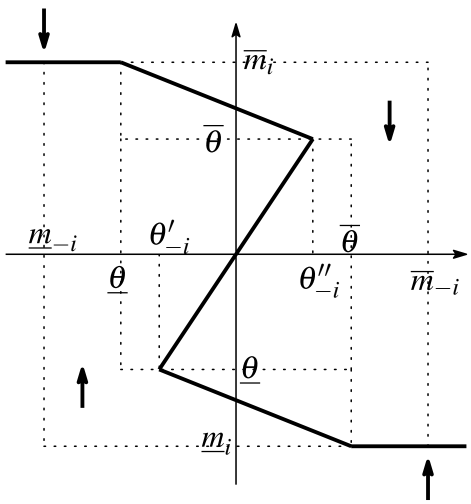

Figure 7: Motion of $m_i(t)$ for Fixed m_{-i}

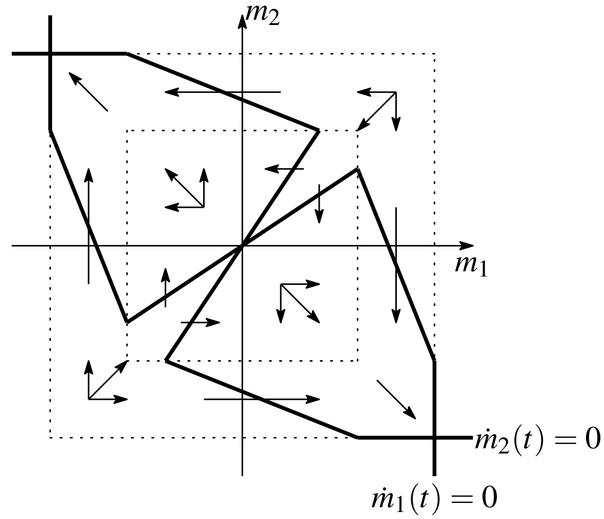

Figure 8: Motion of $m(t)$

As can be seen from the figure, the polygonal lines intersect three times, and these are the steady states of the ODE. That is, the ODE have one interior steady state (the origin) and two boundary steady states ($(\underline{m}_1, \overline{m}_2)$ and $(\overline{m}_1, \underline{m}_2)$). From the figure, it is easy to check that given any initial value (m, ξ), the solution to the ODE eventually converges to one of these steady states. However, this does not imply that the set of steady states is globally attracting; a problem is that in a neighborhood of the origin (the interior steady state), $(\frac{dm_1(t)}{dt}, \frac{dm_2(t)}{dt})$ is approximately $(0,0)$, meaning that the motion of $m(t)$ can be very slow. Accordingly, for some initial value, it takes arbitrarily long time for the solution to reach a neighborhood of the boundary steady state, so we cannot find a uniform bound T appearing in the definition of attracting sets.

Nonetheless, we can show that m^t converge to the boundary steady states. This implies the result we want, as in such a case the actual belief $\tilde{N}(m_i^t, \frac{1}{t\xi_i^t})$ converges to $1_{\underline{\theta}}$ or $1_{\overline{\theta}}$.

Formally, our goal is to prove the following lemma. Let $B = \{(\underline{m}_1, \overline{m}_2, \underline{I}, \overline{I}), (\overline{m}_1, \underline{m}_2, \overline{I}, \underline{I})\}$ denote the set of the boundary steady states. Also, let $M = (\times_{i=1,2}[\underline{m}_i, \overline{m}_i]) \times [\underline{I}, \overline{I}]^2$.

Lemma 7. *Pick a particular path $\boldsymbol{w} : \boldsymbol{R} \to \boldsymbol{R}^4$ such that (i) $\boldsymbol{w}$ is an asymptotic* *pseudotrajectory of the ODE, (ii)* $\lim_{t\to\infty} d(\boldsymbol{w}(t), M) = 0$*, and (iii)* $\lim_{t\to\infty} \boldsymbol{w}(t) \neq p$*. (Note that these properties hold with probability one, as shown by the earlier lemmas.) Then* $\lim_{t\to\infty} d(\boldsymbol{w}, B) = 0$.

Proof. Pick $\boldsymbol{w}$ as stated. Since $\lim_{t\to\infty} \boldsymbol{w}(t) \neq p$, there is $\varepsilon > 0$ such that for any $T > 0$, there is $t > T$ such that $\boldsymbol{w}(t) \notin (\times_{i=1,2}[m_i^* - \varepsilon, m_i^* + \varepsilon]) \times [\underline{I}, \bar{I}]^2$. Pick such ε.

Now, note that the inverse function f_i^{-1} is increasing and $f_i^{-1}(m_i^*) = m_{-i}^*$. Hence we have $f_i^{-1}(m_i^* - \varepsilon) < f_i^{-1}(m_i^*) < f_i^{-1}(m_i^* + \varepsilon)$. Then there is $\eta > 0$ such that

$$f_i^{-1}(m_i^* - \varepsilon) + 2\eta < f_i^{-1}(m_i^*) < f_i^{-1}(m_i^* + \varepsilon) - 2\eta \tag{38}$$

for all i. Pick such $\eta > 0$. Then let $A \subset \boldsymbol{R}^4$ be such that

$$A = \{(m,\xi) \in M \,|\, \min\{m_1 - m_1^*, m_2 - m_2^*\} \leq \eta\} \cap \{(m,\xi) \,|\, \max\{m_1 - m_1^*, m_2 - m_2^*\} \geq -\eta\}.$$

See Figure 9.

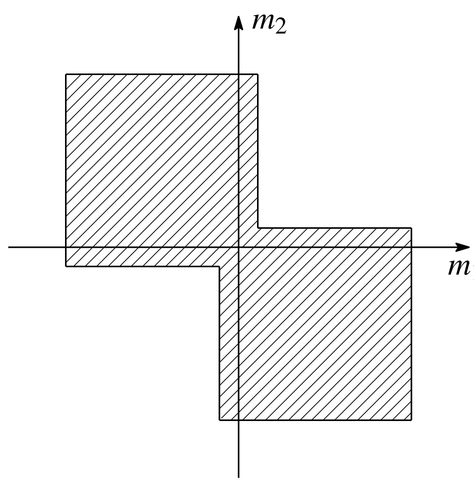

Figure 9: The projection of the set A.

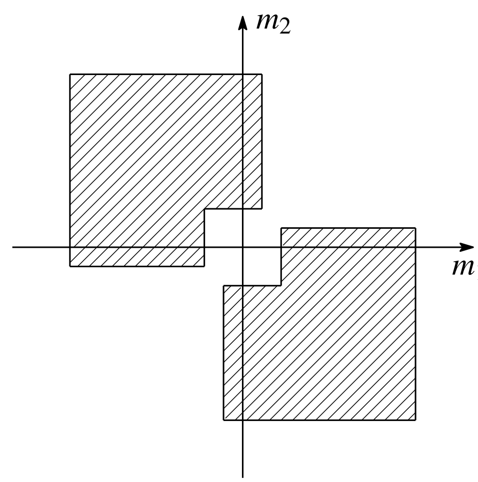

Figure 10: The projection of the set A'.

From Figure 8, given any initial value chosen from the ε-neighborhood of M, the solution to the ODE converges to this set A. Also, the solution does not enter a neighborhood of the origin on the way to a neighborhood of A; this means that the solution reaches a neighborhood of A by some time T, which is independent of the initial value. Thus the set A is attracting, and its basin is the ε-neighborhood of M.

Theorem 6.10 of Benaïm (1999) asserts that if a path $\boldsymbol{w}$ visits the basin W of an attracting set A infinitely often and if W is compact, then $\boldsymbol{w}$ converges to the set A. Since we assume that $\lim_{t\to\infty} d(\boldsymbol{w}(t), M) = 0$, our path $\boldsymbol{w}$ indeed visits the ε-neighborhood of M infinitely often (actually $\boldsymbol{w}$ stays there forever, after a long time). Also ε-neighborhood of M is compact. Hence $\boldsymbol{w}$ converges to the set A, i.e., $\lim_{t\to\infty} d(\boldsymbol{w}(t), A) = 0$. This in particular implies that there is $T > 0$ such that for any $t > T$, $\boldsymbol{w}(t)$ stays in the η-neighborhood of the set A.

At the same time, by the assumption $\boldsymbol{w}$ leaves the set $(\times_{i=1,2}[m_i^* - \varepsilon, m_i^* + \varepsilon]) \times [\underline{I}, \overline{I}]^2$ infinitely often. This means that $\boldsymbol{w}$ visits the set

$$A' = \{(m,\xi) | d((m,\xi), A) \le \eta \text{ and } m \notin \times_{i=1,2}(m_i^* - \varepsilon, m_i^* + \varepsilon)\}$$

infinitely often. See Figure 10.

Note that this set A' is compact and is a basin of the set B of the boundary steady states.[34] Hence again from Theorem 6.10 of Benaïm (1999), $\boldsymbol{w}$ converges to B, as desired. *Q.E.D.*

C.3 Proof of Proposition 4

Let $(x_i(\theta_i), \hat{x}_{-i}(\theta_i))$ denote the one-shot Nash equilibrium $(x_i, \hat{x}_{-i})$ given a state θ_i and a parameter A_i. By the assumption, for the steady state belief $\theta_2 = m_2^*$, $f_1(\theta_2)$ solves $Q(x_1(\theta_1), \hat{x}_2(\theta_1), \theta_1) = Q(x_1(\theta_1), x_2(\theta_2), \theta^*)$. So by the implicit function theorem,

$$\frac{\partial f_1}{\partial \theta_2} = \frac{\frac{\partial Q^*}{\partial x_2}\frac{\partial x_2(\theta_2)}{\partial \theta_2}}{\frac{\partial Q}{\partial x_1}\frac{\partial x_1(\theta_1)}{\partial \theta_1} + \frac{\partial Q}{\partial \hat{x}_2}\frac{\partial \hat{x}_2(\theta_1)}{\partial \theta_1} + \frac{\partial Q}{\partial \theta_1} - \frac{\partial Q^*}{\partial x_1}\frac{\partial x_1(\theta_1)}{\partial \theta_1}},$$

where $Q = Q(x_1(\theta_1), \hat{x}_2(\theta_1), \theta_1)$ and $Q^* = Q(x_1(\theta_1), x_2(\theta_2), \theta^*)$.

[34] To see that A' is a basin of B, pick any point $(m,\xi) \in A'$. If (m,ξ) is in the fourth quadrant, we have $\frac{dm_1(0)}{dt} > 0$ and $\frac{dm_2(0)}{dt} < 0$, i.e., the solution $m(t)$ to the ODE move toward the south-east direction, and eventually converge to the boundary point $(\overline{m}_1, \underline{m}_2)$. See Figure 8. Also the solution does not enter the ε-neighborhood of the origin, so it reaches a neighborhood of the boundary point by some time T which is independent of the initial value. Next, consider the case in which (m,ξ) is in the first quadrant. In this case we have either $m_1 < m_1^* + 2\eta$ or $m_2 < m_2^* + 2\eta$, and without loss of generality, we will focus on the case with $m_2 < m_2^* + 2\eta$. Then from (38), the point (m,ξ) is below the graph of f_1 (the flatter upward-sloping line in Figure 8). Then again we have $\frac{dm_1(0)}{dt} > 0$ and $\frac{dm_2(0)}{dt} < 0$, so that the solution $m(t)$ moves toward the south-east direction and eventually converges to the boundary point $(\overline{m}_1, \underline{m}_2)$. A similar argument applies when (m,ξ) is in the second or the third quadrant. Hence A' is indeed a basin of B.

On the other hand, given (m_1, m_2), the KL minimizer $\theta_1(m_1, m_2)$ solves $Q(x_1(m_1), \hat{x}_2(m_1), \theta_1) = Q(x_1(m_1), x_2(m_2), \theta^*)$. So by the implicit function theorem,

$$\frac{\partial\theta_1}{\partial m_1} = -\frac{\frac{\partial Q}{\partial x_1}\frac{\partial x_1(\theta_1)}{\partial\theta_1} + \frac{\partial Q}{\partial\hat{x}_2}\frac{\partial\hat{x}_2(\theta_1)}{\partial\theta_1} - \frac{\partial Q^*}{\partial x_1}\frac{\partial x_1(\theta_1)}{\partial\theta_1}}{\frac{\partial Q}{\partial\theta_1}}.$$

Similarly,

$$\frac{\partial\theta_1}{\partial m_2} = \frac{\frac{\partial Q^*}{\partial x_2}\frac{\partial x_2(\theta_2)}{\partial\theta_2}}{\frac{\partial Q}{\partial\theta_1}}.$$

Then simple algebra shows that

$$\frac{\partial f_1}{\partial\theta_2} = \frac{\partial\theta_1}{\partial m_2}\frac{1}{1 - \frac{\partial\theta_1}{\partial m_1}} \tag{39}$$

To interpret this equation, suppose that player 2's belief θ_2 increases. Then her optimal action x_2 changes, which influences player 1's belief (KL minimizer) by $\frac{\partial\theta_1}{\partial m_2}$. Since player 1's belief changes, her optimal action changes, which influences her own belief (KL minimizer) by $\frac{\partial\theta_1}{\partial m_2}\frac{\partial\theta_1}{\partial m_1}$. Then again player 1's optimal action changes, which influences her own belief by $\frac{\partial\theta_1}{\partial m_2}(\frac{\partial\theta_1}{\partial m_1})^2$, and so on. The total effect of this process is

$$\frac{\partial\theta_1}{\partial m_2}\left\{1 + \frac{\partial\theta_1}{\partial m_1} + \left(\frac{\partial\theta_1}{\partial m_1}\right)^2 + \cdots\right\},$$

which equals the right-hand side of the above equation.

Now, recall that the eigenvalues of the matrix J' solves

$$\left(\frac{\partial\theta_1}{\partial m_1} - 1 - \lambda\right)\left(\frac{\partial\theta_2}{\partial m_2} - 1 - \lambda\right) - \frac{\partial\theta_1}{\partial m_2}\frac{\partial\theta_2}{\partial m_1} = 0,$$

which is equivalent to

$$\lambda^2 - \left(\frac{\partial\theta_1}{\partial m_1} + \frac{\partial\theta_2}{\partial m_2} - 2\right)\lambda + \left(\frac{\partial\theta_1}{\partial m_1} - 1\right)\left(\frac{\partial\theta_2}{\partial m_2} - 1\right) - \frac{\partial\theta_1}{\partial m_2}\frac{\partial\theta_2}{\partial m_1} = 0. \tag{40}$$

Part (i): $\frac{\partial\theta_i}{\partial m_i} - 1 > 0$ for each i.

Suppose that $\frac{\partial \theta_i}{\partial m_i} - 1 > 0$ for each i. Then $\frac{\partial \theta_1}{\partial m_1} + \frac{\partial \theta_2}{\partial m_2} - 2 > 0$, Hence if (40) have real solutions, at least one of them must be positive, implying linear instability. If (40) have imaginary solutions, then the real part of these solutions is $\frac{1}{2}(\frac{\partial \theta_1}{\partial m_1} + \frac{\partial \theta_2}{\partial m_2} - 2) > 0$, which again implies linear instability.

Part (ii): $\frac{\partial \theta_i}{\partial m_i} - 1 < 0$ for each i.

Assume first that $f_1' f_2' > 1$. Plugging (39) into $f_1' f_2' > 1$, we have

$$\frac{\partial \theta_1}{\partial m_2} \frac{1}{1 - \frac{\partial \theta_1}{\partial m_1}} \frac{\partial \theta_2}{\partial m_1} \frac{1}{1 - \frac{\partial \theta_2}{\partial m_2}} > 1.$$

Since we assume $\frac{\partial \theta_i}{\partial m_i} - 1 < 0$, this inequality is equivalent to

$$\frac{\partial \theta_1}{\partial m_2} \frac{\partial \theta_2}{\partial m_1} > \left(1 - \frac{\partial \theta_1}{\partial m_1}\right) \left(1 - \frac{\partial \theta_2}{\partial m_2}\right)$$

This implies that the y-intercept of the quadratic curve appearing in (40) is negative, which in turn implies that (40) has one positive solution and one negative solution. Hence the steady state is linearly unstable.

Next, consider the case with $f_1' f_2' < 1$. Algebra similar to the one above yields

$$\frac{\partial \theta_1}{\partial m_2} \frac{\partial \theta_2}{\partial m_1} < \left(1 - \frac{\partial \theta_1}{\partial m_1}\right) \left(1 - \frac{\partial \theta_2}{\partial m_2}\right),$$

which means that the y-intercept of the curve appearing in (40) is positive. Since we assume $\frac{\partial \theta_i}{\partial m_i} < 1$, we have $\frac{\partial \theta_1}{\partial m_1} + \frac{\partial \theta_2}{\partial m_2} - 2 < 0$. Hence if (40) have real solutions, they must be negative, implying asymptotic stability. If (40) have imaginary solutions, then the real part of these solutions is $\frac{1}{2}(\frac{\partial \theta_1}{\partial m_1} + \frac{\partial \theta_2}{\partial m_2} - 2) < 0$, which again implies asymptotic stability.

C.4 Proof of Proposition 6

Throughout this proof, we will write f_i instead of f_i^*, in order to emphasize that f_i^* is a function. (Note that when f_i^* is a function, it coincides with the function f_i defined in Proposition 4.)

We use the tools developed in Section A. Recall that under double misspecification, there are two real players and two hypothetical players. Let $x =$

$(x_1, x_2, \hat{x}_1, \hat{x}_2)$ denote the action profile of these players, and given a sample path $h = (x^t, y^t)_{t=1}^{\infty}$, let $\sigma^t(h) \in \triangle(X_1 \times X_2 \times X_1 \times X_2)$ denote the action frequency up to period t. Note that this $\sigma^t(h)$ contains information about the past actions of the real players *and* the hypothetical players.

Proposition 7 shows that after a long time, each player i's posterior belief will be concentrated on the KL minimizer $\theta_i^t = \theta_i(\sigma^t(h))$. Also Proposition 13 shows that the motion of these KL minimizers, (θ_1^t, θ_2^t), is approximated by the differential inclusion (28), which can be rewritten as the two dimensional problem

$$\left(\frac{d\boldsymbol{\theta}_1(t)}{dt}, \frac{d\boldsymbol{\theta}_2(t)}{dt}\right) \in \bigcup_{\sigma:\theta(\sigma)=\theta(t)} \left(-\frac{K_1'(\boldsymbol{\theta}_1(t), s(\boldsymbol{\theta}_1(t), \boldsymbol{\theta}_2(t)))}{K_1''(\boldsymbol{\theta}_2(t), \sigma)}, -\frac{K_2'(\boldsymbol{\theta}_2(t), s(\boldsymbol{\theta}_1(t), \boldsymbol{\theta}_2(t)))}{K_2''(\hat{\boldsymbol{\theta}}_1(t), \sigma)}\right) \tag{41}$$

where $s(\theta_1, \theta_2)$ denotes a static equilibrium $x = (x_1, x_2, \hat{x}_1, \hat{x}_2)$ given the beliefs $(\theta_1, \theta_2, \hat{\theta}_1, \hat{\theta}_2)$ with $\hat{\theta}_1 = \theta_2$ and $\hat{\theta}_2 = \theta_1$.

In what follows, we will show that regardless of the initial value, any solution to the differential inclusion (41) converges to the steady state after a long time. This implies that the steady state is globally attracting in the sense of Esponda, Pouzo, and Yamamoto (2021), and their Proposition 2 ensures that θ^t converges there almost surely, as desired.

The following lemma partially characterizes the solution to the differential inclusion (41): It shows that $\boldsymbol{\theta}_2(t)$ moves toward $f_2(\boldsymbol{\theta}_1(t))$ at any time t.

Lemma 8. *Pick any initial value $\boldsymbol{\theta}(0) = (\boldsymbol{\theta}_1(0), \boldsymbol{\theta}_2(0))$ and any solution $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ to the differential inclusion (41). Then for any $t \geq 0$ with $\boldsymbol{\theta}_2(t) > f_2(\boldsymbol{\theta}_1(t))$, we have $\dot{\boldsymbol{\theta}}_2(t) < 0$. Similarly, for any $t \geq 0$ with $\boldsymbol{\theta}_2(t) < f_2(\boldsymbol{\theta}_1(t))$, we have $\dot{\boldsymbol{\theta}}_2(t) > 0$*

Proof. We will prove only the first part of the lemma, because the proof of the second part is symmetric. Suppose that $\boldsymbol{\theta}_2(t) > f_2(\hat{\theta}_1(t))$ at some time t. To prove $\dot{\boldsymbol{\theta}}_2(t) < 0$, it suffices to show that $K_2'(\boldsymbol{\theta}_2(t), s(t)) > 0$, where $s(t)$ denotes the static equilibrium $s(\boldsymbol{\theta}_1(t), \boldsymbol{\theta}_2(t))$ in time t.

Suppose not and $K_2'(\boldsymbol{\theta}_2(t), s(t)) < 0$. (We ignore the case with $K_2'(\boldsymbol{\theta}_2(t), s(\boldsymbol{\theta}_1(t), \boldsymbol{\theta}_2(t))) = 0$, because in such a case, $\boldsymbol{\theta}_2(t) \in f_2(\boldsymbol{\theta}_1(t))$, which contradicts with the uniqueness of $f_2(\boldsymbol{\theta}_1(t))$.) We consider the following two cases:

Case 1: $\boldsymbol{\theta}_2(t) = \overline{\theta}$. In this case, the KL minimizer given the equilibrium $s(t)$ is $\theta_2(s(t)) = \overline{\theta} = \boldsymbol{\theta}_2(t)$ (this follows from the fact that the KL divergence K_2 is single-peaked w.r.t. θ_2). Hence $\boldsymbol{\theta}_2(t) = \overline{\theta}$ is a steady state, i.e., $\boldsymbol{\theta}_2(t) \in f_2(\boldsymbol{\theta}_1(t))$. But this contradicts with the uniqueness of $f_2(\boldsymbol{\theta}_1(t))$.

Case 2: $\theta_2(t) < \overline{\theta}$. An argument similar to that in Case 1 shows that at $\theta_2 = \overline{\theta}$, we have $K_2'(\overline{\theta}, s(\theta_1(t), \overline{\theta})) > 0$. On the other hand, by the assumption, $K_2'(\theta_2(t), s(\theta_1(t), \theta_2(t))) < 0$. Then since $K_2'(\theta, s(\theta_1(t), \theta))$ is continuous in θ, there must be $\theta \in (\theta_2(t), \overline{\theta})$ such that $K_2'(\theta, s(\theta_1(t), \theta)) = 0$. This implies that $\theta \in f_2(\theta_1)$, but it contradicts with the uniqueness of $f_2(\theta_1)$. *Q.E.D.*

Now we will construct a Lyapunov function V to show that any solution to the differential inclusion (41) converges to the steady state. Without loss of generality, assume that the steady state is $(\theta_1^*, \theta_2^*) = (0,0)$. From assumption (iii), there is $\kappa > 0$ such that $\max_{\theta_1} |\frac{f_2(\theta_1)}{\partial \theta_1}| < \kappa < \frac{1}{\max_{\theta_2} |\frac{f_1(\theta_2)}{\partial \theta_2}|}$. Pick such κ, and for each $\theta = (\theta_1, \theta_2)$, let

$$V(\theta) = \max\left\{|\theta_2|, |\kappa \hat{\theta}_1|\right\}.$$

We will show that given any initial value $\theta(0)$ and given any solution θ to the differential inclusion (27),

$$\dot{V}(\theta(t)) < 0$$

for all t with $\theta(t) \neq (0,0)$. We will consider the following cases separately:

Case 1: $|\theta_2(t)| > |\kappa\theta_1(t)|$. Assume first that $\theta_2(t) > 0$. Then by the definition of κ and $f_2(0) = 0$, we have $f_2(\theta_1(t)) < |\kappa\theta_1(t)| < \theta_2(t)$. Then from Lemma 8 and $\theta_2(t) > 0$, we have $\dot{V}(\theta(t)) = \dot{\theta}_2(t) < 0$.

Assume next that $\theta_2(t) < 0$. By the definition of κ and $f_2(0) = 0$, we have $f_2(\hat{\theta}_1(t)) > -|\kappa\hat{\theta}_1(t)| > \theta_2(t)$. Then from Lemma 8 and $\theta_2(t) < 0$, we have $\dot{V}(\theta(t)) = -\dot{\theta}_2(t) < 0$.

Case 2: $|\theta_2(t)| < |\kappa\theta_1(t)|$. An argument similar to those for Case 1 shows that $\dot{V}(\theta(t)) < 0$.

Case 3: $|\theta_2(t)| = |\kappa\theta_1(t)|$. We will focus on the case with $\theta_2(t) > 0$ and $\theta_1(t) > 0$, because a similar argument applies to all other cases. Then as in the first half of Case 1, we have $\dot{\theta}_2(t) < 0$. Also, a similar argument shows that $\dot{\theta}_1(t) < 0$. Hence we have $\dot{V}(\theta(t)) = \{\dot{\theta}_2(t), \kappa\dot{\theta}_1(t)\} < 0$. *Q.E.D.*

C.5 Proof of Lemma 1

For the case in which X is finite, this is exactly the same as Lemma 1 of Esponda, Pouzo, and Yamamoto (2021). For the case in which X is continuous, we need a minor modification of the proof. We first prove a preliminary lemma:

Lemma 9. *Assume that X is continuous. Under Assumption 1(iii) and (iv), $\int_Y g(x,y)Q(dy|x)$ is bounded and continuous in x.*

Proof. Take a sequence x^n converging to x. Then

$$\begin{aligned}&\int_Y g(x^n,y)Q(dy|x^n) - \int_Y g(x,y)Q(dy|x)\\ &\leq \left|\int_Y g(x^n,y)Q(dy|x^n) - \int_Y g(x^n,y)Q(dy|x)\right|\\ &\quad + \left|\int_Y g(x^n,y)Q(dy|x) - \int_Y g(x,y)Q(dy|x)\right|.\end{aligned}$$

From Assumption 1(iii), $Q(dy|x^n)$ weakly converges to $Q(dy|x)$, so the first term of the right-hand side converges to zero. Also from Assumption 1(iv-a), $g(x^n,y)$ pointwise converges to $g(x,y)$, so the second term converges to zero. *Q.E.D.*

As shown in the display in EPY, we have

$$\begin{aligned}K_{i,k}(\theta^n_{i,k},\sigma^n) - K_i(\theta^n_{i,k},\sigma) \leq &\int_X\int_Y g(x,y)Q(dy|x)\sigma^n_{\hat{X}_{1,1}\times\hat{X}_{2,1}}(dx)\\ &- \int_X\int_Y g(x,y)Q(dy|x)\sigma_{\hat{X}_{1,1}\times\hat{X}_{2,1}}(dx)\end{aligned}$$

where $\sigma_{\hat{X}_{1,1}\times\hat{X}_{2,1}}$ and $\sigma^n_{\hat{X}_{1,1}\times\hat{X}_{2,1}}$ are the marginals of σ and σ^n on $\hat{X}_{1,1}\times\hat{X}_{2,1}$, respectively. From Lemma 9, the right-hand side converges to zero as $\sigma^n \to \sigma$. The rest of the proof is exactly the same as in EPY. *Q.E.D.*

C.6 Proof of Proposition 7

For the special case in which X is finite, Theorem 1 of Esponda, Pouzo, and Yamamoto (2021) proves the same result. We need a minor modification to their proof, as they use finiteness of X in Step 2 in the proof of Lemma 2.

Pick i, k, $\theta_{i,k}$. Then let

$$f_l(\hat{x}) = E_{Q(\cdot|\hat{x}_{1,1},\hat{x}_{2,1})}\left[\sup_{\theta'_{i,k}\in O(\theta_{i,k},\frac{1}{l})}\left|\frac{q(y|\hat{x}_{1,1},\hat{x}_{2,1})}{q_{\theta_{i,k}}(y|\hat{x}_{i,k},\hat{x}_{i,k+1})} - \frac{q(y|\hat{x}_{1,1},\hat{x}_{2,1})}{q_{\theta'_{i,k}}(y|\hat{x}_{i,k},\hat{x}_{i,k+1})}\right|\right]$$

where $O(\theta_{i,k},\frac{1}{l})$ is a $\frac{1}{l}$-neighborhood of $\theta_{i,k}$. Then as explained at the end of the the first paragraph in EPY's step 2, $\lim_{l\to\infty} f_l(\hat{x}) \to 0$ for each $\hat{x}$. In what follows, we will show that this convergence is uniform in $\hat{x}$; then there is $\delta(\theta_{i,k},\varepsilon)$ with which (16) of EPY holds, and the rest of the proof is exactly the same as EPY's.

Pick an arbitrary $\varepsilon > 0$. For each $\hat{x}$, let $F(\hat{x}) = \{l \in [0,\infty) | f_l(\hat{x}) \geq \varepsilon\}$. Then we have the following lemma:

Lemma 10. *For each $\hat{x}$, there is $l(\hat{x}) > 0$ such that $F(\hat{x}) = [0, l(\hat{x})]$. Also $F(\hat{x})$ is upper hemi-continuous in $\hat{x}$.*

Proof. The first part follows from the fact that $f_l(\hat{x})$ is continuous and decreasing in l, and $\lim_{l\to\infty} f_l(\hat{x}) = 0$.

To prove the second part, pick $\hat{x}$ and an arbitrary small $\eta > 0$. Then $f_{l(\hat{x})+\eta}(\hat{x}) < \varepsilon$. Since $f_l(\hat{x})$ is continuous in $\hat{x}$, there is an open neighborhood U of $\hat{x}$ such that $f_{l(\hat{x})+\eta}(\hat{x}') < \varepsilon$ for all $\hat{x}' \in U$. This implies that $l(\hat{x}') < l(\hat{x}) + \eta$ for all $\hat{x}' \in U$. *Q.E.D.*

The above lemma implies that $l(\hat{x})$ is an upper hemi-continuous function, and from the Maximum theorem, $l(\hat{x})$ is bounded; $l(\hat{x}) < l^*$ for some l^*. Hence $f_l(\hat{x}) \leq \varepsilon$ for all $\hat{x}$ and $l \geq l^*$, implying uniform convergence. *Q.E.D.*

C.7 Proof of Proposition 9

This is very similar to the first step of the proof of Proposition 2 in EPY. However, we need a minor modification, as X may not be finite in our setup. We first prove upper hemi-continuity of $B_\varepsilon(\sigma)$.

Lemma 11. *$B_\varepsilon(\sigma)$ is upper hemi-continuous in (ε, σ).*

Proof. Since $\prod_{i=1}^{2}\prod_{k=1}^{k_i+1} \triangle\Theta_{i,k}$ is compact, it is sufficient to show that $(\varepsilon^n, \sigma^n, \hat{\mu}^n) \to (\varepsilon, \sigma, \hat{\mu})$ and $\hat{\mu}^n \in B_{\varepsilon^n}(\sigma^n)$ for each n imply $\hat{\mu} \in B_\varepsilon(\sigma)$. Note that

$$\begin{aligned}
&\lim_{n\to\infty}\left(\int_{\Theta_{i,k}} (K_{i,k}(\theta_{i,k}, \sigma^n)\hat{\mu}^n_{i,k}(d\theta_{i,k}) - \int_{\Theta_{i,k}} (K_{i,k}(\theta_{i,k}, \sigma)\hat{\mu}_{i,k}(d\theta_{i,k})\right)\\
&\quad= \lim_{n\to\infty}\left(\int_{\Theta_{i,k}} (K_{i,k}(\theta_{i,k}, \sigma^n)\hat{\mu}^n_{i,k}(d\theta_{i,k}) - \int_{\Theta_{i,k}} (K_{i,k}(\theta_{i,k}, \sigma)\hat{\mu}^n_{i,k}(d\theta_{i,k})\right)\\
&\qquad+ \lim_{n\to\infty}\left(\int_{\Theta_{i,k}} (K_{i,k}(\theta_{i,k}, \sigma)\hat{\mu}^n_{i,k}(d\theta_{i,k}) - \int_{\Theta_{i,k}} (K_{i,k}(\theta_{i,k}, \sigma)\hat{\mu}_{i,k}(d\theta_{i,k})\right).
\end{aligned}$$

The first term of the right-hand side is zero, because $K_{i,k}(\cdot, \sigma^n)$ pointwise converges to $K_{i,k}(\cdot, \sigma)$ (which follows from the fact that σ^n weakly converges to σ). Also the second term of the right-hand side is zero, as $\mu^n_{i,k}$ weakly converges to $\mu_{i,k}$.

$$\lim_{n\to\infty}\int_{\Theta_{i,k}} (K_{i,k}(\theta_{i,k}, \sigma^n)\hat{\mu}^n_{i,k}(d\theta_{i,k}) = \int_{\Theta_{i,k}} (K_{i,k}(\theta_{i,k}, \sigma)\hat{\mu}_{i,k}(d\theta_{i,k}).$$

Since $\hat{\mu}^n \in B_{\varepsilon^n}(\sigma^n)$,

$$\int_{\Theta_{i,k}} (K_{i,k}(\theta_{i,k}, \sigma^n) - K^*_{i,k}(\sigma^n))\hat{\mu}^n_{i,k}(d\theta_{i,k}) \leq \varepsilon^n.$$

Taking $n \to \infty$ and using continuity of $K^*_{i,k}(\sigma)$ (which follows from the theory of maximum),

$$\int_{\Theta_{i,k}} (K_{i,k}(\theta_{i,k}, \sigma) - K^*_{i,k}(\sigma))\hat{\mu}_{i,k}(d\theta_{i,k}) \leq \varepsilon.$$

Hence $\mu \in B_\varepsilon(\sigma)$, which implies upper hemi-continuity of $B_\varepsilon(\sigma)$. *Q.E.D.*

Now we show that $S_\varepsilon(\sigma)$ is upper hemi-continuous at $\varepsilon = 0$. Since X is compact, it suffices to show that $(\varepsilon^n, \sigma^n, x^n) \to (0, \sigma, x)$ and $x^n \in S_{\varepsilon^n}(\sigma^n)$ for each n, imply $x \in S_\varepsilon(\sigma)$. As noted earlier, we already know that $S_0(\sigma)$ is upper hemi-continuous in σ. So without loss of generality, we assume $\varepsilon^n > 0$ for all n.

Since $x^n \in S_{\varepsilon^n}(\sigma^n)$, there is $\hat{\mu}^n \in B_{\varepsilon^n}(\sigma^n)$ with $x^n = \hat{s}(\hat{\mu}^n)$. The sequence $(\varepsilon^n, \sigma^n, x^n, \hat{\mu}^n)$ is in a compact set, so there is a convergent subsequence, still denoted by $(\varepsilon^n, \sigma^n, x^n, \hat{\mu}^n)$. Let $\hat{\mu} = \lim_{n\to\infty} \hat{\mu}^n$. Then $\hat{\mu} \in B_0(\sigma)$, as $B_\varepsilon(\sigma)$ is upper hemi-continuous and $\hat{\mu}^n \in B_{\varepsilon^n}(\sigma^n)$ for each n. Also, we have $x \in \hat{S}(\hat{\mu})$, because $\hat{S}$ is upper hemi-continuous and $x^n \in \hat{S}(\hat{\mu}^n)$ for each n. Hence $x \in S_0(\sigma)$. *Q.E.D.*

C.8 Proof of Proposition 10

The proof is very similar to that of Theorem 2 of EPY. (Their statement of the theorem is incorrect, and we need to take the infimum over the set of all solutions of the differential inclusion, rather than the solutions for some initial value. See Esponda, Pouzo, and Yamamoto (2022) for details.) In EPY, the proof consists of three steps. In the first two steps, they show that $\boldsymbol{w}$ is a perturbed solution of the differential inclusion. Then in the last step, they show that a perturbed solution is an asymptotic pseudotrajectory (i.e., it satisfies (26)).

Our Propositions 8 and 9 imply that $\boldsymbol{w}$ is indeed a perturbed solution in the sense of EPY. We can also show that a perturbed solution is indeed an asymptotic pseudotrajectory. The proof is omitted because, other than replacing the Euclidean norm with the dual bounded-Lipschitz norm, it is exactly the same as the last step of EPY.[35] *Q.E.D.*

[35] This parallels Perkins and Leslie (2014), who show that the stochastic approximation technique of Benaïm (1999) for the Euclidean space extends to Banach spaces with the same proof.

C.9 Proof of Lemma 2

We will show that $\theta(\sigma)$ is Lipschitz continuous in σ. Under Assumptions 4(i) and (iii), the inverse $(\nabla^2 K_{i,k}(\theta_{i,k}(\sigma),\sigma))^{-1}$ of the Hessian matrix exists for each σ, and is continuous in σ. This means that $\|(\nabla^2 K_{i,k}(\theta_{i,k}(\sigma),\sigma))^{-1}\|$ is bounded and continuous in σ, where $\|C\| = \max_{ij}|c_{ij}|$ denotes the max norm of a matrix $C = \{c_{ij}\}$, . Since $\triangle\hat{X}$ is compact, there is L_1 such that $\|(\nabla^2 K_{i,k}(\theta_{i,k}(\sigma),\sigma))^{-1}\| < L_1$ for all i, k, and σ. Pick such L_1.

Under Assumption 4(ii), there is $L_2 > 1$ such that

$$\left|\frac{\partial K_{i,k}(\theta_{i,k},\hat{x})}{\partial\theta_{i,k,m}} - \frac{\partial K(\theta_{i,k},\hat{x}')}{\partial\theta_{i,k,m}}\right| < L_2|\hat{x}-\hat{x}'|$$

for all i, k, m, $\theta_{i,k}$, $\hat{x}$, and $\hat{x}'$. Also, under Assumption 4(i), there is $L_3 > 1$ such that

$$\left|\frac{\partial K_{i,k}(\theta_{i,k},\hat{x})}{\partial\theta_{i,k,m}}\right| < L_3$$

for all i, k, m, $\theta_{i,k}$, and $\hat{x}$. Then for each σ and σ', we have

$$\begin{aligned}&\left|\frac{\partial K_{i,k}(\theta_{i,k},\sigma)}{\partial\theta_{i,k,m}} - \frac{\partial K_{i,k}(\theta_{i,k},\sigma')}{\partial\theta_{i,k,m}}\right| \\ &= \left|\int \frac{\partial K_{i,k}(\theta_{i,k},\hat{x})}{\partial\theta_{i,k,m}}\sigma(d\hat{x}) - \int \frac{\partial K_{i,k}(\theta_{i,k},\hat{x})}{\partial\theta_{i,k,m}}\sigma'(d\hat{x})\right| \leq 4L_2L_3\|\sigma-\sigma'\|\end{aligned}$$

where the inequality follows from the definition of the dual bounded-Lipschitz norm and the fact that $\frac{1}{4L_2L_3}\frac{\partial K_{i,k}(\theta_{i,k},\hat{x})}{\partial\theta_{i,k,m}} \in BL(\hat{X})$. This in turn implies that $\nabla K_{i,k}(\theta_{i,k},\sigma)$ is equi-Lipschitz continuous, that is, there is $L_4 > 0$ such that $|\nabla K_{i,k}(\theta_{i,k},\sigma) - \nabla K_{i,k}(\theta_{i,k},\sigma')| < L_4\|\sigma-\sigma'\|$ for all i, k, $\theta_{i,k}$, σ, and σ'.

Let $L = L_1L_4$. We will show that $\theta(\sigma)$ is Lipschitz continuous with the constant L. To do so, pick two action frequencies σ and $\sigma' \neq \sigma$ arbitrarily. For each $\beta \in [0,1]$, let $\sigma_\beta = \beta\sigma + (1-\beta)\sigma'$ denote a convex combination of σ and σ'. From Assumption 4(iii), the KL minimizer $\theta_{i,k}(\sigma_\beta)$ must solve the first-order condition

$$\nabla K_{i,k}(\theta_{i,k},\sigma_\beta) = 0,$$

Our result differs from Perkins and Leslie (2014) in that we consider a differential inclusion, rather than a differential equation. But this does not cause any technical difficulty, because (i) $\triangle\hat{X}$ is a compact subset of a banach space with the dual bounded Lipschitz norm and (ii) Mazur's lemma, which is used to establish the result for differential inclusions in Euclidean spaces (Benaïm, Hofbauer, and Sorin (2005) and Esponda, Pouzo, and Yamamoto (2021)), is valid even in Banach spaces.

which is equivalent to

$$\beta \nabla K_{i,k}(\theta_{i,k}, \sigma) + (1-\beta)\nabla K_{i,k}(\theta_{i,k}, \sigma') = 0.$$

Then by the implicit function theorem,

$$\frac{d\theta(\sigma_\beta)}{d\beta} = -(\nabla^2 K_{i,k}(\theta_{i,k}(\sigma_\beta), \sigma_\beta))^{-1} \left(\nabla K_{i,k}(\theta_{i,k}(\sigma_\beta), \sigma) - \nabla K_{i,k}(\theta_{i,k}(\sigma_\beta), \sigma')\right). \tag{42}$$

Using the fundamental theorem of calculus, we have

$$\begin{aligned}
&\theta(\sigma) - \theta(\sigma') \\
&= \theta(\sigma_1) - \theta(\sigma_0) \\
&= -\int_0^1 (\nabla^2 K_{i,k}(\theta_{i,k}(\sigma_\beta), \sigma_\beta))^{-1} \left(\nabla K_{i,k}(\theta_{i,k}(\sigma_\beta), \sigma) - \nabla K_{i,k}(\theta_{i,k}(\sigma_\beta), \sigma')\right) d\beta.
\end{aligned}$$

Then by the definition of L_1 and L_4,

$$|\theta(\sigma) - \theta(\tilde{\sigma})| \le \int_0^1 L_1 L_4 \|\sigma - \sigma'\| d\beta = L\|\sigma - \sigma'\|. \quad \textit{Q.E.D.}$$

C.10 Proof of Proposition 13

We will first present a preliminary lemma. Pick an arbitrary action frequency $\boldsymbol{\sigma}(0) \in \triangle \hat{X}$ and a solution $\boldsymbol{\sigma} \in Z(\boldsymbol{\sigma}(0))$ to the differential inclusion (26) starting from this $\boldsymbol{\sigma}(0)$. Let $\boldsymbol{\theta}(t) = \theta(\boldsymbol{\sigma}(t))$ for each t. The following lemma shows that $\{\boldsymbol{\theta}(t)\}_{t \ge 0}$ solves (27).

Lemma 12. *Pick $t \ge 0$ such that (26) holds. Then $\dot{\boldsymbol{\theta}}(t)$ exists and satisfies (27).*

Proof. Pick t as stated, and pick $\sigma^* \in \triangle S_0(\boldsymbol{\sigma}(t))$ such that $\dot{\boldsymbol{\sigma}}(t) = \sigma^* - \boldsymbol{\sigma}(t)$. Let $\sigma_\beta = \beta \sigma^* + (1-\beta)\boldsymbol{\sigma}(t)$ for each $\beta \in [0,1]$. Then we have

$$\frac{\theta(\boldsymbol{\sigma}(t+\varepsilon)) - \theta(\boldsymbol{\sigma}(t))}{\varepsilon} = \left(\frac{\theta(\sigma_\varepsilon) - \theta(\sigma_0)}{\varepsilon} + \frac{\theta(\boldsymbol{\sigma}(t+\varepsilon)) - \theta(\sigma_\varepsilon)}{\varepsilon} \right).$$

All we need to show is that the right-hand side has a limit as $\varepsilon \to 0$, and the limit is in the right-hand side of (27). Then $\frac{\theta(\boldsymbol{\sigma}(t+\varepsilon)) - \theta(\boldsymbol{\sigma}(t))}{\varepsilon}$ also has a limit $\dot{\boldsymbol{\theta}}(t)$ and this limit value satisfies (27).

Note first that $\lim_{\varepsilon\to 0}\frac{\theta(\sigma_\varepsilon)-\theta(\sigma_0)}{\varepsilon}$ exists and is in the right-hand side of (27). Indeed, from (42),

$$\begin{aligned}&\lim_{\varepsilon\to 0}\frac{\theta(\sigma_\varepsilon)-\theta(\sigma_0)}{\varepsilon}=\left.\frac{d\theta(\sigma_\beta)}{d\beta}\right|_{\beta=0}\\&=-(\nabla^2 K_{i,k}(\theta_{i,k}(\sigma_0),\sigma_0))^{-1}\left(\nabla K_{i,k}(\theta_{i,k}(\sigma_0),\sigma_1)-\nabla K_{i,k}(\theta_{i,k}(\sigma_0),\sigma_0)\right)\\&=-(\nabla^2 K_{i,k}(\theta_{i,k}(\boldsymbol{\sigma}(t)),\boldsymbol{\sigma}(t)))^{-1}\left(\nabla K_{i,k}(\theta_{i,k}(\boldsymbol{\sigma}(t)),\boldsymbol{\sigma}^*)\right)\end{aligned}$$

where the second equality follows from the fact that $\theta_{i,k}(\sigma_0)$ solves the first-order condition.

We conclude the proof by showing that $\lim_{\varepsilon\to 0}\frac{\theta(\boldsymbol{\sigma}(t+\varepsilon))-\theta(\sigma_\varepsilon)}{\varepsilon}=0$. Since $\theta(\sigma)$ is Lipschitz continuous, there is $L>0$ such that

$$\begin{aligned}\left|\frac{\theta(\boldsymbol{\sigma}(t+\varepsilon))-\theta(\sigma_\varepsilon)}{\varepsilon}\right|&\le L\left\|\frac{\boldsymbol{\sigma}(t+\varepsilon))-\sigma_\varepsilon}{\varepsilon}\right\|\\&=L\left\|\frac{(\boldsymbol{\sigma}(t+\varepsilon))-\boldsymbol{\sigma}(t))-(\sigma_\varepsilon-\sigma_0)}{\varepsilon}\right\|.\end{aligned}$$

Taking $\varepsilon\to 0$,

$$\begin{aligned}\lim_{\varepsilon\to 0}\left|\frac{\theta(\boldsymbol{\sigma}(t+\varepsilon))-\theta(\sigma_\varepsilon)}{\varepsilon}\right|&=L\left\|\lim_{\varepsilon\to 0}\frac{\boldsymbol{\sigma}(t+\varepsilon)-\boldsymbol{\sigma}(t)}{\varepsilon}-\lim_{\varepsilon\to 0}\frac{\sigma_\varepsilon-\sigma_0}{\varepsilon}\right\|\\&=L\left\|\frac{d\boldsymbol{\sigma}(t)}{dt}-\left.\frac{d\sigma_\beta}{d\beta}\right|_{\beta=0}\right\|=0\end{aligned}$$

Q.E.D.

Now we prove the proposition. Pick $T>0$ and $h\in\mathscr{H}$ arbitrary. Pick any small $\varepsilon>0$. Since $\theta(\sigma)$ is uniformly continuous in σ (this follows from Lipschitz continuity of θ), there is $\eta>0$ such that $|\theta(\sigma)-\theta(\tilde{\sigma})|<\varepsilon$ for any σ and $\tilde{\sigma}$ with $\|\sigma-\tilde{\sigma}\|<\eta$. From Proposition 10, there is t^* such that for any $t>t^*$, there is $\boldsymbol{\sigma}\in Z$ (for the special case with $S_0(\triangle\theta(\sigma))$ is a singleton for all σ, $\boldsymbol{\sigma}\in Z(\boldsymbol{w}(h)[t])$) such that

$$\|\boldsymbol{w}(h)[t+\tau]-\boldsymbol{\sigma}(\tau)\|<\eta$$

for all $\tau\in[0,T]$. Pick such $\boldsymbol{\sigma}$, and consider the corresponding $\boldsymbol{\theta}$, i.e., let $\boldsymbol{\theta}(t)=\theta(\boldsymbol{\sigma}(t))$ for each t. Then by the definition of η, we have

$$\|\boldsymbol{w}_\theta(h)[t+\tau]-\boldsymbol{\theta}(\tau)\|<\varepsilon$$

for all $\tau \in [0,T]$. Also this $\boldsymbol{\theta}$ solves (27).[36] This implies the result we want. *Q.E.D.*

C.11 Proof of Proposition 14

Let $\mathscr{N}$ be a neighborhood of p, and choose a function $\eta : \mathscr{N} \to \boldsymbol{R}_+$ as in Section 3 of Pemantle (1990), given the ODE (30). Roughly, $\eta(v)$ measures the distance between a point v and (the set of) the paths pointing to the steady state p. For example, any point v with $\eta(v) = 0$ is on such a path, so starting from this point v, a solution to the ODE (30) converges to p.

On the other hand, any point v with $\eta(v) > 0$ is not on such a path. So the solution to the ODE does not converge to p. Indeed, as shown by Proposition 3(v) of Pemantle (1990), we have $D_v(\eta)(F(v)) > 0$ for any v with $\eta(v) > 0$. So a solution to the ODE moves away from the paths converging to p. (Here, the notation for multidimensional derivatives uses $D_v(\eta)$ for the differential of η at a point v.)

Let $S_t = \eta(v(t))$ and $X_t = S_t - S_{t-1}$. Lemma 1 of Pemantle (1990) shows that after every history $\mathscr{F}_t$, the stochastic process $\{S_k\}$ can exceed $\frac{c^*}{\sqrt{t}}$ (i.e., $v(t)$ leaves a neighborhood of the paths converging to p) at some point in the future with probability at least 0.5. The following lemma shows that the same result holds in our setup. The proof can be found in Section C.12

Lemma 13. *There is a constant $c^* > 0$ and t^* such that for any $t > t^*$ and $\mathscr{F}_t$,*

$$\Pr\left(\sup_{k>t} S_k > \frac{c^*}{\sqrt{t}} \text{ or } v(k) \notin \mathscr{N} \text{ for some } k > t \,\middle|\, \mathscr{F}_t\right) > 0.5.$$

Lemma 2 of Pemantle (1990) shows that once the process $\{v(t)\}$ leaves a $\frac{c^*}{\sqrt{t}}$-neighborhood of p as stated in the lemma above, then it fails to return to p with positive probability. The proof can be found in Section C.13

Lemma 14. *Let $c^* > 0$ be as in Lemma 13. Then there is $a > 0$ such that*

$$\Pr\left(\inf_{k>t} S_k > \frac{c^*}{2\sqrt{t}} \text{ or } v(k) \notin \mathscr{N} \text{ for some } k \geq t \,\middle|\, \mathscr{F}_t, S_t \geq \frac{c^*}{\sqrt{t}}\right) \geq a.$$

[36]Note that $\boldsymbol{\theta}$ is absolutely continuous because $\boldsymbol{\sigma}$ is absolutely continuous and $\theta(\sigma)$ is Lipschitz continuous. Also from Lemma 12, $\boldsymbol{\theta}$ satisfies (27) for almost all t.

The rest of the proof is exactly the same as the argument in the full paragraph on page 711 of Pemantle (1990): Suppose that $\Pr(v(t) \to p) > 0$. Then there is some history $\mathscr{F}_t$ after which the probability that $v(M)$ converges to p and never leaves the neighborhood $\mathscr{N}$ is at least $1 - \frac{a}{2}$. However, Lemmas 13 and 14 imply that the probability that $v(M)$ fails to converge to p or leaves $\mathscr{N}$ is at least $\frac{a}{2}$ conditional on any history $\mathscr{F}_t$. This is a contradiction.

C.12 Proof of Lemma 13

Without loss of generality, assume that $\mathscr{N}$ (the domain of the "distance function" η) is a closed ball surrounding p. (This is so because given a neighborhood U of the point p, we can always find a closed ball $\mathscr{N} \subseteq U$ containing p.) Then enlarge the domain of η by letting $\eta(v) = \eta(\arg\max_{\tilde{v} \in \mathscr{N}} d(\tilde{v}, \mathscr{N}))$ for each $v \notin \mathscr{N}$. Here $d(v, \mathscr{N})$ measures the Euclidean distance between v and the ball $\mathscr{N}$. This function η is well-defined because $\mathscr{N}$ is a closed ball. Since η is Lipschitz in $\mathscr{N}$, it is so in the entire space $\boldsymbol{R}^n$.

Pick a sufficiently large t, and define a stopping time $\tau = \{M \geq t | S_M > \frac{c^*}{\sqrt{t}}\}$. We will show that $\Pr(\tau = \infty | \mathscr{F}_t) < 0.5$.

Step 1: Inequalities (12) and (14) of Pemantle (1990).

In the proof Pemantle (1990), he shows that there is $k_2 > 0$ such that for any $M > t$ with $S_M \leq \frac{c^*}{\sqrt{t}}$,

$$E[2X_{M+1}S_M | \mathscr{F}_M] \geq \frac{k_2 S_M^2}{M+1} + O\left(\frac{S_M}{M^2}\right), \tag{43}$$

$$E[X_{M+1}^2 | \mathscr{F}_M] \text{ is at least } \frac{\text{const.}}{M^2}. \tag{44}$$

See (12) and (14) of Pemantle (1990). His proof relies on the assumption that the noise term has a bounded support (and hence the step size is of order $\frac{1}{t+1}$). We will show that the same result holds in our setup where the noise is Gaussian.

Note that for any $v, \tilde{v} \in \boldsymbol{R}^n$ and sufficiently large M,

$$\begin{aligned}
& E\left[\eta\left(v+\frac{b(M,\tilde{v})\varepsilon}{M+1}\right)\middle|\mathscr{F}_M\right] \\
&= E[\eta(v+zb(M,\tilde{v})\varepsilon)], \quad \text{where } z=\tfrac{1}{M+1} \\
&= \eta(v)+\left.\frac{\partial E[\eta(v+zb(M,\tilde{v})\varepsilon)]}{\partial z}\right|_{z=0} z+O(z^2) \\
&= \eta(v)+\sum_{i=1}^{n}\frac{\partial\eta(v)}{\partial v_i}b_i(M,\tilde{v})E[\varepsilon]z+O(z^2) \\
&= \eta(v)+O(z^2) \\
&= \eta(v)+O\left(\frac{1}{M^2}\right)
\end{aligned}$$

To obtain the second equation, we regard the whole term as a function of z and apply Taylor expansion at $z=0$. Intuitively, this shows that the impact of the noise ε in period M on the expected value of $\eta(v(M+1))$ is of order $O(\frac{1}{M^2})$. Then we have

$$\begin{aligned}
& E[S_{M+1}|\mathscr{F}_M] \\
&= E\left[\eta\left(v(M)+\frac{1}{M+1}\left(\tilde{F}(t,v(M))+b(M,v(M))\varepsilon\right)\right)\middle|\mathscr{F}_M\right] \\
&= \eta\left(v(M)+\frac{\tilde{F}(t,v(M))}{M+1}\right)+O\left(\frac{1}{M^2}\right) \\
&= \eta\left(v(M)+\frac{F(v(M))}{M+1}\right)+O\left(\frac{1}{M^2}\right) \\
&\geq \frac{k_2 S_M}{M+1}+O\left(\frac{1}{M^2}\right),
\end{aligned}$$

which immediately implies (43). Here the third equation follows from the Lipschitz continuity of η, and $|F(v)-\tilde{F}(M,v)|<\frac{K}{M}$. The last inequality follows from Proposition 3(iv) of Pemantle (1990).

To obtain (44), note that

$$\begin{aligned}
E[X_{M+1}^2|\mathscr{F}_M] &= (E[X_{M+1}^+|\mathscr{F}_M])^2 \\
&\geq (\Pr(|\varepsilon(M)|<1|\mathscr{F}_M)E[X_{M+1}^+|\mathscr{F}_M,|\varepsilon(M)|<1])^2.
\end{aligned}$$

Conditional on $|\varepsilon(M)| < 1$, the step size $v(M+1) - v(M)$ is of order $\frac{1}{M+1}$. Hence as in the first display on page 709 of Pemantle (1990), we have

$$
\begin{aligned}
& E[X_{M+1}^{+}|\mathscr{F}_M, |\varepsilon(M)| < 1] \\
& \geq E\left[\left(D_{v(M)}(\eta)\left(\frac{\tilde{F}(M,v(M)) + b(M,v(M))\varepsilon}{M+1}\right) + O(|v(M+1) - v(M)|^2)\right)^{+}\middle|\mathscr{F}_M, |\varepsilon(M)| < 1\right] \\
& = E\left[\left(D_{v(M)}(\eta)\left(\frac{F(v(M)) + b(M,v(M))\varepsilon}{M+1}\right) + O\left(\frac{1}{M^2}\right)\right)^{+}\middle|\mathscr{F}_M, |\varepsilon(M)| < 1\right] \\
& \geq E\left[\left(D_{v(M)}(\eta)\left(\frac{b(M,v(M))\varepsilon}{M+1}\right) + O\left(\frac{1}{M^2}\right)\right)^{+}\middle|\mathscr{F}_M, |\varepsilon(M)| < 1\right] \\
& \geq \frac{\text{const.}}{M+1} + O\left(\frac{1}{M^2}\right)
\end{aligned}
$$

Here the equality follows from linearity of $D_v(\eta)$, $|F(v) - \tilde{F}(M,v)| < \frac{K}{M}$, and the fact that the step size $v(M+1) - v(M)$ is of order $\frac{1}{M+1}$. The second to the last inequality follows from Proposition 3(v) of Pemantle (1990). The last inequality uses the fact that the gradient of η at p is $c'h$ for some $c' > 0$ and $h \in H^*$, which implies $D_v(\eta)(b(M,v(M))) \geq c'\kappa$ for any $v(M)$ in a neighborhood of p.

Substituting this inequality to the previous one, we obtain (44).

Step 2: Main Proof.

As argued in the full paragraph on page 709 of Pemantle (1990), combining (43) and (44) yields

$$E[2X_{M+1}S_M + X_{M+1}^2|\mathscr{F}_M] \geq \frac{\text{const.}}{M^2},$$

which in turn implies

$$
\begin{aligned}
E[S_{\tau\wedge(M+1)}^2|\mathscr{F}_t] - E[S_{\tau\wedge M}^2|\mathscr{F}_t] &= E[1_{\tau>M}(2X_{M+1}S_M + X_{M+1}^2)|\mathscr{F}_t] \\
&= E[E[1_{\tau>M}(2X_{M+1}S_M + X_{M+1}^2)|\mathscr{F}_M]|\mathscr{F}_t] \\
&\geq \frac{\text{const.}}{M^2} E[1_{\tau>M}|\mathscr{F}_t] \\
&\geq \frac{\text{const.}}{M^2} \Pr(\tau = \infty|\mathscr{F}_t)
\end{aligned}
$$

for $M > t$. Pemantle (1990) applies this inequality iteratively and obtains

$$E[S^2_{\tau\wedge M}|\mathscr{F}_t] \geq S_t^2 + \text{const.} \cdot \Pr(\tau = \infty|\mathscr{F}_t) \sum_{i=t}^{M-1} \frac{1}{i^2}$$
$$\geq \text{const.} \cdot \Pr(\tau = \infty|\mathscr{F}_t) \left(\frac{1}{t} - \frac{1}{M}\right). \tag{45}$$

Then in the first paragraph on page 710, Pemantle (1990) shows that

$$\frac{4(c^*)^2}{t} \geq E(S^2_{M\wedge\tau}|\mathscr{F}_t), \tag{46}$$

using the assumption that the noise has a bounded support (which ensures that the step size is of order $\frac{1}{M+1}$). We can show that the same result holds in our setup, the proof can be found at the end.

Then the rest of the proof is the same as Pemantle (1990): Combining (45) and (46),

$$\frac{4(c^*)^2}{t} \geq \text{const.} \cdot \Pr(\tau = \infty|\mathscr{F}_t) \left(\frac{1}{t} - \frac{1}{M}\right).$$

This inequality holds for all M, and when $M \to \infty$, it reduces to

$$4(c^*)^2 \geq \text{const.} \cdot \Pr(\tau = \infty|\mathscr{F}_t).$$

By taking c^* small enough, we have $\Pr(\tau = \infty|\mathscr{F}_t) \leq 0.5$, as desired.

Step 3: Proof of (46).

We will show that (46) holds in our setup: Let $L > 0$ be the Lipschitz constant of η, and $\hat{c} > 0$ be such that $|\tilde{F}(M,v) - v| < \hat{c}$ for all $M > t$ and for all v in a neighborhood of p. Then

$$|S_{M\wedge\tau} - S_{(M\wedge\tau)-1})| < L|v(M\wedge\tau) - v((M\wedge\tau)-1)|$$
$$\leq \frac{\hat{c}L + n\bar{b}|\varepsilon|L}{M\wedge\tau}$$
$$\leq \frac{\hat{c}L + n\bar{b}|\varepsilon|L}{t}. \tag{47}$$

whenever $v(M)$ is in the neighborhood of p. Since the mean of the half-normal distribution is $\frac{\sqrt{2}}{\sqrt{\pi}}$ and its variance is $1-\frac{2}{\pi}$, we have

$$E[|S_{M\wedge\tau}-S_{(M\wedge\tau)-1}||\mathscr{F}_t] < \frac{\text{const.}}{t},$$
$$E[(S_{M\wedge\tau}-S_{(M\wedge\tau)-1})^2|\mathscr{F}_t] < \frac{\text{const.}}{t^2}.$$

Then we have

$$\begin{aligned} E[S_{M\wedge\tau}^2|\mathscr{F}_t] &= E[\{S_{(M\wedge\tau)-1}+(S_{M\wedge\tau}-S_{(M\wedge\tau)-1})\}^2|\mathscr{F}_t] \\ &\leq \left(\frac{c^*}{\sqrt{t}}\right)^2 + 2\frac{c^*}{\sqrt{t}}\frac{\text{const.}}{t} + \frac{\text{const.}}{t^2} \end{aligned}$$

where the inequality uses $S_{(M\wedge\tau)-1} \leq \frac{c^*}{\sqrt{t}}$ (which follows from the definition of τ), the Lipschitz-continuity of η, and the previous inequalities. When t is large, the last line is less than $\frac{4(c^*)^2}{t}$, and hence (46) follows.

C.13 Proof of Lemma 14

The proof is almost the same as that of Pemantle (1990). However, at some places, his proof uses the assumption that the noise has a bounded support (which ensures that the step size of the process is of order $\frac{1}{t+1}$). In what follows, we will explain how to extend his argument to our setup with Gaussian noise.

Enlarge the domain of η as in the proof of Lemma 13. Pick t large enough, and assume that $S_t \geq \frac{c^*}{\sqrt{t}}$. Let $\tau = \inf\{k \geq t | S_k \leq \frac{c^*}{2\sqrt{t}}\}$. Recall that $X_k = S_k - S_{k-1}$ is a difference sequence. Let $\mu_k = E[X_k|\mathscr{F}_{k-1}]$. Consider a martingale $\{Z_k\}_{k=t}^{\infty}$ defined as $Z_k = S_t + \sum_{j=t+1}^{k} Y_j$, where $Y_k = 0$ for $\tau > k$ and $Y_k = X_k - \mu_k$ for $\tau \leq k$.

In the seventh to the last line on page 710, Pemantle (1990) argues that if the step size is of order $\frac{1}{k}$, then $\{Z_k\}$ is L^2-bounded (and hence the martingale convergence theorem applies).

In our setup, we can still prove that $\{Z_k\}$ is L^2-bounded. It is well-known that a martingale $\{Z_k\}$ is L^2-bounded if and only if

$$\sum_k E[(Z_k - Z_{k-1})^2] < \infty.$$

We will show that this inequality holds in our model. Using the argument similar to (47), we have

$$|S_k - S_{k-1}| < \frac{\hat{c}L + n\bar{b}|\varepsilon|L}{k}, \tag{48}$$

and hence

$$E[(S_k - S_{k-1})^2 | \mathscr{F}_{k-1}] < \frac{\text{const.}}{k^2},$$
$$E[|S_k - S_{k-1}| | \mathscr{F}_{k-1}] < \frac{\text{const.}}{k}.$$

These inequalities imply

$$\begin{aligned}
&E[(Z_k - Z_{k-1})^2 | \mathscr{F}_{k-1}] \\
&\le E[(X_k - \mu_k)^2 | \mathscr{F}_{k-1}] \\
&= E[(S_k - S_{k-1})^2 - 2\mu_k (S_k - S_{k-1}) + \mu_k^2 | \mathscr{F}_{k-1}] \\
&= E[(S_k - S_{k-1})^2 - 2E[S_k - S_{k-1} | \mathscr{F}_{k-1}](S_k - S_{k-1}) + (E[S_k - S_{k-1} | \mathscr{F}_{k-1}])^2 | \mathscr{F}_{k-1}] \\
&< \frac{\text{const.}}{k^2}.
\end{aligned}$$

Hence we have $E[(Z_k - Z_{k-1})^2] < \frac{\text{const.}}{k^2}$ for every k. Then obviously $\sum_k E[(Z_k - Z_{k-1})^2] < \infty$, as desired.

Also in the last line on page 710, Pemantle (1990) shows that if the step size is of order $\frac{1}{k}$, then

$$\mathrm{Var}\left(\sum_{k=t+1}^{\tau} Y_k\right) \le \sum_{k=t+1}^{\infty} \frac{\text{const.}}{k^2},$$

In our model, we can still prove the same inequality. It is well-known that the covariance of the martingale difference (Y_i, Y_j) is zero, and hence

$$\begin{aligned}
\mathrm{Var}\left(\sum_{k=t+1}^{\tau} Y_k\right) &= \mathrm{Var}\left(\sum_{k=t+1}^{\infty} Y_k\right) \\
&= \sum_{k=t+1}^{\infty} \mathrm{Var}(Y_k).
\end{aligned}$$

We have seen that $E[(Z_k - Z_{k-1})^2] < \frac{\text{const.}}{k^2}$. Since $\mathrm{Var}(Y_k) = E[(Z_k - Z_{k-1})^2]$, the desired inequality holds.