Misspecified Bayesian Learning by Strategic Players: First-Order Misspecification and Higher-Order Misspecification*

Takeshi Murooka
Osaka University

Yuichi Yamamoto
Hitotsubashi University

October 8, 2021

Abstract

We consider strategic players who may have a misspecified view about the world, and investigate their long-run behavior when they learn an unknown state from public signals over time. Our framework is flexible and allows for *higher-order misspecification*, in that a player may have a bias about the physical environment, a bias about the opponent's bias about the physical environment, and so on. We provide a condition under which players' beliefs and actions converge to a steady state, and then characterize how one's misspecification influences the long-run (steady-state) outcome. We apply these results to various economic examples such as Cournot competition, team production, and discrimination to study when one's misspecifcation improves her own payoff and how it influences the opponent's behavior. We also find that higher-order misspecification can have a significant impact on the equilibrium outcome, e.g., one's bias about the physical environment can have opposite effects on their payoffs and actions, depending on whether the opponent is aware of this bias or not.

*We thank Matthias Fahn, Drew Fudenberg, Kevin He, 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). 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, 2021). 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 proposes a general model which allows for such higher-order misspecification, and studies its economic consequences.

Specifically, we consider an infinite-horizon game in which players take actions each period and learn an unknown state from public signals over time. Players are Bayesian and maximize the expected payoffs just as in the standard game-theoretic model, but they evaluate information using misspecified models. We assume myopic players in order to rule out the folk-theorem type result.[2] Actions are unobservable. Our goal is to understand how a misspecified player behaves differently than the unbiased player in the long run, and how it influences the behavior of other players.

Why should we be interested in the long-run behavior of misspecified players,

[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.

[2] Our results are valid even for forward-looking players by assuming that they play a Markov perfect equilibrium.

rather than an equilibrium in a one-shot model? When players have a misspecified view about the world, they observe outcomes which are systematically different from the anticipation. Accordingly, it is likely that they eventually change the belief about some economic variable. For example, if a firm is persistently overconfident about some aspect of the demand function (e.g., the intercept of the inverse demand curve), on average, actual prices are lower than the firm's anticipation.[3] So after a long time, this firm becomes (unrealistically) pessimistic about other aspect of the demand (e.g., the slope of the inverse demand curve). Similarly, in tournaments, if an agent is persistently overconfident about her own capability, after a series of unexpected losses, she may think that the tournament is unfair. Our framework is useful to understand players' long-run behavior in these cases, and we show that this learning feature has a substantial impact on the equilibrium outcomes in various applications.

In Section 2, we consider a benchmark case in which there is no higher-order misspecification. Specifically, we assume that players have first-order misspecification only, in that they may have misspecified views about the physical environment and these first-order beliefs about the environment are common knowledge. This setup covers a wide range of applications, such as Cournot duopoly with misspecified demand and team production with overconfidence/prejudice. We show that players' beliefs and actions converge to a steady state under some condition, and then characterize how one's misspecification influences the steady-state outcomes. A novelty here is that we *quantify* the impact of misspecification on the steady-state outcomes, which allows us to discuss how each parameter of the model influences the equilibrium outcome, and how strategic interaction amplifies/reduces the impact of misspecification. For example, our result implies that in any symmetric game, both strategic substitutes and strategic complements *am-*

[3] Recent evidence suggests that overconfidence can be persistent: Hoffman and Burks (2020) find that workers are persistently overconfident about their own productivity, and Huffman, Raymond, and Shvets (2019) find that managers are persistently overconfident about future performance. In a laboratory experiment, Grossman and Owens (2012) report that subjects' beliefs are consistent with Bayesian updating and overconfident prior beliefs about their ability, and the overconfidence is persistent in the face of repeated feedback.

plify the impact of first-order misspecification.

Then in Section 3, we consider a model with higher-order misspecification. There are many types of higher-order misspecification we can think of, and here we focus on a particular one which seems economically relevant.[4] Specifically, we assume that each player may have a misspecified view about the physical environment as in the case of first-order misspecification, and on top of that, each player naively thinks that the opponent has the same view. In this setup, players are not aware of the opponent having a different view about the world; they think that their view about the world is absolutely correct. This describes, for example, a worker who is unaware of a bias of his colleague. We find that even with such higher-order misspecification, players' beliefs and actions still converge to a steady state under an additional assumption. We then quantify how one's misspecification influences this steady-state outcome.

In Section 4, we apply these results to more specific examples. We find that the presence of higher-order misspecification (i.e., unawareness of the opponent's bias) can have a significant impact on the equilibrium outcome. For example, when Alice and Bob work on a joint project, it is possible that Bob's overconfidence (about his own capability) improves his equilibrium payoff if his overconfidence is common knowledge, but reduces his payoff if Alice is not aware of the overconfidence. A point is that with higher-order misspecification, Alice faces *inferential naivety*: She makes an incorrect prediction about Bob's action. This leads to incorrect learning, e.g., if Alice overestimates Bob's effort, she finds that actual outputs are systematically worse than her anticipation, and becomes (unrealistically) pessimistic about the fundamental. Accordingly, Alice may take an action different from the one she would take if Bob's overconfidence was common knowledge, which may result in a qualitative difference in equilibrium payoffs.

We also find that our framework of higher-order misspecification is useful to explain *bias transmission*. In Section 4.3, we consider a teacher who has a bias

[4]We propose a general model of higher-order misspecification in Appendix A. In the working paper version (Murooka and Yamamoto, 2021), we present the analysis of different types of higher-order misspecification, as well as other applications such as a tournament.

against a particular type of students (e.g., female students). We show that the teacher's bias can endogenously induce these students' negative self-stereotypes, if the students are not aware of the teacher's bias.

Section 5 summarizes the related literature and concludes. Appendix A presents a general model and characterize the asymptotic behavior of players' actions and beliefs. Appendix B provides proofs.

2 First-Order Misspecification

2.1 Setup

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 $\mu \in \triangle\Theta$. In each period t, each player i has a belief $\mu_i^t \in \triangle\Theta$ about 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 other player $j \neq i$. 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 which describes a physical environment (e.g., a parameter which determines a market demand, a player's capability, etc) and ε is a random noise whose distribution is $N(0,1)$. Each player i receives a payoff $u_i(x_i,y)$. Both Q and u_i are twice continuously differentiable.

We assume that one of the players (player 2) has a biased view about the parameter a, while the other player is unbiased and knows the parameter a. We call it *first-order misspecification*, because player 2 has an incorrect first-order belief about the parameter a. Specifically, consider the following information structure:

- Player 1 believes that for each parameter θ, the signal y is given by $y = Q(x_1,x_2,a,\theta) + \varepsilon$.
- Player 2 (incorrectly) believes that for each parameter θ, the signal y is given by $y = Q(x_1,x_2,A,\theta) + \varepsilon$, where $A \neq a$.

- The above beliefs are common knowledge (e.g., player 2 believes that player 1 believes that $y = Q(x_1, x_2, a, \theta) + \varepsilon$, and the like).

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.[5] 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.[6]

In period one, both players have the same belief $\mu_1^1 = \mu_2^1 = \mu$, so a Nash equilibrium (x_1^1, x_2^1) solves the first-order condition $\frac{\partial E[U_i(x,\theta)|\mu]}{\partial x_i} = 0$ for each i, where the expectation is taken with respect to θ. At the end of period one, players

[5] 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.

[6] Condition (ii) is inessential if the game is dominance solvable (e.g., Cournot with linear demand in our application).

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_i^2 in period two is given by

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

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 $Q(x^1, a, \theta)$. These posteriors are common knowledge among players.[7] 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 $t \geq 3$, 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.

2.2 Steady-State Analysis

In this subsection, we will *assume* that the actions and the beliefs converge to a steady state in the long run, and characterize how one's misspecification influences this long-run (steady-state) outcome. In the next subsection, we will show that the actions and the beliefs indeed converge to this steady state under some conditions. As will be seen, these conditions are satisfied in many economic examples such as Cournot competition and team production.

A *steady state* in this model is a pair $(x_1^*, x_2^*, \mu_1^*, \mu_2^*)$ of an action profile and a

[7]This is because the players' information structure about the parameter a is common knowledge.

belief profile which satisfies the following four conditions:

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

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

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

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

The first two conditions are incentive compatibility, which requires that each player maximizes her payoff given some beliefs. The next two conditions require that these beliefs satisfy consistency: (3) asserts that the unbiased player 1 correctly learns the true state θ^* in a steady state. (4) requires that player 2's belief be concentrated on a state θ_2 with which her subjective signal distribution coincides with the true distribution. This condition must be satisfied in a steady state, because otherwise, player 2 is "surprised" by observed signals being different from what she thinks, and changes her belief about θ accordingly. In general, this steady-state belief θ_2 is different from the true state θ^*, i.e., the biased player 2 cannot learn θ^* correctly.

We assume that for each (x, A), there is a unique state θ_2 which solves the consistency condition $Q(x, a, \theta^*) = Q(x, A, \theta)$, and we denote it by $\theta_2(x, A)$. Intuitively, this $\theta_2(x, A)$ is player 2's long-run belief given an action profile x; if players choose the same action profile x every period, then almost surely, player 2's belief will be concentrated on the state $\theta_2(x, A)$ after a long time (Berk (1966)). Player 1's long-run belief is defined as $\theta_1(x, A) = \theta^*$ for all x and A, because she is unbiased and can learn the true state θ^* regardless of players' play.

Our goal is to quantify how player 2's misspecification A influences the steady-state action defined above. Before doing so, it is useful to think about how one's action influence the opponent's steady-state action. Consider player i's *asymptotic best response correspondence*, which is defined as

$$BR_i(x_{-i}) = \left\{ x_i \,\middle|\, x_i \in \arg\max_{x_i'} U_i\left(x_i', x_{-i}, \theta_i(x, A)\right) \right\}. \tag{5}$$

Intuitively, $BR_i(x_{-i})$ describes player i's steady-state action in a *single-agent learning problem* where player i learns the state while the opponent simply chooses the same action x_{-i} every period. Player 2's asymptotic best response BR_2 is different from the standard best response, as it describes the optimal action *in the long run* where the belief $\theta_2(x,A)$ is endogenously determined. On the other hand, player 1's asymptotic best response BR_1 coincides with the standard best response correspondence given the state θ^*, because her long-run belief is constant, i.e., $\theta_1(x,A) = \theta^*$ for all x (correct learning regardless of actions). By a fixed-point theorem, $BR_i(x_{-i})$ is non-empty for all x_{-i}. Also a standard argument shows that BR_i is upper hemi-continuous in x_{-i}.

When the asymptotic best response is a function (rather than a correspondence), its slope BR_i' can be computed by

$$BR_i' = -\frac{M_{ij}}{M_{ii}},$$

where for each i and j (possibly $i = j$),

$$M_{ij} = \left.\frac{\partial^2 U_i(x,\theta)}{\partial x_i \partial x_j}\right|_{\theta=\theta_i(x,A)} + \left.\frac{\partial^2 U_i(x,\theta)}{\partial x_i \partial \theta}\right|_{\theta=\theta_i(x,A)} \frac{\partial \theta_i(x,A)}{\partial x_j}.$$

measures how player j's action influences player i's marginal utility *in the long run*. To be precise, suppose that players choose the same action every period, and that player j increases the action x_j at bit. This influences player i's marginal utility directly and indirectly through the belief $\theta_i(x,A)$ in the long run. The first term of M_{ij} represents this direct effect, and the second term represents the indirect effect. For $i = 1$, the indirect effect is zero, because player 1's long-run belief is constant and does not depend on the actions ($\theta_1(x,A) = \theta^*$ for all x). Hence $BR_1' = \frac{\partial^2 U_1/\partial x_1 \partial x_2}{\partial^2 U_1/\partial^2 x_1}$, which is precisely the slope of the standard best-response curve. Also, for $i = 2$, the indirect effect disappears in the limit as $A \to a$, because $\theta_2(x,a) = \theta^*$ for all x. So when misspecification is small (i.e., A is close to a), each M_{ij} is approximated by $\frac{\partial^2 U_i}{\partial x_i \partial x_j}$, which means that BR_i' is approximately the same as the slope of the standard best-response function.

Let

$$M_{2A} := \left.\frac{\partial^2 U_2(x,A,\theta)}{\partial x_2 \partial A}\right|_{\theta=\theta_2(x,A)} + \left.\frac{\partial^2 U_2(x,A,\theta)}{\partial x_2 \partial \theta}\right|_{\theta=\theta_2(x,A)} \frac{\partial \theta_2(x,A)}{\partial A} \tag{6}$$

denote how player 2's bias A influences her marginal utility in the long run. Again, the first term $\frac{\partial^2 U_2}{\partial x_2 \partial A}$ measures the direct effect, while the second term $\frac{\partial^2 U_2}{\partial x_2 \partial \theta_2}\frac{\partial \theta_2}{\partial A}$ measures the indirect effect through the belief. Our first proposition quantifies the impact of player 2's first-order misspecification on the steady-state action.

Definition 1. A steady state x^* is *regular* if the following conditions are satisfied in x^*: (i) the steady-state action x_i^* is uniquely optimal, i.e., $U_i(x^*, \theta_i(x^*,A)) > U_i(x_i, x_{-i}^*, \theta_i(x^*,A))$ for all i and $x_i \neq x_i^*$, (ii) x^* and $\theta_2(x^*,A)$ are interior points, (iii) $BR_1' BR_2' \neq 1$, and (iv) $M_{ii} < 0$ for each i.

Proposition 1 (Steady State under First-Order Misspecification)**.** *Let x^* be a regular steady state for some parameter A^*. Then there is an open neighborhood of A^* such that for any value A in this neighborhood, there is a regular steady state x^* which is continuous with respect to A, and we have*

$$\frac{\partial x_2^*}{\partial A} = -\frac{M_{2A}}{M_{22}} \cdot \frac{1}{1 - BR_1' BR_2'},$$
$$\frac{\partial x_1^*}{\partial A} = \frac{\partial x_2^*}{\partial A} \cdot BR_1'.$$

Suppose in addition that given the parameter A^*, the steady state is unique and each asymptotic best response BR_i is a continuous function. Then, $BR_1' BR_2' < 1$.[8]

Note that the regularity conditions (i) and (ii) are standard, and the condition (iii) is satisfied for generic parameters. The condition (iv) reduces to player i's

[8] If these additional assumptions do not hold, there may be a steady state with $BR_1' BR_2' > 1$. But it seems that such a steady state is unstable in an evolutionary sense, especially when misspecification is small. Indeed, in a one-shot game with correctly specified model, a Nash equilibrium with $BR_1' BR_2' > 1$ is not stable under the replicator dynamics (hence it is not an ESS) or the best response dynamics. So in practice, if players' play converge after a long time, it is natural to expect that $BR_1' BR_2' < 1$ in the steady state.

second-order condition for incentive compatibility when misspecification is small (i.e., A is close to a).[9][10]

Under this regularity condition, the proposition above shows that the impact of first-order misspecification on the steady-state action is represented as the *base misspecification effect* $-\frac{M_{2A}}{M_{22}}$ times the *multiplier effect* $\frac{1}{1-BR_1'BR_2'}$. The base misspecification effect measures how player 2's bias influences her steady-state action x_2^* *in the absence of strategic interaction*. To see what it means, suppose that player 1 chooses the same fixed action each period, so player 2 faces a single-agent problem. Suppose that player 2's bias A increases a bit. This influences player 2's marginal utility by M_{2A} (recall that this includes the indirect effect through the belief in the long run), and hence her optimal long-run action changes. The base misspecification effect $-\frac{M_{2A}}{M_{22}}$ measures this change.

The multiplier effect $\frac{1}{1-BR_1'BR_2'}$ in Proposition 1 measures how strategic interaction between two players amplifies/weakens the base misspecification effect. To better understand the nature of this multiplier effect, suppose that player 2 changes her action by Δ. Then player 1 best-responds to it and changes her action by $BR_1'\Delta$, which in turn has a feedback effect of $BR_1'BR_2'\Delta$ on player 2's steady-state action; note that player 1's action influences player 2's optimal action directly and indirectly through her belief $\theta_2(x,A)$, and both these effects are taken into account in the asymptotic best response BR_2'.

This process continues multiple times; the feedback effect on player 2's action influences player 1's action, which again causes a feedback effect of $(BR_1'BR_2')^2\Delta$ on player 2's action, and so on. Summing all these feedback effects, player 2's action changes by

$$\sum_{k=0}^{\infty}(BR_1'BR_2')^k\Delta = \frac{1}{1-BR_1'BR_2'}\Delta. \tag{7}$$

[9]Heidhues, Kőszegi, and Strack (2018) impose a similar assumption: They consider a single-agent learning problem and assume a unique steady state, which requires $M_{ii} \leq 0$ in the steady state.

[10]We conjecture that a steady state with $M_{22} > 0$ is unstable in the sense that players' actions converge there with zero probability. (We can actually show that if player 1 chooses the steady-state action every period and player 2 learns an unknown state θ, then player 2's action never converges to a steady state with $M_{22} > 0$. The proof is available upon request.)

So the multiplier $\frac{1}{1-BR_1'BR_2'}$ can be seen as a result of the infinite adjustment process between the two strategic players.

The following corollary is an immediate consequence of Proposition 1:

Corollary 1. *Suppose that all the assumptions in Proposition 1 (including the ones in the second part) are satisfied. Then we have the following results:*

(i) *The multiplier $\frac{1}{1-BR_1'BR_2'}$ is positive. So a strategic interaction influences the size of the impact of misspecification, but not the direction.*

(ii) *If $\text{sgn}(BR_1') = \text{sgn}(BR_2')$, then the multiplier $\frac{1}{1-BR_1'BR_2'}$ is greater than one. So both strategic substitutes and strategic complements amplify the impact of misspecification.*

(iii) *If $\text{sgn}(BR_1') \neq \text{sgn}(BR_2')$, then the multiplier $\frac{1}{1-BR_1'BR_2'}$ is less than one. So a strategic interaction reduces the impact of misspecification.*

This corollary characterizes when a strategic interaction amplifies/reduces the impact of misspecification. An interesting special case is symmetric games, such as Cournot duopoly and team production. When $A = a$, we have $\text{sgn}(BR_1') = \text{sgn}(BR_2')$ in any symmetric equilibrium of a symmetric game. So part (ii) of the corollary implies that a strategic interaction always amplifies the impact of misspecification in these games, if misspecification is small. On the other hand, part (iii) shows that a strategic interaction reduces the impact of misspecification if $\text{sgn}(BR_1') \neq \text{sgn}(BR_2')$. This condition is satisfied, for example, in a tournament model.[11]

Remark 1. So far we have assumed that player 1 knows the true parameter a, but the result similar to Proposition 1 still holds even when both players have first-order misspecification. Suppose that player 1 believes that the true parameter is $A_1 \neq a$, and player 2 believes that the true parameter is $A \neq a$. Suppose also

[11] Consider a tournament model by Lazear and Rosen (1981) in which player i's payoff is $P_i(e_i - e_j)w - c(e_i)$ where $w > 0$ is the prize for a winner, $c(\cdot)$ is an increasing and convex cost, and $P_i(\cdot)$ is i's probability of winning which satisfies $P_1(\cdot) = 1 - P_2(\cdot)$. Then, so long as $P_i''(0) \neq 0$, the standard best response curves have opposite signs.

that these first-order beliefs are common knowledge. Then the impact of player 2's misspecification A on the steady-state actions is still described by the formula presented in Proposition 1, with a minor modification on the definition of BR_1'; now it must involve an indirect learning effect, in order to take into account the endogeneity of her long-run belief θ_1.

2.3 Sufficient Condition for Convergence

In a single-agent finite-action setup, Esponda, Pouzo, and Yamamoto (2021) provide a fairly general condition for convergence; they show that the agent's belief converges to a steady state almost surely if an "identifiability condition" holds. In this subsection, we will show that the same result holds in our two-player continuous-action model. Having two players does not cause a serious difficulty, because our problem is essentially a single-agent problem; since player 1 is unbiased and learns the true state, we only need to take care of player 2's belief evolution.

On the other hand, having continuous actions causes a technical complication. When actions are finite, an agent's *action frequency* is represented as a finite-dimensional vector. Esponda, Pouzo, and Yamamoto (2021) show that the motion of this action frequency is approximated by a differential inclusion, and use this result to prove convergence. In our continuous-action model, the action frequency is an infinite-dimensional vector, and it moves in a Banach space (rather than a Euclidean space). Thus we need to think about whether this differential-inclusion approach is valid even in a Banach space. As we show in Appendix A, such an extension is indeed possible, if we choose an appropriate norm for the Banach space

Let us define the identifiability condition in our environment. For each action profile x, define player 2's *surprise function* as

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

Intuitively, this surprise function measures how player 2's subjective expectation $Q(x,\theta,A)$ about the output is different from the truth, when she believes that the state is θ.[12] Then for each probability measure $\sigma \in \triangle X$ on the set of action profiles, define a *weighted surprise function* as

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

Intuitively, this function measures how player 2's subjective expectation is different from the truth on average, when players take different actions in different periods. The *identifiability* requires that for each σ, the weighted surprise function $K_2(\theta,\sigma)$ has a unique minimizer $\theta_2(\sigma)$ and it is an interior point. The following proposition shows that this identifiability condition ensures convergence.

Proposition 2. *Suppose that there is a unique steady state* $(x_1^*, x_2^*, \theta_1, \theta_2)$ *and it is regular.*[13] *Suppose also that the identifiability condition holds. Then almost surely, player* 2*'s belief converges to the steady state belief, i.e.,* $\lim_{t\to\infty} \mu_2^t = 1_{\theta_2}$.

Remark 2. Identifiability is sufficient for convergence, but not necessary. For example, Proposition 9 of Esponda, Pouzo, and Yamamoto (2021) show that in a single-agent problem, the agent's action converges if payoffs and information are "monotone" in some sense. We can show that a similar result holds in our environment, which ensures that players' actions and beliefs converge to a steady state in the team production problem studied in Section 4.2 for general Q. See the working paper version for more details.

3 Higher-Order Misspecification

In the previous section, we have studied the case in which players have incorrect views about the environment, assuming that they correctly understand what the

[12]This surprise function is exactly the Kullback-Leibular divergence between the true output distribution and the subjective distribution. See Appendix A for the general definition of the Kullback-Leiblar divergence.

[13]Assuming a unique steady state is not essential, but it simplifies the statement of the proposition. In the proof, we actually show that the belief converges even when there are multiple steady states.

opponent thinks about the environment. However, economic agents often have *higher-order misspecification*, in that they may have a biased view about the opponent's view about the environment (second-order misspecification), a biased view about the opponent's second-order misspecification, and so on.[14]

In this section, we will focus on a special form of higher-order misspecification: We will assume that each player has a biased view about the environment, 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). Our goal in this section is to characterize how such misspecification influences players' long-run behavior. 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.

3.1 Setup: Double Misspecification

We will consider the following information structure:

- Each player i (incorrectly) believes that for each parameter θ, the signal y is given by $y = Q(x_1, x_2, A_i, \theta) + \varepsilon$, where $A_i \neq a$.
- 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.

This setup is substantially different from the first-order misspecification in the previous section, because now players have *inferential naivety* and make incorrect

[14] 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.

predictions about the opponent's play. Indeed, while player i believes that the opponent (player j) maximizes the payoff conditional on the parameter A_i, in reality, the opponent maximizes the payoff conditional on the parameter A_j. 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*. Hypothetical player 1 is player 1 who thinks that it is common knowledge that the true technology is A_2. Intuitively, player 2 thinks that this hypothetical player is her opponent, and hence each period, player 2 chooses a N ash equilibrium action against this hypothetical player. Similarly, hypothetical player 2 is player 2 who thinks that it is common knowledge that the true technology is A_1. Each period, player 1 chooses a Nash equilibrium against this hypothetical player.

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 solves 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 $\hat{x}_{-i}^1$ about the opponent's play. Then in period two, players play a Nash equilibrium given this belief profile $\mu^2 = (\mu_1^2, \mu_2^2, \hat{\mu}_1^2, \hat{\mu}_2^2)$.[15] Likewise, in any subsequent period $t \geq 3$, players play a Nash equilibrium given the posterior beliefs computed by Bayes' rule.

3.2 Steady-State Analysis

As in the case of first-order misspecification, we assume that players' actions and beliefs converge to a steady state, and study how one's misspecification influences this steady-state outcome. We will provide a sufficient condition for convergence in Section 3.3.

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

$$Q(x_i, \hat{x}_j, A_i, \theta) = Q(x_1, x_2, , a, \theta^*),$$

so that player i's subjective model (the left-hand side) exactly explains the actual output (the right-hand side). A critical difference from the case of first-order misspecification is that player i has inferential naivety and uses $\hat{x}_j$ (rather than x_j) when evaluating the average output. In what follows, we will assume that $\theta_2(x, A)$ is unique for each x and A_i.

With this notation, a *steady state* under double misspecification is defined as

[15] 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. The same is true for player 2 and hypothetical player 1.

$(x_1^*, x_2^*, \hat{x}_1^*, \hat{x}_2^*, \mu_1^*, \mu_2^*, \hat{\mu}_1^*, \hat{\mu}_2^*)$ which satisfies

$$x_i^* \in \arg\max_{x_i} U_i(x_i, \hat{x}_{-i}^*, A_i, \theta_i) \quad \forall i, \tag{8}$$

$$\hat{x}_i^* \in \arg\max_{\hat{x}_i} \hat{U}_i(\hat{x}_i, x_{-i}^*, A_{-i}, \theta_{-i}) \quad \forall i, \tag{9}$$

$$\mu_1^* = \hat{\mu}_2^* = 1_{\theta_1(x,A_1)}, \tag{10}$$

$$\mu_2^* = \hat{\mu}_1^* = 1_{\theta_2(x,A_2)}. \tag{11}$$

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.

As in the case of first-order misspecification, we will quantify how one's misspecification influences the steady-state outcome, using the slope of asymptotic best response curve. For notational convenience, let player 3 refer to hypothetical player 1, and player 4 refer to hypothetical player 2. Then define the slope of player i's asymptotic best response curve with respect to player j's action as

$$BR'_{ij} = -\frac{M_{ij}}{M_{ii}}$$

where for each $i, j = 1, 2, 3, 4$ (possibly $i = j$),

$$M_{ij} = \left.\frac{\partial^2 U_i(x,\theta)}{\partial x_i \partial x_j}\right|_{\theta=\theta_i(x,A_i)} + \left.\frac{\partial^2 U_i(x,\theta)}{\partial x_i \partial \theta}\right|_{\theta=\theta_i(x,A_i)} \cdot \frac{\partial \theta_i(x,A)}{\partial x_j}$$

denotes the impact of player j's action on player i's marginal utility in the long run. Here the first term is the direct effect, and the second term is the indirect effect through the steady-state belief θ_i.

Intuitively, BR'_{ij} measures how player j's action influences player i's optimal long-run action, when other players' actions are fixed. The mathematical definition of BR'_{ij} is exactly the same as that for first-order misspecification, but there are two important remarks. First, each M_{ij} here involves the indirect effect caused by inferential naivety, and thus BR'_{ij} need not be zero even when player i does not think that player j is the opponent.[16] For example, consider $BR'_{12} = -\frac{M_{12}}{M_{11}}$. Since player 1 does not think that player 2 is the opponent, she does not best-respond to player 2's action, which means that the direct effect in M_{12} is zero. However, a change in player 2's action influences player 1's steady-state belief θ_1; it leads to misguided learning, because player 1 is not aware of a change in player 2's action. Hence the indirect effect in M_{12} is non-zero, and so is BR'_{12}.

Second, this indirect effect from inferential naivety does not vanish even when A_i approaches a. So even in the limit case with $A_i = a$, the slope BR'_{ij} is not approximated by the slope of the standard best response curve. This is in a sharp contrast with the case of first-order misspecification, where all indirect effects vanish in the limit as $A \to a$.

For each $i = 2,3$, let

$$M_{iA} := \left.\frac{\partial^2 U_i(x,\theta,A_2)}{\partial x_i \partial A_2}\right|_{\theta=\theta_i(x,A_2)} + \left.\frac{\partial^2 U_i(x,\theta,A_2)}{\partial x_i \partial \theta}\right|_{\theta=\theta_i(x,A_2)} \frac{\partial \theta_i(x,A_2)}{\partial A_2}$$

denote the impact of player i's first-order misspecification on her marginal utility. The following proposition characterizes how player 2's misspecification influences the steady-state actions.

Definition 2. A steady state x^* is *regular* if the following conditions are satisfied in x^*: (i) the steady-state action x_i^* is uniquely optimal, (ii) x^* and $\theta_i(x^*,A_i)$ are interior points, (iii) $BR'_{14}BR'_{41} \neq 1$, $BR'_{23}BR'_{32} \neq 1$, and $BR'_{14}BR'_{41} + BR'_{23}BR'_{32} + (BR'_{21} + BR'_{23}BR'_{31})(BR'_{12} + BR'_{14}BR'_{42}) \neq 1$, and (iv) $M_{ii} < 0$ for each i.[17]

Proposition 3 (Steady State under Double Misspecification)**.** *Let x^* be a regular steady state for some parameter $A^* = (A_1^*, A_2^*)$. Then there is an open neighbor-*

[16] More precisely, we always have $BR'_{13} = BR'_{24} = BR'_{34} = BR'_{43} = 0$, but other slopes BR'_{ij} are non-zero in general.

[17] As in the case with first-order misspecification, the regularity conditions (i) and (ii) ensure that the steady state is continuous with respect to the parameter A_i and the first-order condition for the incentive compatibility is satisfied there. The condition (iii) is needed for the multiplier effect to be well-defined. The condition (iv) ensures that the base misspecification effect and the slope of the asymptotic best response curve are well-defined. This condition is also useful when we interpret the base misspecification effect.

hood of A_2^* such that for any value A_2 in this neighborhood, there is a regular steady state x^* which is continuous with respect to A_2, and we have

$$\frac{\partial x_2^*}{\partial A_2} = \left(-\frac{M_{2A}}{M_{22}} - BR'_{23}\frac{M_{3A}}{M_{33}}\right)\left(\frac{1}{1-BR'_{23}BR'_{32}}\right)\left(\frac{1}{1-NE'_1NE'_2}\right),$$
$$\frac{\partial x_1^*}{\partial A_2} = \frac{\partial x_2^*}{\partial A_2}\cdot NE'_1,$$

where

$$NE'_1 = \frac{BR'_{12}+BR'_{14}BR'_{42}}{1-BR'_{14}BR'_{41}} \quad \textit{and} \quad NE'_2 = \frac{BR'_{21}+BR'_{23}BR'_{31}}{1-BR'_{23}BR'_{32}}.$$

The term $-\frac{M_{2A}}{M_{22}} - BR'_{23}\frac{M_{3A}}{M_{33}}$ in the first equation is the base misspecification effect of double misspecification. Suppose that the bias A_2 increases a bit. This influences player 2's steady-state action in two ways: First, it increases player 2's bias about the physical environment a, which influences her optimal action directly and indirectly through the belief. This effect is measured by $-\frac{M_{2A}}{M_{22}}$, and note that this part is exactly the same as what we explained in the case of first-order misspecification. Second, when A_2 increases, hypothetical player 1's bias about the physical environment a increases. Hence this hypothetical player modifies the action, and player 2 best-responds to it. This effect is measured by $-BR'_{23}\frac{M_{3A}}{M_{33}}$. This second effect is a consequence of player 2's second-order misspecification (inferential naivety).

Proposition 3 shows that this base misspecification effect is further amplified by the two multipliers, $\frac{1}{1-BR'_{23}BR'_{32}}$ and $\frac{1}{1-NE'_1NE'_2}$. The first multiplier effect, $\frac{1}{1-BR'_{23}BR'_{32}}$, is similar to the multiplier effect appearing in Proposition 1, and it represents how the strategic interaction between player 2 and hypothetical player 1 (which happens in player 2's mind) amplifies the base misspecification effect, holding other players' actions being fixed. So the term $(-\frac{M_{2A}}{M_{22}} - BR'_{23}\frac{M_{3A}}{M_{33}})(\frac{1}{1-BR'_{23}BR'_{32}})$ in the equation measures how player 2's misspecification influences her own action, when player 1's action is fixed.

The second multiplier effect, $\frac{1}{1-NE'_1NE'_2}$, captures what happens when player 1's action is not fixed and she changes her action over time. The economic interpretation of this multiplier effect is very different from the first one, in that it is *not* about the impact of strategic interaction; rather, it measures how incorrect learning due to inferential naivety amplifies the impact of misspecification.

To see what it means, suppose that player 2's action changes by Δ due to misspecification. This does not influence player 1's action immediately, because player 1 is not aware of player 2's misspecification, and thus does not best-respond to this change. However, in the long run, it causes incorrect learning and influences player 1's steady-state belief μ_1, which in turn influences player 1's action directly and indirectly through hypothetical player 2's action. (Note that the belief μ_1 influences hypothetical player 2's optimal action, and player 1 best-responds to it.) This effect is $BR_{12}\Delta + BR_{14}BR_{42}\Delta$. Since this effect is further amplified by $\frac{1}{1-BR_{14}BR_{41}}$ due to the strategic interaction between player 1 and hypothetical player 2, in total, player 1's action changes by $NE_1'\Delta$. Then for the same reason, this change in player 1's action influences player 2's action, which influences player 1's action, and so on. This infinite process leads to the multiplier $\frac{1}{1-NE_1'NE_2'}$.[18]

Remark 3. In this section, we have assumed two-sided misspecification, in that both players 1 and 2 are misspecified. For some applications, it is also important to think about one-sided misspecification; e.g., one can think of a seller-buyer problem where only a buyer is misspecified while a seller is fully rational. It turns out that even with such one-sided misspecification, the result similar to Proposition 3 still holds. Specifically, the equations in Proposition 3 are still valid, if we replace NE_1' with the slope of the best-response function BR_1'. This is because when player 1 is fully rational, then she correctly learns the state and simply best-

[18] NE_1' can be also seen as the slope of *player 1's asymptotic Nash equilibrium correspondence* defined as

$$NE_1(x_2) = \{x_1 | \exists \hat{x}_2 \text{ satisfying (8) for } i=1, \text{ (9) for } i=2, (10)\}.$$

In words, $NE_1(x_2)$ denotes player 1's steady-state action, when player 2 chooses the same action x_2 every period while the other players learn the state and adjust actions. Note that player 2 is not player 1's opponent, but nonetheless her action x_2 influences player 1's steady-state action due to the incorrect learning: If player 2 changes the action and player 1 is not aware of it, player 1's long-run belief is affected, and so is her long-run optimal action. So its slope, NE_1, measures how a marginal change in player 2's (constant) action x_2 influences player 1's steady-state action.

responds to player 2's action.

3.3 Sufficient Condition for Convergence

Now we will think about whether players' actions and beliefs indeed converge to a steady state. The problem here is two-dimensional, in that both player 1's belief μ_1^t and player 2's belief μ_2^t evolve in a non-trivial way.[19] This makes our analysis significantly more complicated than that of first-order misspecification, where there is only one belief which moves in a non-trivial way. In particular, the proof techniques developed for a single-agent learning problem in the literature are not applicable. Nonetheless, we find that the belief converges if the identifiability condition and some additional assumption hold.

Recall that in the case with first-order misspecification, identifiability requires each (weighted) surprise function to have a unique minimizer. Under double misspecification, player i's *surprise function* is defined as

$$K_i(\theta,x) = \frac{(Q(x_i,\hat{x}_{-i},\theta,A_i) - Q(x_i,x_{-i},\theta,a))^2}{2}$$

for each action profile $x = (x_1,x_2,\hat{x}_1,\hat{x}_2)$, and her *weighted surprise function* is defined as

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

for each probability measure $\sigma \in \triangle(X_1 \times X_2 \times X_1 \times X_2)$. These surprise functions are a bit different from those under first-order misspecification because of inferential naivety; player i thinks that players play $(x_i,\hat{x}_{-i})$, but the actual actions are (x_i,x_{-i}). Identifiability requires that each of the above surprise functions has a unique minimizer.

Under double misspecification, we need an additional assumption for convergence. To state our condition formally, consider the *single-agent* learning problem in which player 1 (and hypothetical player 2) learns the state over time, while the belief of player 2 (and of hypothetical player 1) is fixed at some value θ_2. Let

[19] Recall that the beliefs of the hypothetical players are exactly the same as those of the real players, i.e., we have $\mu_1^t = \hat{\mu}_2^t$ and $\mu_2^t = \hat{\mu}_1^t$.

$f_1(\theta_2)$ denote the set of steady-state beliefs of player 1 in this problem, that is, $f_1(\theta_2)$ is the set of all θ_1 such that there is $(x_1, x_2, \hat{x}_1, \hat{x}_2)$ satisfying the consistency condition (10) and the incentive-compatibility condition (8) and (9) given $\mu_2 = \hat{\mu}_1 = 1_{\theta_2}$.

Likewise, consider the single-agent learning problem in which player 2 (and hypothetical player 1) learns the state over time, while the belief of player 1 (and of hypothetical player 2) is fixed at some value θ_1. Then let $f_2(\theta_1)$ denote the set of steady-state belief of player 2, that is, $f_2(\theta_1)$ is the set of all θ_2 such that there is $(x_1, x_2, \hat{x}_1, \hat{x}_2)$ satisfying the consistency condition (11) and the incentive-compatibility condition (8) and (9) given $\mu_1 = \hat{\mu}_2 = 1_{\theta_1}$.

The following proposition present a sufficient condition for convergence under double misspecification.

Proposition 4. *Suppose that there is a unique steady state* $(x_1^*, x_2^*, \theta_1, \theta_2)$ *and it is regular. Suppose also that for each* i *and* σ*, the weighted surprise function* $K_i(\theta, \sigma)$ *has a unique minimizer* $\theta_2(\sigma)$ *and it is an interior point. 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.

Assumption (i) in this proposition is identifiability, and it ensures that the belief converge in every one-dimensional problem. 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. The proposition above shows that the beliefs indeed converge under these conditions.

4 Applications

4.1 Cournot duopoly

Consider a symmetric Cournot duopoly. Each firm $i = 1,2$ simultaneously chooses its quantity x_i, and then they observe a market price $y = Q(x_1 + x_2, a, \theta) + \varepsilon$, where a is a parameter which influences the demand and θ is an unknown economic state. Firm i's payoff is $u_i(x_i, y) = x_i y - c(x_i)$, where $x_i y$ is firm i's revenue and $c(x_i)$ is firm i's production cost. We assume that the inverse demand function Q is strictly decreasing and weakly concave in the first element, and the cost function c is strictly increasing and weakly convex.[20]

Kyle and Wang (1997), Heifetz, Shannon, and Spiegel (2007), and Englmaier (2010) study (a variant of) one-shot Cournot competition with linear demand, and show that (a moderate level of) overconfidence about the market demand is beneficial, in that an overconfident firm earns higher equilibrium payoffs than the unbiased rival firm. Intuitively, the overconfident firm is willing to produce more than in the correctly specified model. Knowing that, the unbiased firm reduces its production level in equilibrium, which yields higher profits to the overconfident firm. This mechanism is similar to the commitment effect in the Stackelberg duopoly.

However, their result relies on two implicit assumptions. First, they assume that the game is one-shot. It is not a priori clear if their result persists in the long run, because when the game is repeatedly played, the overconfident firm is "surprised" by a realized price being lower than its anticipation, and modifies the (subjective) view about the demand function. Second, they assume that a firm's overconfidence is common knowledge, while in reality the opponent may not recognize it. In this section, we will relax these assumptions and investigate how it changes the result.

[20]These assumptions imply a concave payoff function, a downward-sloping best response curve, and a unique Nash equilibrium under the correctly specified model. See, for example, Tirole (1988).

First-order misspecification. To begin with, we will relax the first assumption only, and consider a dynamic model in which one's overconfidence is common knowledge. Specifically, we consider the model of first-order misspecification in which firm 2 incorrectly believes that the true parameter is $A > a$. We assume that $Q_A > 0$ and $Q_{xA} \geq 0$ for all x with $x_1 + x_2 > 0$, which means that firm 2 is overconfident about the price level Q and (weakly) overconfident about the slope of the inverse demand curve Q_x. Note that a similar assumption is imposed in Kyle and Wang (1997). Firm 1 knows that the true parameter is a, and the firms' first-order beliefs are common knowledge. We also assume that $Q_\theta > 0$ and $Q_{x\theta} \geq 0$ for all x with $x_1 + x_2 > 0$, i.e., the state θ has positive impacts on the price level and the slope of the inverse demand function.

Here are two examples which satisfy the assumptions above:[21]

$$Q(x_1 + x_2, a, \theta) = a - (1 - \theta)(x_1 + x_2), \tag{12}$$

$$Q(x_1 + x_2, a, \theta) = \theta - (1 - a)(x_1 + x_2). \tag{13}$$

In the first example (12), firm 2 is overconfident about the intercept of the demand function and learns its slope.[22] Conversely, in the second example (13), firm 2 is overconfident about the slope and learns the intercept.[23]

We will consider how firm 2's overconfidence influences the long-run steady state outcome. Recall from Proposition 1 that the impact of firm 2's overconfidence on its own steady-state action is represented as the base misspecification

[21]These examples satisfy the regularity condition for first-order misspecification with $A = a$, and for double misspecification with $A_1 = A_2 = a$. They also satisfy the conditions stated in Propositions 2 and 4, so the firms' beliefs and actions converge to the steady state. The proof can be found in the working paper version (Murooka and Yamamoto, 2021).

[22]This happens, for example, when the firm is overconfident about the preference of the representative customers and learns their number. Suppose that there are $\frac{1}{1-\theta}$ customers, and each of them purchases $a - p$ units of products, where p is a price. Then, the total demand is $x = \frac{a-p}{1-\theta}$, which results in the inverse demand function $p = a - (1 - \theta)x$.

[23]This happens, for example, when the firm is overconfident about the number of the customers and learns their preference. Suppose that there are $\frac{1}{1-a}$ customers, and each of them purchases $\theta - p$ units of products. Then, the total demand is $x = \frac{\theta - p}{1-a}$, which results in the inverse demand function $p = \theta - (1 - a)x$.

effect $-\frac{M_{2A}}{M_{22}}$ times the multiplier. For ease of exposition, we assume that misspecification is small (i.e., A is close to a) so that $M_{22} < 0$ and the multiplier is positive. Simple algebra shows that the base misspecification effect in our Cournot model is written as

$$-\frac{1}{M_{22}}\left[\underbrace{\overbrace{Q_A(x_1^*+x_2^*,A,\theta_2)}^{\text{direct effect}}+\overbrace{\frac{\partial\theta_2}{\partial A}Q_\theta(x_1^*+x_2^*,A,\theta_2)}^{\text{indirect effect}}}_{\text{on the price level}}+x_2^*\left(\underbrace{\overbrace{Q_{xA}(x_1^*+x_2^*,A,\theta_2)}^{\text{direct effect}}+\overbrace{\frac{\partial\theta_2}{\partial A}Q_{x\theta}(x_1^*+x_2^*,A,\theta_2)}^{\text{indirect effect}}}_{\text{on the slope}}\right)\right]. \tag{14}$$

This expression shows that the long-run behavior of the overconfident firm is governed by two countervailing forces. The first one is the direct effect, $Q_A + x_2 Q_{xA}$, which measures how firm 2's misspecification directly influences its marginal utility. Since we assume $Q_A > 0$ and $Q_{xA} \geq 0$, this effect is positive, and hence *boosts* the firm's incentive to produce. This is exactly the effect studied in Kyle and Wang (1997). The second one is the indirect effect, $\frac{\partial\theta_2}{\partial A}(Q_\theta + x_2 Q_{x\theta})$, which measures how firm 2's learning (about θ) influences the marginal utility in the long run. Since $\frac{\partial\theta_2}{\partial A} < 0$, $Q_\theta > 0$, and $Q_{x\theta} \geq 0$, this effect is negative, and hence *weakens* the firm's incentive to produce.

If the direct effect is larger than the indirect effect, the overconfident firm is willing to produce more even in the long run. This means that the result of Kyle and Wang (1997) persists, i.e., in the long-run steady state, the rival firm best-responds by producing less, which yields a higher profit to the overconfident firm.[24]

On the other hand, if the indirect effect outweighs the direct effect, the overconfident firm is willing to produce *less* in the long run. Then the commitment effect works towards the opposite direction, i.e., the rival firm best-responds by producing more, which harms the overconfident firm's profit. So in this case, the result of Kyle and Wang (1997) is overturned and a firm's overconfidence is detrimental in the long run.

[24]However, due to the indirect effect, the overconfident firm's profit is less than that in the one-shot game: By the incorrect learning, the commitment effect is weakened in the long run.

There is another interpretation of the base misspecification effect (14). Note that the first two terms in the brackets cancel out, because the overconfident firm correctly predicts the price level in the steady state.[25] Accordingly, the base misspecification effect is rewritten as

$$-\frac{x_2^*}{M_{22}}\left(Q_{xA}+\frac{\partial\theta_2}{\partial A}Q_{x\theta}\right)=-\frac{x_2^*Q_A}{M_{22}}\left(\frac{Q_{xA}}{Q_A}-\frac{Q_{x\theta}}{Q_\theta}\right).$$

This expression implies that the long-run behavior of the overconfident firm is determined by its steady-state belief about the demand slope: If $\frac{Q_{xA}}{Q_A}-\frac{Q_{x\theta}}{Q_\theta}>0$ so that the firm is optimistic about the demand slope, then it produces more than in the correctly-specified model, and obtains a higher profit. This happens in example (13) where the firm is persistently overconfident about the demand slope.

On the other hand, if $\frac{Q_{xA}}{Q_A}-\frac{Q_{x\theta}}{Q_\theta}<0$ so that the firm is pessimistic about the demand slope, then it produces less and earns a lower profit. This happens in example (12) where the overconfident firm becomes pessimistic about the demand slope through learning. This discussion leads to the following corollary:

Corollary 2. *Consider the model of first-order misspecification. Suppose that there is a unique steady state at $A=a$ and it is regular.*[26] *Then at $A=a$, we have* $\text{sgn}\frac{\partial x_2^*}{\partial A}=\text{sgn}\left(-\frac{\partial x_1^*}{\partial A}\right)=\text{sgn}\frac{\partial \pi_2^*}{\partial A}=\text{sgn}\left(-\frac{\partial \pi_1^*}{\partial A}\right)=\text{sgn}\left(\frac{Q_{xA}}{Q_A}-\frac{Q_{x\theta}}{Q_\theta}\right)$, $|\frac{\partial x_2^*}{\partial A}|>|\frac{\partial x_1^*}{\partial A}|$, *and* $|\frac{\partial \pi_1^*}{\partial A}|>|\frac{\partial \pi_2^*}{\partial A}|$.

Double misspecification. Now we will relax the common knowledge assumption, and consider the case in which firm 1 is unaware of the rival firm's overconfidence. Specifically, consider double misspecification with $A_1=a$ and $A_2>a$, so that the firm's beliefs about the parameter a are the same as before, but each firm (incorrectly) believes that the other firm shares the same view about the parameter a.

How does the overconfident firm 2 behave in such a situation? Proposition 3 shows that the impact of one's overconfidence on its own steady-state action is the

[25] By the implicit function theorem, $\frac{\partial\theta_2}{\partial A}=-\frac{Q_A}{Q_\theta}<0$, and hence we indeed have $Q_A+\frac{\partial\theta_2}{\partial A}Q_\theta=0$.

[26] Because a steady state is an intersection of asymptotic best response correspondences BR_1 and BR_2, the steady state is unique if $BR_i'(x_j)\in(-1,1)$ for all i and x_j where $j\neq i$.

base misspecification effect, $-\frac{M_{2A}}{M_{22}} - BR'_{23}\frac{M_{3A}}{M_{33}}$, times the multipliers. As shown in Lemma 4 in Appendix B.7, when the game is symmetric, the sign of this base misspecification effect is the same as that of first-order misspecification.[27] So under double misspecification, firm 2 produces more than in the correctly-specified model if it does so in the case of first-order misspecification, and produces less if it does so in the case of first-order misspecification.

How about the behavior of the rival firm? A critical difference from the case of first-order misspecification is that the commitment effect does not exist in this environment; although firm 2's overconfidence influences its behavior, the opponent is not aware of it and does not best-respond to it. This in particular implies that if we look at the one-shot game, firm 2's overconfidence does not influence the opponent's behavior, and hence never be beneficial.

However, in our dynamic model, firm 2's overconfidence can still improve the equilibrium payoff. Indeed, we have the following result:

Corollary 3. *Consider the model of double misspecification with $A_1 = a$. Suppose that there is a unique steady state at $A_2 = a$ and it is regular. Then at $A_2 = a$, all the results stated in Corollary 2 still hold.*

So in the long-run, first-order misspecification and double misspecification lead to similar steady-state outcomes, i.e., firm 1's unawareness about firm 2's overconfidence does not have a significant impact on their long-run behavior. In particular, when $\frac{Q_{xA}}{Q_A} - \frac{Q_{x\theta}}{Q_\theta} > 0$, firm 2's overconfidence improves its long-run equilibrium payoff, regardless of whether the opponent is aware of it.

Why do we have such a result, even though the commitment effect does not exist under double misspecification? A key is that firm 1's long-run behavior is influenced by incorrect learning. To illustrate the idea, suppose that the base misspecification effect is positive, so that firm 2 produces more than in the correctly-specified model. Under double misspecification, firm 1 is not aware of it, and hence observes prices which are systematically lower than the anticipation. Accordingly, firm 1 becomes pessimistic about θ and produces less; this yields a

[27]Intuitively, this happens because the impact $|BR'_{23}\frac{M_{3A}}{M_{33}}|$ of second-order misspecification is smaller than the impact $|\frac{M_{2A}}{M_{22}}|$ of first-order misspecification in symmetric games.

higher profit to the overconfident firm 2 *even in the absence of the commitment effect*.

4.2 Team production

In the Cournot example, we have seen that one's unawareness about the opponent's overconfidence *does not* have a significant impact on the equilibrium outcome. However, this result does not hold in general; there are many economic examples in which first-order misspecification and double misspecification lead to qualitatively different outcomes. In this subsection, we will present one of such examples: a team production problem.

Consider two players working on a joint project. Each period, player $i = 1, 2$ chooses an effort level x_i, and observes a stochastic output $y = Q(x, a, \theta^*) + \varepsilon$ where a is a fixed parameter (e.g., one's capability) and θ^* is an unknown fundamental. We assume that Q is twice-continuously differentiable, $Q_{x_i} > 0$, $Q_a > 0$, and $Q_\theta > 0$. Player i's payoff is $y - c(x_i)$, where $c(x_i)$ is the effort cost satisfying $c' > 0$ and $c'' > 0$. Assume also that there is a unique Nash equilibrium in the one-shot game. This setup is fairly general, and includes the following examples as special cases.[28]

Example 1. Let a_i denote player i's capability, and let $a = a_1 + a_2$ denote the total capability. Let $c_i(x_i) = x_i^2$ and

$$Q = \theta(x_1 + x_2 + kx_1x_2 + a),$$

where $k \in (-\frac{2}{\theta^*}, \frac{2}{\theta^*})$ is a fixed parameter.[29] Note that efforts are complements if $k > 0$, and substitutes if $k < 0$. Each player i may have a bias and incorrectly believe that the total capability is $A_i \neq a$. When $A_i > a$, it represents one's overconfidence. When $A_i < a$, it represents one's underconfidence or prejudice about

[28] Again, these examples satisfy the regularity condition for first-order misspecification with $A = a$ and for double misspecification with $A_1 = A_2 = a$, and the sufficient conditions for convergence stated in Propositions 2 and 4. The proof can be found in the working paper version (Murooka and Yamamoto, 2021).

[29] This assumption ensures that the equilibrium is an interior point.

the opponent's capability. Players learn the profitability θ of the business over time. This setup corresponds to a multi-player version of Example 2 of Heidhues, Kőszegi, and Strack (2018).

Example 2. Let a denote player 1's capability. Let $c_i(x_i) = x_i^2$ and

$$Q = \theta(ax_1 + x_2 + kx_1x_2 + 2),$$

where $k \in (-\frac{2}{\theta^*}, \frac{2}{\theta^*})$ is a fixed parameter. A difference from Example 1 is that player 1's capability a influences her marginal productivity, which makes the function Q asymmetric, in that $Q(x_1, x_2, a, \theta) \neq Q(x_2, x_1, a, \theta)$ for $a \neq 1$. As will be explained, this property has a qualitative impact on the steady-state outcome under double misspecification.

First-Order Misspecification. Again, we start with the benchmark case in which player 2 has first-order misspecification, in that she incorrectly believes $A \neq a$. Simple algebra shows that the base misspecification effect of first-order misspecification is

$$-\frac{1}{M_{22}}\left(\underbrace{\overbrace{Q_{x_iA}}^{\text{direct effect}} + \overbrace{\frac{\partial\theta_2}{\partial A}Q_{x_i\theta}}^{\text{indirect effect}}}_{\text{on marginal productivity}}\right). \tag{15}$$

So the base misspecification effect is determined by the biased player 2's subjective view about the marginal productivity in the steady state; the term Q_{x_iA} measures how player 2's misspecification influences her view about the marginal productivity, and the term $\frac{\partial\theta_2}{\partial A}Q_{x_i\theta}$ measures how the incorrect learning modifies it.

In what follows, we will assume $Q_{x_2A} \leq 0$ and $Q_{x_2\theta} > 0$, i.e., the marginal return Q_{x_i} of the misspecified player is negatively correlated with the capability, and positively correlated with the fundamental. It is easy to check that this assumption is satisfied in Examples 1 and 2 above. Under this assumption, both the direct effect and the indirect effect in (15) are negative. Thus the overconfident player 2 works less in the static model, and even less in the long run (so we have

$\frac{\partial x_2^*}{\partial A} < 0$). Intuitively, when a player is overconfident about her own capability, she observes outputs systematically lower than the anticipation. Accordingly, as time goes, she becomes pessimistic about the state θ and reduces the effort.[30]

How about the equilibrium payoffs? There are two cases to be considered. First, suppose that $Q_{x_1x_2} > 0$ so that efforts are complements. (In the above examples, this corresponds to $k > 0$.) Then the rational player 1 *reduces* the effort as a response to the lower effort of the overconfident player, which reduces the overconfident player's payoff. So one's overconfidence is detrimental in this case.

Next, suppose that $Q_{x_1x_2} < 0$ so that efforts are substitutes. (In the above examples, this corresponds to $k < 0$.) In this case, one's overconfidence is beneficial; indeed, the rational player 1 *increases* the effort as a response, which improves the overconfident player's payoff. The mechanism here is exactly the same as the commitment effect in the Cournot model, i.e., one's misspecification may influence the opponent's behavior, which may improve the misspecified player's payoff. This result is in a sharp contrast with the single-agent case studied by Heidhues, Kőszegi, and Strack (2018), where one's overconfidence is always detrimental. In their model, the overconfident player reduces the effort just as in our model, but it never improves the overconfident player's payoff due to the lack of the commitment effect.

Double Misspecification. Now we consider the case of double misspecification, where player 1 is not aware of player 2's overconfidence. We will focus on Examples 1 and 2 above, and show that first-order misspecification and double

[30] Of course, the argument here can be extended to a more general setup. The bottom-line is that the short-run effect of player 2's overconfidence on her own action is determined by Q_{x_2A}, and the long-run effect is determined by $Q_{x_2A} + \frac{\partial \theta_2}{\partial A} Q_{x_2\theta} = Q_{x_2A} - \frac{Q_A}{Q_\theta} Q_{x_2\theta}$. For example, suppose that $Q < 0$ is the damage from drought and agents invest to irrigation which mitigate the damage, and it takes a form of

$$Q = -\frac{1}{\theta}\left(\frac{1}{x_1 + x_2} + \frac{1}{a}\right).$$

In this case, $Q_{x_iA} \geq 0$ and $Q_{x_i\theta} < 0$, so both the direct effect and the indirect effect are positive. This means that player 2's overconfidence about her capability *increases* her effort in the one-shot game, and she makes even more effort in the long run.

misspecification can have opposite effects on players' actions and payoffs.

First, consider Example 1. This game is symmetric, in that $Q(x_1,x_2,a,\theta) = Q(x_2,x_1,a,\theta)$ and $u_1(x_1,y) = u_2(x_1,y)$. As discussed in the Cournot model, in such a case, the sign of the base misspecification effect $-\frac{M_{2A}}{M_{22}} - BR'_{23}\frac{M_{3A}}{M_{33}}$ of double misspecification coincides with that of first-order misspecification. So the overconfident player 2 reduces the effort (i.e., $\frac{\partial x_2^*}{\partial A_2} < 0$) just as in the case of first-order misspecification.

How does it influence player 1's action? Under double misspecification, player 1's long-run behavior is influenced by her incorrect learning: Player 1 is not aware of the overconfident player reducing the effort, and thus observes outputs lower than the anticipation on average. This makes her pessimistic about the state over time, and she reduces the effort in the end. So in this example, one's overconfidence lowers both players' efforts and payoffs.

Note that this result does not rely on whether efforts are complements or substitutes. This is in a sharp contrast with the case of first-order misspecification, where an agent's overconfidence improves her profit when the efforts are substitutes. That is, first-order misspecification and double misspecification have opposite effects on the overconfident player's payoff (and player 1's action) when $k < 0$.

Next, consider Example 2. Here the difference between first-order misspecification and double misspecification is even more striking, in that the base misspecification effects of these two misspecifications can have opposite signs. For example, let $\Theta = [0.1, 0.3]$, $\theta^* = 0.2$, $a = A_1 = 1$, and $k = 4$. Then at $A_2 = a$, the base misspecification effect of double misspecification is *positive*, and thus the overconfident player 2 increases the effort.[31] Intuitively, in the case of double misspecification, player 2 (incorrectly) believes that player 1 is overconfident about a and makes higher effort than in the reality. Because efforts are complements (recall $k = 4 > 0$), with this perception, player 2 makes more effort. Given the

[31] In this example, we have $M_{2A} = -\frac{1}{44}$, $M_{22} = M_{33} = -\frac{47}{22}$, $BR'_{23} = \frac{63}{235}$, and $M_{3A} = \frac{39}{220}$. Hence, the base misspecification effect of double misspecification is about 0.012, which is positive as claimed in the main text.

specified parameters, this effect dominates all the other effects coming from the misspecification, so the overconfident player 2 makes higher effort in the steady state.

4.3 Bias Transmission and Self-fulfilling Prophecies

Recent evidence suggests that the gender gap in math achievement arises from culture and social conditioning rather than from biological reasons (such as brain functioning). For example, Lavy and Sand (2018) and Carlana (2019) find that the gender gap in performance in math exam substantially increases when students are assigned to math teachers with stronger gender stereotypes. In particular, Carlana (2019) argues that this effect is at least partially driven by lower self-confidence on math ability of female students who are exposed to gender-biased teachers.[32] We will show that our framework is useful to explain such a *bias transmission* from teachers to students.[33]

Suppose that player 1 (she) is a student and player 2 (he) is a teacher. The student's achievement (e.g., math test performance) is given by $y = a(x_1 + x_2 + b) + \varepsilon$, where $a > 0$ represents a gender-specific capability and $b > 0$ is the student's own capability. The student knows her own capability b, but does not know the gender-specific capability θ_1. So she thinks that the outcome is given by $y = \theta_1(x_1 + x_2 + b) + \varepsilon$ and learns θ_1 over time. On the other hand, the teacher has a biased view about the gender specific capability, and he thinks that the outcome is given by $y = A_2(x_1 + x_2 + \theta_2) + \varepsilon$, where $A_2 < a$ represents his bias. He does not know the student's individual capability θ_2, and learns it over time. We assume that each player (incorrectly) thinks that the opponent has the same view about the

[32] Relatedly, Gong, Lu, and Song (2018) report that a male math teacher in their survey is more likely to question and praise male students than female students compared with a female math teacher, and that having such a teacher lowers female students' beliefs about gender-specific capability: They find that female students who have a male math teacher are more likely to agree with a question "boys are more talented in learning math than girls" than those who have a female math teacher.

[33] See Giuliano (2020) for a survey of the transmissions of gender-biased norms and beliefs.

world. This means that the student is not aware of the teacher's gender-stereotype $A_2 < a$.

This setup is different from the one presented in Section 3, in that different players learn different parameters. However, this does not has a substantial impact on the property of the steady state. Similarly to the analysis in Section 4.2, it is straightforward to show that *both* the teacher and the student exert less effort than in the correctly specified model in the steady state, and the student becomes underestimating the gender-specific capability θ_1.

A notable feature in this framework is that the student initially has an unbiased view about the environment, but nonetheless, the teacher's gender bias is eventually transmitted to the student.[34] A key driving force is the student's inferential naivety; a biased teacher secretly reduces the effort, and the student is not aware of it. Then on average, the realized outcomes are lower than the student's expectation, which makes the student unrealistically pessimistic about the gender-specific capability θ_1.

It seems that bias transmissions and self-fulfilling prophecies are as prevalent in workplaces as they are in school classrooms. Livingston (1969) finds that a manager's high expectation of the subordinates improves the subordinates' job performance, and it is confirmed by many subsequent papers.[35] This phenomenon is known as the *Pygmalion effect* in management, and Eden (1984) and Davidson and Eden (2000) argue that this effect stems from bias transmission: a manager's

[34]Recent work by Heidhues, Kőszegi, and Strack (2020) also argue that a gender bias (more generally, a group discrimination) can endogenously arise as a consequence of misspecified learning. Formally, they develop a single-agent learning model, and show that an underconfident (resp. overconfident) agent tends to underestimate (resp. overestimate) the capability of her in-group members. So in their setup, the source of a group discrimination is one's misconfidence about her own capability. Our result complements their work by considering the case in which an agent does not have underconfidence, or more generally, any bias about the physical environment. Our analysis shows that one's existing prejudice may induce *other players'* negative self-stereotypes through learning.

[35]Kierein and Gold (2000) and McNatt (2000) provide meta-analysis results of the Pygmalion effect in management. See Bertrand and Duflo (2017) for a survey of the self-fulfilling prophecies in economics.

higher expectation raises workers' beliefs about their own self-efficacy, which lead them to greater motivation and achievement.[36] Our framework here is useful to better understand why such a bias transmission occurs. Roughly speaking, if an optimistic manager makes an extra effort and a worker is not (fully) aware of it, then on average, the worker observes an output better than his expectation. This makes the worker more confident of his own capability, which in turn improves the outcome further.

Of course, if a manager is negatively biased, the mechanism above can work in the opposite direction. For example, Hoobler, Wayne, and Lemmon (2009) report that in many industries, managers tend to think that female workers are unfit for promotion compared to male workers. Our analysis suggests that such a bias can be transmitted to female workers, and they become underconfident about their own capabilities. This is consistent with the recent work by Born, Ranehill, and Sandberg (forthcoming) who find that women are less confident than men in their relative ability as being a leader position. This mechanism may help understand why a "glass ceiling," an invisible barrier that discourages women and minorities, persists in various institutions.

5 Related Literature and Conclusion

There is a rapidly growing literature on Bayesian learning with model misspecification. Nyarko (1991) presents an example 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. Heidhues, Kőszegi, and Strack (2018), Heidhues, Kőszegi, and Strack (2021), and

[36]Of course, there are other mechanisms which explain the correlation between a manager's bias and a worker's performance. For example, Glover, Pallais, and Pariente (2017) find that minority workers in French grocery stores tend to perform worse when they work with biased managers, while working with biased managers does not activate self-stereotyping of the minority workers. Glover, Pallais, and Pariente (2017) argue that this result can be explained by the fact that biased managers are less comfortable around minorities: such managers do not monitor minority workers frequently and do not ask them to stay after the end of their shifts.

He (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 robustness of steady states. All these papers look at a single-agent problem and focus on first-order misspecification.

More recently, Ba and Gindin (2021) consider two-player team production in which both players are overconfident about their own capability. They show that if efforts are complements and information has a positive externality, then learning is mutually reinforcing, i.e., one's strategic play reduces *both* players' efforts and results in a worse outcome. Our work strengthens their result, in three ways. First, our Proposition 1 gives a necessary and sufficient condition for mutually-reinforcing learning: When the base misspecification effect is negative, our proposition shows that player 1's strategic play reduces both players' efforts if and only if the two asymptotic best response curves are upward-sloping (i.e., $BR_1' > 0$ and $BR_2' > 0$).[37] Second, our Proposition 1 does not impose any assumptions on payoffs and information structure, and allows us to study a wide range of applications such as Cournot duopoly and tournaments. Third and most importantly, we develop a model of higher-order misspecification and study how each type of misspecification influences players' beliefs and actions.

Misspecified learning has also been studied in other settings. In the literature on social learning, many papers study how inferential naivety or model misspecification influences the asymptotic outcomes (e.g., DeMarzo, Vayanos, and Zwiebel, 2003; Eyster and Rabin, 2010; Gagnon-Bartsch and Rabin, 2016; Bohren and Hauser, 2021; Frick, Iijima, and Ishii, 2020). Molavi (2020) considers a general equilibrium model in which a representative agent has a misspecified view about the world. Cho and Kasa (2017) study an asset-pricing model in which an agent incorrectly believes that the environment is not stationary.

[37] Strategic complementarity and positive information externality assumed by Ba and Gindin (2021) imply upward-sloping asymptotic best response curves.

A prominent example of the first-order misspecification is overconfidence on one's own capability. Plenty of experimental and empirical papers report that people exhibit overconfidence on their own ability in various economic activities, as discussed in Section 1.[38] Furthermore, recent empirical evidence suggests that overconfidence on a particular aspect of one's own capability persists even after a long time and a plenty of feedback (Grossman and Owens, 2012; Hoffman and Burks, 2020; Huffman, Raymond, and Shvets, 2019), which calls for the analysis of long-run behavior under model misspecifications.

References

Aghion, P., P. Bolton, C. Harris, and B. Jullien (1991): "Optimal Learning by Experimentation" *Review of Economic Studies*, 58 (4), 621-654.

Ba, C. and A. Gindin (2021): "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.

Benaïm, M., J. Hofbauer, and S. Sorin (2005): "Stochastic approximations and differential inclusions," *SIAM Journal on Control and Optimization* 44, 328-348.

Benoît, J.- P. and Dubra, J. (2011): "Apparent Overconfidence," *Econometrica* 79, 1591-1625.

Benoît, J.- P., Dubra, J. and Moore, D.A. (2015): "Does the Better-than-Average Effect Show that People are Overconfident?: Two Experiments," *Journal of the European Economic Association* 13, 293-329.

[38] Benoît and Dubra (2011) point out that most of these works may not be sufficient to conclude that people have overconfidence, because it can also be explained by considering a model in which people do not have overconfidence but face uncertainty about some aspect of the environment. However, this explanation is not supported by Benoît, Dubra, and Moore (2015); they conduct a laboratory experiment which separates out the effect of uncertainty, and find that subjects still exhibit overconfidence.

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

Bertrand, M. and E. Duflo (2017): "Field Experiments on Discrimination," In A.V. Banerjee & E. Duflo, (Eds.), *Handbook of Economic Field Experiments*, 1, 309-393, North Holland.

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

Bohren, J.A. and D.N. Hauser (2021): "Learning with Model Misspecification: Characterization and Robustness," forthcoming in *Econometrica*.

Born A., E Ranehill, and A. Sandberg (forthcoming): "Gender and Willingness to Lead – Does the Gender Composition of Teams Matter?," *Review of Economics and Statistics*.

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

Carlana, M. (2019): "Implicit Stereotypes: Evidence from Teachers' Gender Bias," *Quarterly Journal of Economics*, 134 (3), 1163-1224.

Chevalier, A., S. Gibbons, A. Thorpe, M. Snell, and S. Hoskins (2009): "Students' Academic Self-Perception." *Economics of Education Review*, 28 (6): 716–27.

Cho, I-K. and K. Kasa (2017): "Gresham's Law of Model Averaging," *American Economic Review*, 107 (11), 3589-3616.

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

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.

Davidson, O.B. and D. Eden (2000): "Remedial Self-Fulfilling Prophecy: Two Field Experiments to Prevent Golem Effects Among Disadvantaged Women," *Journal of Applied Psychology*, 85 (3), 386-398.

Eden, D. (1984): "Self-Fulfilling Prophecy as a Management Tool: Harnessing Pygmalion," *Academy of Management Review*, 9 (1), 64-73.

Englmaier, F. (2010): "Managerial Optimism and Investment Choice," *Managerial and Decision Economics*, 31 (4), 303-310.

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.

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.

Giuliano, P. (2020): "Gender and Culture," *Oxford Review of Economic Policy*, 36, (4), 944–961.

Glover, D., A. Pallais, and W. Pariente (2017): "Discrimination as a Self-Fulfilling Prophecy: Evidence from French Grocery Stores," *Quarterly Journal of Economics* 132 (3), 1219-1260.

Gong, J., Y. Lu, and H. Song (2018): "The Effect of Teacher Gender on Students' Academic and Noncognitive Outcomes," *Journal of Labor Economics* 36 (3), 743-778.

Grossman, Z. and D. Owens (2012): “An Unlucky Feeling: Overconfidence and Noisy Feedback,” *Journal of Economic Behavior & Organization* 84 (2), 510-524.

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

He, K. (2021): “Mislearning from Censored Data: The Gambler ’ s Fallacy and Other Correlational Mistakes in Optimal-Stopping Problems,” forthcoming in *Theoretical Economics*.

Heidhues, P. and B. Kőszegi (2018): “Behavioral Industrial Organization,” In D.B. Bernheim, S. DellaVigna & D. Laibson, (Eds.), *Handbook of Behavioral Economics: Foundations and Applications 1*, 517-612, North Holland.

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

Heidhues, P., B. Kőszegi, and P. Strack (2020): “Overconfidence and Prejudice,” Working Paper.

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

Heifetz, A., C. Shannon, and Y. Spiegel (2007): “The Dynamic Evolution of Preferences,” *Economic Theory*, 32 (2), 251-286.

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

Hoffman, M. and S.V. Burks (2020): “Worker Overconfidence: Field Evidence and Implications for Employee Turnover and Firm Profits,” *Quantitative Economics* 11 (1), 315-348.

Hoobler, J.M., S.J. Wayne, and G. Lemmon (2009): “Bosses’ Perceptions of Family-Work Conflict and Women’s Promotability: Glass Ceiling Effects,” *Academy of Management Journal*, 52 (5), 939-957.

Huffman, D., C. Raymond, and J. Shvets (2019): “Persistent overconfidence and biased memory: Evidence from managers,” Working Paper.

Kierein, N. M. and M.A. Gold (2000): “Pygmalion in Work Organizations: A Meta-Analysis,” *Journal of Organizational Behaviour*, 21, 913-928.

Kőszegi, B. (2014): “Behavioral Contract Theory,” *Journal of Economic Literature*, 52 (4), 1075-1118.

Kyle, A. S. and F. A. Wang (1997): “Speculation Duopoly with Agreement to Disagree: Can Overconfidence Survive the Market Test?,” *Journal of Finance*, 52 (5), 2073–2090.

Lavy, V. and E. Sand (2018): “On the Origins of Gender Gaps in Human Capital: Short- and Long-Term Consequences of Teachers’ Biases,” *Journal of Public Economics*, 167, 263-279.

Lazear, E.P. and S. Rosen (1981): “Rank-Order Tournaments as Optimum Labor Contracts,” *Journal of Political Economy*, 89 (5), 841-864.

Livingston, J.S. (1969): “Pygmalion in Management,” *Harvard Business Review*, 47, 81-89.

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

Madarász, K. (2012): “Information Projection: Model and Applications,” *Review of Economic Studies*, 79 (3), 961-985.

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.

McNatt, D.B. (2000): “Ancient Pygmalion Joins Contemporary Management: A Meta-Analysis of the Result,” *Journal of Applied Psychology*, 85, 314-322.

Molavi, P. (2020): “Macroeconomics with Learning and Misspecification: A General Theory and Applications,” Working paper.

Murooka, T and Y. Yamamoto (2021): “Multi-Player Bayesian Learning with Misspecified Models,” OSIPP Discussion Paper DP-2021-E-001.

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

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

Tirole, J. (1988): *The Theory of Industrial Organization*, MIT Press.

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.

Appendices

A Asymptotic Behavior of Misspecified Players

In this appendix, we will provide 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. Then we show that the motion of players' beliefs is also approximated by a solution to a differential inclusion. This result is new, and we use it to derive a sufficient condition for belief convergence.

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 Propisition 7. The second property is crucial in order to use a stochastic approximation technique in the proof of Proposition 8. 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.[39] In the general model here, we will impose a (similar but) weaker assumption:

[39]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.

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.[40]

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 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

[40]This assumption is needed to establish Propositions 7 and 8. 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.

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:

$$\hat{s}_{i,k}(\hat{\mu}_i) \in \arg\max_{\hat{x}_{i,k} \in \hat{X}_{i,k}} \int_{\Theta_{i,k}} U_{\theta_{i,k}}(\hat{x}_{i,k}, \hat{s}_{i,k+1}(\hat{\mu}_i))\hat{\mu}_{i,k}(d\theta_{i,k}) \quad \forall k \forall \hat{\mu}_i. \tag{16}$$

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

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

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

$$\lim_{t\to\infty} \int_{\Theta_{i,k}} (K_{i,k}(\theta_{i,k}, \sigma^t(h)) - K^*_{i,k}(\sigma^t(h)))\hat{\mu}^{t+1}_{i,k}(h)[d\theta_{i,k}] = 0. \tag{17}$$

Let $\mathscr{H}$ denote the set of sample paths h which satisfy (17). By Proposition 5, $P^{\hat{s}}(\mathscr{H}) = 1$.

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

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

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 (17), 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 6. *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

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

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 7. *$S_\varepsilon(\sigma)$ 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 6 and 7 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

$$\dot{\boldsymbol{\sigma}}(t) \in \triangle S_0(\boldsymbol{\sigma}(t)) - \boldsymbol{\sigma}(t). \tag{18}$$

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 8 below, the differential inclusion (18) 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 (18).[41]

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

[41] There is also a technical reason: In the proof of Proposition 8, 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 (18) with an initial value* $\sigma \in \triangle\hat{X}$ if it is absolutely continuous in all compact intervals, $\boldsymbol{\sigma}(0) = \sigma$, and (18) 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 (18) has a solution. (See Theorem 9 of Deimling (1992) on page 117.) Let $Z(\sigma)$ denote the set of all these solutions.

Proposition 8. *Pick a Markov strategy $\hat{s}$. Then for any $T > 0$ and any sample path $h \in \mathscr{H}$,*

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

A.3.2 Steady State and Generalized Berk-Nash Equilibrium

$\sigma \in \triangle\hat{X}$ is a *steady state* of the differential inclusion (18) 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 9. *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 (18).*

In all the examples in this paper, we assume that the agents are myopically optimal so that the strategy profile $\hat{s}$ satisfies (16). In this special case, steady states of our differential inclusion are *generalized Berk-Nash equilibria* in the following sense:

Definition 3. 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 $\hat{\mu}$. We will discuss more on this later.

Proposition 10. *Suppose that the strategy profile $\hat{s}$ satisfies (16). Then any steady state of our differential inclusion (18) 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 5, 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) 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.

We view GBNE as a natural extension of BNE of Esponda and Pouzo (2016) to our setup. For comparison, let us think about the information structure considered in Esponda and Pouzo (2016), that is, suppose that the subjective signal distribution $Q_{\theta_{i,k}}(\cdot|x)$ is independent of the opponent's action (i.e., $Q_{\theta_{i,k}}(\cdot|x) = Q_{\theta_{i,k}}(\cdot|x_i)$ for each i, k, and x). In this special case, all higher-level agents (i,k) with $k \geq 2$ are irrelevant, in the sense that they do not influence the actions and the beliefs of the actual player $(i,1)$. This is so because each player i's subjective expected payoff $U_{\theta_{i,1}}(x_i) = \sum_{y\in Y} Q_{\theta_{i,1}}(y|x_i)u_i(x_i,y)$ and her Bayes' formula $\mu_i^{t+1}(\theta_{i,1}) = \frac{\mu_i^t(\theta_{i,1})q_{\theta_{i,1}}(y^t|x_i^t)}{\int_{\Theta_{i,1}} \mu_i^t(\theta_{i,1})q_{\theta_{i,1}}(y^t|x_i^t)d\theta_{i,1}}$ are independent of the opponent j's action. Accordingly, GBNE reduces to a probability measure $\sigma \in \triangle(X_1 \times X_2)$ on the set $X_1 \times X_2$ such that for each pure action profile $x = (x_1, x_2)$ in the support of σ and for each i, there is a belief $\mu_i \in \triangle\Theta_{i,1}(\sigma)$ such that

$$x_i \in \arg\max_{x_i'} \int_{\Theta_{i,1}} U_{\theta_{i,1}}(x_i')\mu_i(d\theta_{i,1})$$

where $\Theta_{i,1}(\sigma)$ is the minimizers of the Kullback-Leibular divergence

$$\int_X \int_Y \log \frac{q(y|x)}{q_{\theta_{i,1}}(y|x)} Q(dy|x)\sigma(dx).$$

It is easy to check that GBNE is a weakening of BNE of Esponda and Pouzo (2016), that is, any GBNE σ is a BNE. In particular, degenerate BNE is equivalent to pure-strategy BNE, in that σ is a pure-strategy BNE if and only if it is a degenerate GBNE. However, a non-degenerate GBNE need not be a mixed-strategy BNE, because in a GBNE, (i) different actions x_i may be supported by different beliefs μ_i, and (ii) GBNE distribution σ allows correlation between x_1 and x_2. This difference comes from the fact that GBNE is a limit point of the action frequency, while BNE is a limit point of the action itself. Specifically, in Esponda and Pouzo (2016), there is an i.i.d. payoff perturbation each period, so that each player (independently) mix actions each period. A mixed-strategy BNE σ is regarded as a limit point of this mixed action. In this case, in a steady state, the mixed strategy σ_i must be optimal (with a payoff perturbation) given a single belief, and there is no correlation between actions of different players. In contrast, in our model, each player chooses a pure action, so there is perfect correlation between x_1 and x_2. Also as noted earlier, different action profiles x and x' which appear in non-degenerate GBNE σ are played in different periods t and t'; so they are supported by different beliefs μ^t and $\mu^{t'}$.

A.4 Motion of the KL Minimizer

A.4.1 Identifiability and Differential Inclusion

Our Proposition 8 shows that the asymptotic motion of the action frequency σ^t is described by the differential inclusion (18). However, solving the differential inclusion (18) 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 $|\frac{\partial K_{i,k}(\theta_{i,k},\hat{x})}{\partial \theta_{i,k,m}} - \frac{\partial K_{i,k}(\theta_{i,k},\hat{x}')}{\partial \theta_{i,k,m}}| < L|\hat{x}-\hat{x}'|$ 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

$$\dot{\boldsymbol{\theta}}(t) \in \bigcup_{\sigma:\theta(\sigma)=\boldsymbol{\theta}(t)} \bigcup_{\sigma' \in \triangle S_0(\boldsymbol{\theta}(t))} -(\nabla^2 K(\boldsymbol{\theta}(t), \sigma))^{-1} \left(\nabla K(\boldsymbol{\theta}(t)), \sigma'\right). \tag{19}$$

Let $Z_{\theta}(\boldsymbol{\theta}(0))$ be the set of solutions to the differential inclusion (19) with the initial value $\boldsymbol{\theta}(0)$.

Proposition 11. *Suppose that Assumptions 3 and 4 hold. Then for any $T > 0$ and any sample path $h \in \mathscr{H}$,*

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

To interpret the differential inclusion (19), 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 (18), we have

$$\dot{\boldsymbol{\theta}}_{i,k}(t) \in \bigcup_{\sigma:\theta(\sigma)=\boldsymbol{\theta}(t)} \bigcup_{\sigma' \in \triangle S_0(\boldsymbol{\theta}(t))} -\frac{K'_{i,k}(\boldsymbol{\theta}_{i,k}(t), \sigma')}{K''_{i,k}(\boldsymbol{\theta}_{i,k}(t), \sigma)} \tag{20}$$

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 reaosns. 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 (20), 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 (18).

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 Proofs

B.1 Proof of Proposition 1

Pick x^* and A^* as stated. Since the steady-state actions (x_1^*, x_2^*) are interior points, they must satisfy the first-order conditions

$$\frac{\partial U_1(x_1, x_2^*, \theta^*)}{\partial x_1} = 0, \tag{21}$$

$$\left.\frac{\partial U_2(x_1^*, x_2, \theta)}{\partial x_2}\right|_{\theta = \theta_2(x^*, A)} = 0. \tag{22}$$

Let M be the Jacobian of this system of the equations. Then each ij-component of the matrix coincides with M_{ij} defined in the main text. That is,

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$$

Since $BR_1' BR_2' \neq 1$, we have $\det M \neq 0$, so the implicit function theorem guarantees that for any parameter A close to A^*, there is an action profile x^* which satisfies the first-order conditions (21) and (22). These action profiles are globally optimal (i.e., maximize the expected payoff given the belief $\theta_1 = \theta^*$ and $\theta_2(x^*, A)$), because of the regularity conditions (i) and (ii). So this x^* is a steady state given the parameter A. The implicit function theorem also asserts that

$$\begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{bmatrix} \frac{\partial x_1^*}{\partial A} \\ \frac{\partial x_2^*}{\partial A} \end{bmatrix} = - \begin{bmatrix} 0 \\ M_{2A} \end{bmatrix},$$

Solving this system of equations,

$$\frac{\partial x_2^*}{\partial A} = -\frac{M_{11} M_{2A}}{\det M},$$
$$\frac{\partial x_1^*}{\partial A} = \frac{M_{12} M_{2A}}{\det M}.$$

Dividing both the numerator and denominator of the first equation by $M_{11}M_{22}$ and using $\det M = M_{11}M_{22} - M_{12}M_{21}$, we have $\frac{\partial x_2^*}{\partial A} = -\frac{1}{1 - BR_1' BR_2'} \frac{M_{2A}}{M_{22}}$. Also by combining the two equations above, we have $\frac{\partial x_1^*}{\partial A} = BR_1 \frac{\partial x_2^*}{\partial A}$.

Next, we prove $BR_1' BR_2' < 1$ by contradiction. Suppose that $BR_1' BR_2' > 1$. Then we have either (i) $BR_1' > 0$ and $BR_2' > 0$, or (ii) $BR_1' < 0$ and $BR_2' < 0$. Consider case (i). Then we have $BR_2' > \frac{1}{BR_1'} > 0$. This means that if we take x_1 on the horizontal axis and x_2 on the vertical axis, then the two asymptotic best response curves are upward-sloping at the steady state action x^*, and BR_2 is steeper than BR_1. This and the continuity of BR_i imply that BR_1 and BR_2 must intersect at some $x_1 > x_1^*$, but this contradicts with the fact that x^* is a unique steady state. The same argument works for case (ii). Hence, we have $BR_1' BR_2' \leq 1$. Also, dividing both sides of $\det M \neq 0$ by $M_{11}M_{22}$, we have $BR_1' BR_2' \neq 1$. *Q.E.D.*

B.2 Proof of Corollary 1

Immediate from Proposition 1. *Q.E.D.*

B.3 Proof of Proposition 2

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 5 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 11 is useful; it shows that the motion of θ_2^t is asymptotically approximated by the differential inclusion (20), which reduces to the one-dimensional differential inclusion

$$\dot{\theta}_2(t) \in \bigcup_{\sigma:\theta(\sigma)=\theta(t)} -\frac{K_2'(\theta_2(t), s(1_{\theta^*}, 1_{\theta_2(t)}))}{K_2''(\theta_2(t), \sigma)} \tag{23}$$

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 intial value $\theta \in \Theta$.

We will consider the following two cases separately.

B.3.1 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 (23) 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.[42] Hence there is $T > 0$ such that

$$\boldsymbol{\theta}_2(t) \geq \theta' + \varepsilon \tag{24}$$

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 11, there is t^* such that for any $t > t^*$, $\boldsymbol{\theta}_2 \in Z_\theta(\boldsymbol{w}_\theta(t))$, and $s \in [0, 2T]$,

$$|\boldsymbol{w}_\theta(t+s) - \boldsymbol{\theta}_2(s)| < \frac{\varepsilon}{2}. \tag{25}$$

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 (24), we have

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

for any $s \geq T$ and for any solution $\boldsymbol{\theta} \in Z_\theta(\boldsymbol{w}_\theta(t^{**}))$. This inequality and (25) 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 (24) 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 (25) 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.

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

B.3.2 Case 2: $\liminf_{t\to\infty}\theta^t_{i,k}(h) = \limsup_{t\to\infty}\theta^t_{i,k}(h)$.

In this case, $\lim_{t\to\infty}\theta^t_{i,k}(h)$ exists. Let $\theta^*_{i,k} = \lim_{t\to\infty}\theta^t_{i,k}(h)$. 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 (24) 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 11, there is t^* such that (25) 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 $\lim_{t\to\infty}\theta^t_{i,k}(h) \geq \theta^*_{i,k} + \frac{\varepsilon}{2}$, which is a contradiction. *Q.E.D.*

B.4 Proof of Proposition 3

Pick A^* and x^* as stated. Since x^* is an interior point, it must satisfy the first-order conditions

$$\frac{\partial U_1(x_1, \hat{x}^*_2, \theta_1)}{\partial x_1} = 0, \tag{26}$$

$$\frac{\partial U_2(\hat{x}^*_1, x_2, \theta_2)}{\partial x_2} = 0, \tag{27}$$

$$\frac{\partial \hat{U}_1(\hat{x}_1, x^*_2, \theta_2)}{\partial \hat{x}_1} = 0, \tag{28}$$

$$\frac{\partial \hat{U}_2(x^*_1, \hat{x}_2, \theta_1)}{\partial \hat{x}_2} = 0. \tag{29}$$

Let M be the Jacobian of this system of the equations. Then each ij-component of the matrix coincides with M_{ij} defined in the main text.

By the regularity condition (iii), $\det M \neq 0$, so the implicit function theorem guarantees that for any parameter A_2 close to A^*_2, there is an action profile x^* which satisfies the first-order conditions (26)-(29). These action profiles are globally optimal, because of the regularity conditions (i) and (ii). So this x^* is a steady state given the parameter A. The implicit function theorem also asserts that

$$\begin{bmatrix} M_{11} & M_{12} & M_{13} & M_{14} \\ M_{21} & M_{22} & M_{23} & M_{24} \\ M_{31} & M_{32} & M_{33} & M_{34} \\ M_{41} & M_{42} & M_{43} & M_{44} \end{bmatrix} \begin{bmatrix} \frac{\partial x_1^*}{\partial A_2} \\ \frac{\partial x_2^*}{\partial A_2} \\ \frac{\partial \hat{x}_1^*}{\partial A_2} \\ \frac{\partial \hat{x}_2^*}{\partial A_2} \end{bmatrix} = - \begin{bmatrix} 0 \\ M_{2A} \\ M_{3A} \\ 0 \end{bmatrix},$$

Solving this system and using $M_{13} = M_{24} = M_{34} = M_{43} = 0$,

$$\begin{aligned} \frac{\partial x_2^*}{\partial A_2} &= \frac{(M_{14}M_{41}M_{33} - M_{11}M_{33}M_{44})M_{2A} + (M_{11}M_{23}M_{44} - M_{23}M_{14}M_{41})M_{3A}}{\det M} \\ \frac{\partial x_1^*}{\partial A_2} &= \frac{(M_{12}M_{33}M_{44} - M_{33}M_{42}M_{14})M_{2A} - (M_{23}M_{44}M_{12} - M_{23}M_{42}M_{14})M_{3A}}{\det M}. \end{aligned}$$

Dividing both the numerator and the denominator of the first equation by $M_{11}M_{22}M_{33}M_{44}$,

$$\begin{aligned} \frac{\partial x_2^*}{\partial A_2} &= - \left\{ (1 - BR_{14}BR_{41}) \frac{M_{2A}}{M_{22}} + (BR_{23} - BR_{23}BR_{14}BR_{41}) \frac{M_{3A}}{M_{33}} \right\} \frac{M_{11}M_{22}M_{33}M_{44}}{\det M} \\ &= - (1 - BR_{14}BR_{41}) \frac{M_{11}M_{22}M_{33}M_{44}}{\det M} \left(\frac{M_{2A}}{M_{22}} + BR_{23} \frac{M_{3A}}{M_{33}} \right). \end{aligned}$$

Note that

$$\begin{aligned} \det M = & M_{11}M_{22}M_{33}M_{44} + M_{14}M_{21}M_{33}M_{42} - M_{14}M_{22}M_{33}M_{41} - M_{12}M_{21}M_{33}M_{44} \\ & - M_{11}M_{23}M_{32}M_{44} - M_{14}M_{23}M_{31}M_{42} + M_{14}M_{23}M_{32}M_{41} + M_{12}M_{23}M_{31}M_{44}, \end{aligned}$$

so

$$\begin{aligned} \frac{M_{11}M_{22}M_{33}M_{44}}{\det M} &= \frac{1}{(1 - BR'_{14}BR'_{41})(1 - BR'_{23}BR'_{32}) - (BR'_{12} + BR'_{14}BR'_{42})(BR'_{21} + BR'_{23}BR'_{31})} \\ &= \left(\frac{1}{(1 - BR'_{14}BR'_{41})(1 - BR'_{23}BR'_{32})} \right) \left(\frac{1}{1 - NE'_1 NE'_2} \right) \end{aligned}$$

Plugging this into the equation above, we obtain the first equation in the proposition. The second equation can be derived in a similar way. *Q.E.D.*

B.5 Proof of 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 5 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 11 shows that the motion of these KL minimizers, (θ_1^t, θ_2^t), is approximated by the differential inclusion (20), which can be rewritten as the two dimensional problem

$$\left(\frac{d\boldsymbol{\theta}_1(t)}{dt}, \frac{d\boldsymbol{\theta}_2(t)}{dt}\right) \in \bigcup_{\sigma:\theta(\sigma)=\boldsymbol{\theta}(t)} \left(-\frac{K_1'(\boldsymbol{\theta}_1(t), s(\boldsymbol{\theta}_1(t), \boldsymbol{\theta}_2(t)))}{K_1''(\boldsymbol{\theta}_2(t), \boldsymbol{\sigma})}, -\frac{K_2'(\boldsymbol{\theta}_2(t), s(\boldsymbol{\theta}_1(t), \boldsymbol{\theta}_2(t)))}{K_2''(\hat{\boldsymbol{\theta}}_1(t), \boldsymbol{\sigma})}\right) \tag{30}$$

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 (30) 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 (30): It shows that $\boldsymbol{\theta}_2(t)$ moves toward $f_2(\boldsymbol{\theta}_1(t))$ at any time t.

Lemma 3. *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 (30). 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 $\boldsymbol{\theta}_2(t) > f_2(\hat{\boldsymbol{\theta}}_1(t))$ 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: $\boldsymbol{\theta}_2(t) < \overline{\theta}$. An argument similar to that in Case 1 shows that at $\theta_2 = \overline{\theta}$, we have $K_2'(\overline{\theta}, s(\boldsymbol{\theta}_1(t), \overline{\theta})) > 0$. On the other hand, by the assumption, $K_2'(\boldsymbol{\theta}_2(t), s(\boldsymbol{\theta}_1(t), \boldsymbol{\theta}_2(t))) < 0$. Then since $K_2'(\boldsymbol{\theta}, s(\boldsymbol{\theta}_1(t), \boldsymbol{\theta}))$ is continuous in $\boldsymbol{\theta}$, there must be $\boldsymbol{\theta} \in (\boldsymbol{\theta}_2(t), \overline{\theta})$ such that $K_2'(\boldsymbol{\theta}, s(\boldsymbol{\theta}_1(t), \boldsymbol{\theta})) = 0$. This implies that $\theta \in f_2(\boldsymbol{\theta}_1)$, but it contradicts with the uniqueness of $f_2(\boldsymbol{\theta}_1)$. *Q.E.D.*

Now we will construct a Lyapunov function V to show that any solution to the differential inclusion (30) 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

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

We will show that given any initial value $\boldsymbol{\theta}(0)$ and given any solution $\boldsymbol{\theta}$ to the differential inclusion (19),

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

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

Case 1: $|\boldsymbol{\theta}_2(t)| > |\kappa\boldsymbol{\theta}_1(t)|$. Assume first that $\boldsymbol{\theta}_2(t) > 0$. Then by the definition of κ and $f_2(0) = 0$, we have $f_2(\boldsymbol{\theta}_1(t)) < |\kappa\boldsymbol{\theta}_1(t)| < \boldsymbol{\theta}_2(t)$. Then from Lemma 3 and $\boldsymbol{\theta}_2(t) > 0$, we have $\dot{V}(\boldsymbol{\theta}(t)) = \dot{\boldsymbol{\theta}}_2(t) < 0$.

Assume next that $\boldsymbol{\theta}_2(t) < 0$. By the definition of κ and $f_2(0) = 0$, we have $f_2(\hat{\boldsymbol{\theta}}_1(t)) > -|\kappa\hat{\boldsymbol{\theta}}_1(t)| > \boldsymbol{\theta}_2(t)$. Then from Lemma 3 and $\boldsymbol{\theta}_2(t) < 0$, we have $\dot{V}(\boldsymbol{\theta}(t)) = -\dot{\boldsymbol{\theta}}_2(t) < 0$.

Case 2: $|\boldsymbol{\theta}_2(t)| < |\kappa\boldsymbol{\theta}_1(t)|$. An argument similar to those for Case 1 shows that $\dot{V}(\boldsymbol{\theta}(t)) < 0$.

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

B.6 Proof of Corollary 2

Consider the infinite-horizon model with first-order misspecification. Let M_{ii}, M_{ij} be the ones in Proposition 1. Then we have

$$M_{ii} = 2Q_x + x_i Q_{xx} - c'' < 0,$$
$$M_{ij} = Q_x + x_i Q_{xx} < 0,$$
$$-\frac{M_{2A}}{M_{22}} = -\frac{x_2^* Q_A}{M_{22}} \left(\frac{Q_{xA}}{Q_A} - \frac{Q_{x\theta}}{Q_\theta} \right).$$

The first two inequalities imply $M_{ii} < M_{ij} < 0$, and hence we have $BR_1' = -\frac{M_{12}}{M_{11}} \in (-1,0)$ and $\frac{1}{1-BR_1' BR_2'} > 1$. Thus it follows from Proposition 1 that $\text{sgn}\frac{\partial x_2^*}{\partial A} = \text{sgn}\left(-\frac{\partial x_1^*}{\partial A}\right) = \text{sgn}\left(\frac{Q_{xA}}{Q_A} - \frac{Q_{x\theta}}{Q_\theta}\right)$ and $|\frac{\partial x_2^*}{\partial A}| > |\frac{\partial x_1^*}{\partial A}|$.

For payoffs, note that at $A = a$, we have

$$\frac{\partial \pi_i^*}{\partial A} = \frac{\partial \pi_i^*}{\partial x_i}\frac{\partial x_i^*}{\partial A} + \frac{\partial \pi_i^*}{\partial x_{-i}}\frac{\partial x_{-i}^*}{\partial A} = \frac{\partial \pi_i^*}{\partial x_{-i}}\frac{\partial x_{-i}^*}{\partial A} = x_i^* Q_x \frac{\partial x_{-i}^*}{\partial A},$$

where the second inequality follows from $\frac{\partial \pi_i^*}{\partial x_i} = 0$. Since $x_1^* = x_2^*$ at $A = a$ and $Q_x < 0$, we have $\text{sgn}\frac{\partial x_2^*}{\partial A} = \text{sgn}\frac{\partial \pi_2^*}{\partial A} = \text{sgn}\left(-\frac{\partial \pi_1^*}{\partial A}\right)$ and $|\frac{\partial \pi_1^*}{\partial A}| > |\frac{\partial \pi_2^*}{\partial A}|$. *Q.E.D.*

B.7 Proof of Corollary 3

We first prove a lemma which is useful to analyze a symmetric game, where $X_1 = X_2$, $u_1(x_1, y) = u_2(x_2, y)$ for all x_1 and x_2 with $x_1 = x_2$, and $Q(x_1, x_2, a, \theta) = Q(x_2, x_1, a, \theta)$. Let x_i^{correct} and π^{correct} denote firm i's steady-state action and payoff in the correctly specified model. Let x^{first} denote the steady-state action profile for first-order misspecification, where player 1 believes that the true parameter is $A_1 = a$ and player 2 believes that the true parameter is $A_2 \neq a$. Likewise, let x^{double} denote the steady-state action profile for double misspecification with $A_1 = a$ and $A_2 \neq a$. The following lemma relates these two steady states when player 2's misspecification is small.

Lemma 4. *Consider a symmetric game with $x_1^{\text{correct}} = x_2^{\text{correct}}$. Suppose that in the case of first-order misspecification with $A = a$, there is a unique steady state and it is regular. Suppose also that in the case of double misspecification with*

$A_1 = A_2 = a$, there is a unique steady state and it is regular. Then in the case of double misspecification with $A_1 = A_2 = a$, we have $NE_i' \in (-1,1)$, and

$$\frac{\partial x_2^{\text{double}}}{\partial A_2} = \frac{(U_{ii}-L)(U_{ii}+U_{ij}-L)}{U_{ii}(U_{ii}+U_{ij}-2L)}\left(1-\frac{U_{ij}-L}{U_{ii}-L}\right)\frac{\partial x_2^{\text{first}}}{\partial A},$$
$$\frac{\partial x_1^{\text{double}}}{\partial A_2} = -\frac{L}{U_{ii}+U_{ij}-L}\frac{\partial x_2^{\text{double}}}{\partial A_2},$$

where $U_{ii} = \frac{\partial^2 U_1(x^{\text{correct}},\theta^*)}{\partial x_1^2}$, $U_{ij} = \frac{\partial^2 U_1(x^{\text{correct}},\theta^*)}{\partial x_1 \partial x_2}$, $L = \frac{\partial \hat{\theta}_1}{\partial x_1}\frac{\hat{U}_1}{\partial \hat{x}_1 \partial \theta} = \frac{Q_{x_1}}{Q_\theta}\cdot\frac{\partial^2 U_1(x^{\text{correct}},\theta)}{\partial x_1 \partial \theta}$. We also have $\operatorname{sgn}\frac{\partial x_2^{\text{double}}}{\partial A_2} = \operatorname{sgn}\frac{\partial x_2^{\text{first}}}{\partial A}$ and $\operatorname{sgn}\frac{\partial x_1^{\text{double}}}{\partial A_2} = \operatorname{sgn}\frac{\partial x_2^{\text{double}}}{\partial A_2}L$.

Proof. We first prove $NE_i' \in (-1,1)$. Note that $x_1 = x_2 = \hat{x}_1 = \hat{x}_2 = x_i^{\text{correct}}$ constitutes a steady state at $A_1 = A_2 = a$. Then we must have $|NE_1'NE_2'| \leq 1$ at $x^{\text{double}} = (x_1^{\text{correct}}, x_2^{\text{correct}})$; the proof is very similar to that of $BR_1'BR_2' \leq 1$ in the proof of Proposition 1, and hence omitted. (We only need to replace BR_i' in the proof if Proposition 1) with NE_i'.) Also, the regularity condition $\det M \neq 0$ implies $|NE_1'NE_2'| \neq 1$. Accordingly, we have $|NE_1'NE_2'| = |NE_i'| < 1$, which implies $NE_i' \in (-1,1)$.

Let $L_i = \frac{Q_{x_1}}{Q_\theta}\frac{\partial^2 U_i}{\partial x_i \partial \theta}$. When $A_1 = A_2 = a$, the multiplier effect on $\frac{\partial x_2^{\text{double}}}{\partial A_2}$ appearing in the proof of Proposition 3, $(1 - BR_{14}BR_{41})\frac{M_{11}M_{22}M_{33}M_{44}}{\det M}$, can be rewritten as

$$\frac{1-BR_{14}BR_{41}}{(1-BR_{14}'BR_{41}')(1-BR_{23}'BR_{32}')-(BR_{12}'+BR_{14}'BR_{42}')(BR_{21}'+BR_{23}'BR_{31}')}$$
$$=\frac{1-\frac{U_{12}-L_1}{U_{11}}\frac{U_{21}}{U_{22}-L_2}}{\left(1-\frac{U_{12}-L_1}{U_{11}}\frac{U_{21}}{U_{22}-L_2}\right)\left(1-\frac{U_{21}-L_2}{U_{22}}\frac{U_{12}}{U_{11}-L_1}\right)-\left(-\frac{L_1}{U_{11}}+\frac{U_{12}-L_1}{U_{11}}\frac{L_2}{U_{22}-L_2}\right)\left(-\frac{L_2}{U_{22}}+\frac{U_{21}-L_2}{U_{22}}\frac{L_1}{U_{11}-L_1}\right)}$$
$$=\frac{U_{22}(U_{11}-L_1)(U_{11}U_{22}-U_{11}L_2-U_{21}U_{12}+U_{21}L_1)}{(U_{11}U_{22}-U_{11}L_2-U_{21}U_{12}+U_{21}L_1)(U_{11}U_{22}-U_{22}L_1-U_{12}U_{21}+U_{12}L_1)-(U_{12}L_1-U_{22}L_1)(U_{21}L_1-U_{11}L_2)}.$$

When the game is symmetric, this reduces to

$$
\begin{aligned}
&\frac{U_{ii}(U_{ii}-L)(U_{ii}-U_{ij})(U_{ii}+U_{ij}-L)}{(U_{ii}-U_{ij})^2(U_{ii}+U_{ij}-L)^2-L^2(U_{ii}-U_{ij})^2}\\
=&\frac{U_{ii}(U_{ii}-L)(U_{ii}+U_{ij}-L)}{(U_{ii}-U_{ij})(U_{ii}^2+U_{ij}^2+2U_{ii}U_{ij}-2U_{ii}L-2U_{ij}L)}\\
=&\frac{U_{ii}(U_{ii}-L)(U_{ii}+U_{ij}-L)}{(U_{ii}-U_{ij})(U_{ii}+U_{ij})(U_{ii}+U_{ij}-2L)}\\
=&\frac{U_{ii}^2}{U_{ii}^2-U_{ij}^2}\cdot\frac{(U_{ii}-L)(U_{ii}+U_{ij}-L)}{U_{ii}(U_{ii}+U_{ij}-2L)}\\
=&\frac{1}{1-BR_1'BR_2'}\cdot\frac{(U_{ii}-L)(U_{ii}+U_{ij}-L)}{U_{ii}(U_{ii}+U_{ij}-2L)}.
\end{aligned}
$$

Similarly, when the game is symmetric, the base misspecification effect on $\frac{\partial x_2^{\text{double}}}{\partial A_2}$ appearing in the proof of Proposition 3, $\left(\frac{M_{2A}}{M_{22}}+BR_{23}\frac{M_{3A}}{M_{33}}\right)$, can be rewritten as

$$
\left(1-\frac{U_{ij}-L}{U_{ii}-L}\right)\frac{M_{2A}}{M_{22}}.
$$

These results and Proposition 3 imply the first equation in the proposition. Also, the second equation follows from

$$
\begin{aligned}
NE_1' =&\frac{BR_{12}'+BR_{14}'BR_{42}'}{1-BR_{14}'BR_{41}'}=\frac{-\frac{L}{U_{ii}}+\frac{U_{ij}-L}{U_{ii}}\frac{L}{U_{ii}-L}}{1-\frac{U_{ij}-L}{U_{ii}}\frac{U_{ij}}{U_{ii}-L}}=\frac{-L(U_{ii}-L)+L(U_{ij}-L)}{U_{ii}(U_{ii}-L)-U_{ij}(U_{ij}-L)}\\
=&\frac{L(U_{ii}-U_{ij})}{(U_{ii}+U_{ij}-L)(U_{ii}-U_{ij})}=\frac{L}{U_{ii}+U_{ij}-L}.
\end{aligned}
$$

Next, we will show $\text{sgn}\frac{\partial x_2^{\text{double}}}{\partial A_2}=\text{sgn}\frac{\partial x_2^{\text{first}}}{\partial A}$. Recall that

$$
\frac{\partial x_2^{\text{double}}}{\partial A_2}=\frac{(U_{ii}-L)(U_{ii}+U_{ij}-L)}{U_{ii}(U_{ii}+U_{ij}-2L)}\left(1-\frac{U_{ij}-L}{U_{ii}-L}\right)\frac{\partial x_2^{\text{first}}}{\partial A}.
$$

Note that $U_{ii}=M_{11}<0$ and $U_{ii}-L=M_{33}<0$ under the regularity condition, Also $U_{ii}+U_{ij}-L<0$ because if not and $U_{ii}+U_{ij}-L>0$,

$$\begin{aligned} NE_i' \in (-1,1) &\Leftrightarrow -1 < \frac{L}{U_{ii}+U_{ij}-L} < 1 \\ &\Leftrightarrow -(U_{ii}+U_{ij}-L) < L < U_{ii}+U_{ij}-L \\ &\Rightarrow U_{ii}+U_{ij} > 0, \end{aligned}$$

which contradicts with $U_{ii}<0$ and $|BR_i'| = |\frac{U_{ij}}{U_{ii}}| < 1$. Similarly, $U_{ii}+U_{ij}-2L<0$ because

$$\begin{aligned} NE_i' \in (-1,1) &\Leftrightarrow -1 < \frac{L}{U_{ii}+U_{ij}-L} < 1 \\ &\Leftrightarrow U_{ii}+U_{ij}-L < L < -(U_{ii}+U_{ij}-L) \\ &\Rightarrow U_{ii}+U_{ij}-2L < 0, \end{aligned}$$

So the term $\frac{(U_{ii}-L)(U_{ii}+U_{ij}-L)}{U_{ii}(U_{ii}+U_{ij}-2L)}$ appearing in the above display is positive. Similarly, the term $1-\frac{U_{ij}-L}{U_{ii}-L}$ is positive, because $U_{ii}<0$ and $|BR_i'| = |\frac{U_{ij}}{U_{ii}}| < 1$ imply

$$U_{ii}-U_{ij}<0 \Leftrightarrow (U_{ii}-L)-(U_{ij}-L)<0 \Leftrightarrow 1-\frac{U_{ij}-L}{U_{ii}-L}>0$$

where the last inequality uses $U_{ii}-L<0$. Hence we have $\text{sgn}\frac{\partial x_2^{\text{double}}}{\partial A_2} = \text{sgn}\frac{\partial x_2^{\text{first}}}{\partial A}$ as desired. Finally, $\text{sgn}\frac{\partial x_1^{\text{double}}}{\partial A_2} = \text{sgn}\frac{\partial x_2^{\text{double}}}{\partial A}L$ directly follows from $NE_i' = \frac{L}{U_{ii}+U_{ij}-L}$ and $U_{ii}+U_{ij}-L<0$. *Q.E.D.*

Let U_{ii}, U_{ij}, and L be as stated in Lemma 4. That is,

$$\begin{aligned} U_{ii} &= \frac{\partial^2 U_i}{\partial^2 x_i} = 2Q_x + x_i Q_{xx} - c'' < 0, \\ U_{ij} &= \frac{\partial^2 U_i}{\partial x_i \partial x_j} = Q_x + x_i Q_{xx} = Q_x + x_i Q_{xx} < 0, \\ L &= \frac{Q_{x_i}}{Q_\theta}\frac{\partial^2 U_i}{\partial x_i \partial \theta} = \frac{Q_x}{Q_\theta}(x_i Q_{x\theta} + Q_\theta) < 0. \end{aligned}$$

Then, Lemma 4 and the argument similar to the proof of Corollary 2 imply the result. *Q.E.D.*

B.8 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 5. *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

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

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 5, 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.*

B.9 Proof of Proposition 5

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 6. *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.*

B.10 Proof of Proposition 7

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 7. *$B_\varepsilon(\sigma)$ 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.*

B.11 Proof of Proposition 8

The proof is very similar to that of Theorem 2 of EPY. 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 (18)).

Our Propositions 6 and 7 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.[43] *Q.E.D.*

B.12 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

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

[43] 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. 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 difficluty, 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.

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

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

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,

$$\frac{d\theta(\sigma_\beta)}{d\beta}=-(\nabla^2K_{i,k}(\theta_{i,k}(\sigma_\beta),\sigma_\beta))^{-1}\left(\nabla K_{i,k}(\theta_{i,k}(\sigma_\beta),\sigma)-\nabla K_{i,k}(\theta_{i,k}(\sigma_\beta),\sigma')\right). \tag{31}$$

Using the fundamental theorem of calculus, we have

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

Then by the definition of L_1 and L_4,

$$|\theta(\sigma)-\theta(\tilde{\sigma})|\leq\int_0^1 L_1L_4\|\sigma-\sigma'\|d\beta=L\|\sigma-\sigma'\|.\quad \textit{Q.E.D.}$$

B.13 Proof of Proposition 11

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 (18) starting from this $\boldsymbol{\sigma}(0)$. Let $\boldsymbol{\theta}(t) = \theta(\boldsymbol{\sigma}(t))$ for each t. The following lemma shows that $\{\boldsymbol{\theta}(t)\}_{t\geq 0}$ solves (19).

Lemma 8. *Pick $t \geq 0$ such that (18) holds. Then $\dot{\boldsymbol{\theta}}(t)$ exists and satisfies (19).*

Proof. Pick t as stated, and pick $\boldsymbol{\sigma}^* \in \triangle S_0(\boldsymbol{\sigma}(t))$ such that $\dot{\boldsymbol{\sigma}}(t) = \boldsymbol{\sigma}^* - \boldsymbol{\sigma}(t)$. Let $\sigma_\beta = \beta\boldsymbol{\sigma}^* + (1-\beta)\boldsymbol{\sigma}(t)$ 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 (19). 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 (19).

Note first that $\lim_{\varepsilon\to 0}\frac{\theta(\sigma_\varepsilon)-\theta(\sigma_0)}{\varepsilon}$ exists and is in the right-hand side of (19). Indeed, from (31),

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

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

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

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 the continuity of θ and the compactness of $\triangle\hat{X}$), 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 8, there is t^* such that for any $t > t^*$, there is $\boldsymbol{\sigma} \in Z(\boldsymbol{w}(h)[t])$ 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 (19).[44] This implies the result we want. *Q.E.D.*

[44] Note that $\boldsymbol{\theta}$ is absolutely continuous because $\boldsymbol{\sigma}$ is absolutely continuous and $\theta(\sigma)$ is Lipschitz continuous. Also from Lemma 8, $\boldsymbol{\theta}$ satisfies (19) for almost all t.