Bayesian Learning with Bounded Rationality: Convergence to $\varepsilon-$Nash Equilibrium[1]

Yuichi Noguchi

Department of Economics

Kanto Gakuin University

1-50-1 Mutsu-ura-higashi, Kanazawa-ku

Yokohama 236-8501

Japan

E-mail: ynoguchi@kanto-gakuin.ac.jp

First version: May 21, 2006

This version: April 22, 2007

[1]Very preliminary and very incomplete. This paper benefits from conversations with Drew Fudenberg, John H. Nachbar, and H. Peyton Young. I also thank seminar participants at the 17th International Conference on Game Theory at Stony Brook 2006 for their helpful comments.

Abstract

We provide a class of prior beliefs that (almost surely) lead to playing approximate Nash equilibrium, combined with bounded rationality, i.e., smooth approximate optimal behaviors, in *any* infinitely repeated game with perfect monitoring: converging to $\varepsilon-$Nash equilibrium for any (finite normal form) stage game, any discount factors (less than one), and any $\varepsilon > 0$. Furthermore, the class of prior beliefs is smart in the sense that, for any learnable set of opponents strategies, a prior belief in the class $\varepsilon-$weakly merges with all opponents strategies in the learnable set. We also argue the implications of our positive result to impossibility results (Nachbar (1997, 2005) and Foster and Young (2001)). Specifically, we point out that the impossibility in Nachbar (1997, 2005) is obtained because the learnability condition in Nachbar (1997, 2005) requires *uniform* learning such that each player's prior belief weakly merges with opponents *true* strategies *uniformly* in his own *various* strategies, including his true one, and that the impossibility in Foster and Young (2001) crucially depends on perfect rationality, i.e., exact optimal behaviors.

1 Introduction

We provide a class of prior beliefs that (almost surely) lead to playing *approximate* Nash equilibrium, combined with *bounded* rationality, i.e., smooth approximate optimal behaviors, in *any* infinitely repeated game with perfect monitoring: converging to $\varepsilon-$Nash equilibrium for any (finite normal form) stage game, any discount factors (less than one), and any $\varepsilon > 0$. Furthermore, the class of prior beliefs is smart in the sense that, for *any* learnable set of opponents strategies, a prior belief in the class $\varepsilon-$weakly merges with *all* opponents strategies in the learnable set. We also argue the implications of our positive result to impossibility results (Nachbar (1997, 2005) and Foster and Young (2001)). Specifically, we point out that the impossibility in Nachbar (1997, 2005) is obtained because the learnability condition in Nachbar (1997, 2005) requires *uniform* learning such that each player's prior belief weakly merges with opponents various strategies, including opponents true ones, uniformly in his own various strategies, including his true one, and that the impossibility in Foster and Young (2001) crucially depends on *perfect* rationality, i.e., exact optimal behaviors.

Convergence problem has been in trouble for Bayesian learning in repeated games since its study started. Kalai and Lehrer (1993, 1994) and others have contributed to foundations of Bayesian learning in repeated games: formulating basic concepts (e.g., merging, etc.) that are appropriate to Bayesian learning in repeated games and providing characterization conditions (e.g., absolute continuity, etc.) for converging to Nash equilibrium. However, in general, any of those characterization conditions is not practically easy to check; indeed, there has been given no nontrivial example in which a characterization condition plays an important role in showing convergence. Therefore, it has been unknown whether convergence to Nash equilibrium is regular or exceptional in repeated

games. Sandroni (2000) provides a positive result in a specific example. But its purpose is to show emergence of coorperation (in the 2×2 coordination stage game) through Bayesian learning rather than to provide a general result of convergence. On the other hand, Nahchbar (1997, 2005) proposes a rather general result of impossibility for Bayesian learning in repeated games. Nachbar shows that in any repeated game under a certain weak condition, if players are able to learn to predict sufficiently *various* opponents strategies and their learning abilities are symmetric, then at least one of players cannot learn to predict his opponents *true* strategies (i.e., opponents (approximate) optimal ones to their prior beliefs) uniformly with any of his own *various* strategies, including his own true one. As Nachbar (2005) admits, this negative result does *not* immediately imply the impossibility of convergence, but it certainly shows that in general, it is not easy to obtain convergence to even *approximate* Nash equilibrium. Furthermore, Foster and Young (2001) show that under *perfect* rationality (i.e., exact optimal behaviors), any given prior beliefs *cannot* converge to Nash equilibrium for almost all stage games near the matching pennies one: there exist *no* prior beliefs such that the prior beliefs learn to play Nash equilibrium in *any* stage game. In other words, (at least under perfect rationality) it is impossible to obtain a general result of learning to play Nash equilibrium.

This paper gives a general *positive* result of learning to play *approximate* Nash equilibrium, provided that players are *boundedly rational* in the sense that they take smooth approximate optimal behaviors. That is, we construct a class of prior beliefs that converge to approximate Nash equilibrium in *any* infinitely repeated game (with perfect monitoring). Furthermore, our class of prior beliefs are smart in the sense that for any prior belief, there exists our prior belief in the class such that our prior belief (apporoximately) learns to predict *all* opponents strategies that the given prior belief learns to predict. This result

implies that we do *not* have to give up learnability at least in any approximate sense in order to obtain convergence to approximate Nash equilibrium.

The point of this paper is how to construct prior beliefs for our purpose. The construction of our prior beliefs are based on two different research lines. The first research line is Foster and Young (2003)'s random search and testing. Foster and Young apply the method to a non-Bayesian learning model, but as will be shown, the method is also applied to Bayesian learning by introducing bounded rationality; more originally, Arthur (1994) proposes a similar but more intuitive mode of learning, which he calls inductive learning. The second research line is Noguchi (2005), which provides a characterization of learnable set of strategies. Making use of the concepts and technique in Noguchi (2005), we generalize the method of random search and testing so fully that our prior beliefs not only learn to predict as many strategies as possible but also converge to approximate Nash equilibrium in any repeated game.

Our positive result has the implications to the impossibilit results in Nachbar (1997, 2005) and Foster and Young (2001). We obtain that under *bounded* rationality (i.e., smooth approximate optimal behavior), it is fairly possible to learn to play *approximate* Nash equilibrium for *any* stage game and *any* discount factors. From this we conclude that Foster and Young's impossibility crucially depends on perfect rationality, so that their impossibility result is *not* robust to bounded rationality (i.e., approximate optimal behavior). In contrast to this, Nachbar's impossibility is quite robust in the sense that it holds even in the case of bounded rationality and "approximate learning," as will be explained. However, our positive result implies that for any learnable sets there exist prior beliefs such that each of the prior beliefs approximately learn to predict all opponents strategies in the learnable set and those prior beliefs converge to approximate Nash

equilibrium. It means that Nachbar impossibility is different from the impossibility of learning to play approximate Nash equilibrium in a general sense and that the diversity (and symmetry) of players' learnable sets does not prevent Bayesian learning from converging to approximate Nash equilibrium.

This paper is organized as follows. Section 2 describes the basic model and concepts. Section 3 explains main results in this paper. In Section 4, we construct prior beliefs. In Section 5, conditions for convergence are given. In Sections 6 and 7, we show that our prior beliefs converge to approximate Nash equilibrium. In Section 8, we prove that each of prior beliefs approximately learns to predict all opponents strategies in any learnable set. In Section 9, we discuss the implication of our results to the existing impossibility results.

2 The Model and Concepts

2.1 Basic model and notations

A group of players $i = 1, \cdots, I$ repeatedly play a stage game over infinite time horizon $t = 1, 2, \cdots$. Each player i takes a (pure) action a_i in a finite set A_i at each time, and let A denote the set of all action profiles: $A := \Pi_{i=1}^{I} A_i$. Given an action profile $a := (a_i)_i$, the stage game payoff for player i is denoted by $u_i(a)$. Let $\Delta(A_i)$ denote the set of all mixed actions over A_i and $\Delta(A)$ denote the set of all mixed action profiles, i.e., $\Delta(A) := \Pi_{i=1}^{I} \Delta(A_i)$. If a mixed action profile $\pi := (\pi_i)_i$ is played, then the stage game expected payoff for player i is defined by $u_i(\pi) := \sum_a \pi_1[a_1] \cdots \pi_I[a_I] u_i(a_1, \cdots, a_I)$. A history of the repeated game is a sequence of all players' actions. A *finite* history is denoted by h. When the length of a finite history is emphasized, we write h_T for a finite

history up to time T: $h_T := (a^1, \cdots, a^T)$. Let H_T denote the set of all finite histories with time length T. Let H designate the set of all finite histories, including the null history $h_0 := \emptyset$: $H := \bigcup_{T=0}^{\infty} H_T$, where $H_0 := \{h_0\}$. An *infinite* history is denoted by h_∞, and let $\mathbf{H}_\infty$ designate the set of all infinite histories. If a finite history h is an *initial* segment of a history h', then it is denoted by $h \leq h'$. When $h \leq h'$ and $h \neq h'$, it is designated by $h < h'$.

2.2 Behavior strategies

We assume *perfect monitoring*, i.e., every player observes the past history of realized actions of all players at each time. Therefore, the behavior of player i in the repeated game is represented by a *behavior strategy*, denoted by $\sigma_i : H \to \Delta(A_i)$. Let μ_σ designate the probability measure over $\mathbf{H}_\infty$ induced by playing a strategy profile $\sigma := (\sigma_1, \cdots, \sigma_I)$.

2.3 Bayesian learning

All players are Bayesian learners in the sense that each player has his *prior belief* about the other players' behavior strategies; every player knows that all players play (mixed) actions *independently* at each time. Then, as Kalai and Lehrer (1993) show, a prior belief of player i is formally represented by a profile of the other players' behavior strategies, denoted by $\tilde{\rho}^i := (\tilde{\rho}^i_j)_{j \neq i}$.[1]

2.4 Payoffs and bounded rationality

Given a strategy profile σ, the payoff for player i in the repeated game is the (averaged) expected discounted payoff sum $V_i(\sigma) := (1-\delta_i) \sum_{T=1}^{\infty} \delta_i^{T-1} \sum_{h \in H_{T-1}} u_i(\sigma(h)) \mu_\sigma(h)$, where

[1] $\tilde{\rho}^i_j$ is a behavior strategy of player $j(\neq i)$, i.e., $\tilde{\rho}^i_j : H \to \Delta(A_j)$.

δ_i is the discount factor of player i ($0 \leq \delta_i < 1$), $\sigma(h) := (\sigma_1(h), \cdots, \sigma_I(h))$, and $\mu_\sigma(h)$ is the probability of playing h. In the continuation game following a realized past history h, a *continuation* behavior strategy for player i is denoted by $\sigma_{i,h}$: $\sigma_{i,h}(h') := \sigma_i(h \cdot h')$ for all $h' \in H$, where $h \cdot h'$ is the concatenation of h and h'. The *continuation* payoff for player i following h is $V_i(\sigma \mid h) := V_i(\sigma_h)$, where $\sigma_h := (\sigma_{1,h}, \cdots, \sigma_{I,h})$.

The key assumption in this paper is that all players are "boundedly rational" in the sense that they take *smooth approximate* optimal behaviors against their prior beliefs. Specifically, we assume that each player i takes his strategy σ_i to maximize the following (averaged) expected discounted "perturbed" payoff sum against his prior belief $\tilde{\rho}^i$:

$$V_i^{v_i}(\sigma_i, \tilde{\rho}^i) := (1-\delta_i) \sum_{T=1}^{\infty} \delta_i^{T-1} \sum_{h \in H_{T-1}} [u_i(\sigma_i(h), \tilde{\rho}^i(h)) + v_i(\sigma_i(h))] \mu_{(\sigma_i, \tilde{\rho}^i)}(h),$$

where v_i is the payoff perturbation for player i. Payoff perturbation v_i is a smooth and strictly concave function from $Int(\Delta(A_i))$ to $\mathbb{R}$, and v_i also satisfies the *boundary condition* that $\|Dv_i(\pi_i)\| \to \infty$ as π_i approaches the boundary of $\Delta(A_i)$.[2] Furthermore, letting $\mid v_i \mid := \sup_{\pi_i} v_i(\pi_i)$, if $\mid v_i \mid$ is small, we say that player i's payoff perturbation is *small*. For simplicity, in the remaining of this paper we assume that payoff perturbations are *symmetric*.[3]

[2] $Int(\Delta(A_i))$ denotes the interior of $\Delta(A_i)$. $\|Dv_i(\pi_i)\|$ is the standard norm of the derivative $Dv_i(\pi_i)$ of v_i at π_i.

[3] Payoff perturbation v is symmetric if, for any mixed action $\pi := (\pi_1, \cdots, \pi_n)$ and any permutation $\varphi : \{1, \cdots, n\} \to \{1, \cdots, n\}$, $v(\pi_1, \cdots, \pi_n) = v(\pi_{\varphi(1)}, \cdots, \pi_{\varphi(n)})$. For example, the logistic function $-\frac{1}{\kappa} \sum_k \pi_k \log \pi_k$ is symmetric.

2.5 Smooth approximate optimal strategy

For any opponents strategy profile $\rho_{-i} := (\rho_j)_{j \neq i}$, there exists a *unique* smooth approximate optimal strategy to ρ_{-i}, denoted by σ_i^ρ: $\sigma_i^\rho := \arg\max_{\sigma_i} V_i^{v_i}(\sigma_i, \rho_{-i})$.[4] Essentially, all that is necessary for our argument is that, for each player i, there is a *uniform* lower bound on the probability of playing any action after any finite history. In fact, because of the boundary condition on v_i, σ_i^ρ has a uniform lower bound $\underline{l}_i$: for all i, there exists $\underline{l}_i > 0$ such that, for any ρ_{-i}, $\sigma_i^\rho(h)[a_i] \geq \underline{l}_i$ for all $a_i \in A_i$ and all $h \in H$.

2.6 $\varepsilon-$Nash equilibrium

We introduce *approximate* Nash equilibrium: for any $\varepsilon \geq 0$, we define $\varepsilon-$Nash equilibrium as follows.

Definition 1 *A strategy profile $\bar{\sigma}$ is called an $\varepsilon-$Nash equilibrium if, for all i and all σ_i,*

$$V_i(\bar{\sigma}_i, \bar{\sigma}_{-i}) + \varepsilon \geq V_i(\sigma_i, \bar{\sigma}_{-i}).$$

Especially, when $\varepsilon = 0$, $\bar{\sigma}$ is called a Nash equilibrium.

2.7 Conditioning rules and classes

We introduce a key concept to model the learning abilitiy of each player: *conditioning rules*. A conditioning rule represents an (approximate) regularity of opponents behavior strategies. Formally, a conditioning rule is a *finite partition* of H, denoted by $\mathcal{P}$. An element of $\mathcal{P}$ is called a *class* in $\mathcal{P}$, denoted by α. Note that a class is considered as a *subset* of H because it is an element of a partition of H. Also, we will often define a subset of H and call it a *class* by the abuse of language. When a realized history $h_{t-1} \in \alpha$, we

[4] For any h such that $\mu_{(\sigma_i, \rho_{-i})}(h) = 0$ (for all σ_i), $\sigma_{i,h}^\rho := \arg\max_{\sigma_i} V_i^{v_i}(\sigma_i, \rho_{-i,h})$.

say that time t is an $\alpha-$*active period* or that α is *active at time* t. For any player i's opponents strategy profile $\sigma_{-i} := (\sigma_j)_{j\neq i}$, we define its $\varepsilon-$*approximate* conditioning rule.

Definition 2 *A finite partition $\mathcal{P}_\varepsilon^{\sigma_{-i}}$ is called an $\varepsilon-$approximate conditioning rule of σ_{-i} if, for all $\alpha \in \mathcal{P}_\varepsilon^{\sigma_{-i}}$, all $h, h' \in \alpha$, and all $j \neq i$, $\|\sigma_j(h) - \sigma_j(h')\| \leq \varepsilon$.*[5] *Especially, when $\varepsilon = 0$, $\mathcal{P}_\varepsilon^{\sigma_{-i}}$ is called a conditioning rule of σ_{-i}.*

The definition says that mixed actions in active periods of each class α are almost the same. Note that, for *all* $\varepsilon > 0$, *any* opponents strategy profile σ_{-i} has its $\varepsilon-$approximate conditioning rule. In the remainder of this paper, we use the *maximum norm* on any set of mixed actions: $\|x\| := \max_a \mid x[a] \mid$.

Conversely, we may generate strategies from given conditioning rules.

Definition 3 *We say that σ_{-i} is generated by a set $\mathbb{P}$ of conditioning rules if, for all $\varepsilon > 0$, there exists $\mathcal{P} \in \mathbb{P}$ such that $\mathcal{P}$ is an $\varepsilon-$approximate conditioning rule of σ_{-i}.*

The definition says that, for all $\varepsilon > 0$, the regularity of σ_{-i} is $\varepsilon-$approximated by some conditioning rule in $\mathbb{P}$. Let $G(\mathbb{P})$ denote the set of all opponents strategies generated by $\mathbb{P}$. Note that *any* opponents strategy profile σ_{-i} is generated by any *countable* set $\{\mathcal{P}_{\varepsilon_n}^{\sigma_{-i}}\}_n$ of its approximate conditioning rules, i.e., $\sigma_{-i} \in G(\{\mathcal{P}_{\varepsilon_n}^{\sigma_{-i}}\}_n)$, where $\varepsilon_n \to 0$ as $n \to \infty$. Furthermore, σ_{-i} is generated by a conditioning rule $\mathcal{P}$, i.e., $\sigma_{-i} \in G(\mathcal{P})$ if and only if $\mathcal{P}$ is a conditioning rule of σ_{-i}: for all $\alpha \in \mathcal{P}$ and all $h, h' \in \alpha$, $\sigma_{-i}(h) = \sigma_{-i}(h')$.

Finally, since conditioning rules are (finite) partitions (of H), they are ordered with respect to *fineness*: if, for all $\alpha \in \mathcal{P}$ there exists $\beta \in \mathcal{Q}$ such that $\alpha \subset \beta$, we say that $\mathcal{P}$ is *finer* than $\mathcal{Q}$, denoted by $\mathcal{Q} \leq \mathcal{P}$. Clearly, $\leq$ is an order relation over the set of all conditioning rules. It is important to note that a *finer* conditioning rule generates *more*

[5] $\|\cdot\|$ is the maximum norm: $\|x\| := \max_a \mid x[a] \mid$.

opponents strategy profiles. Furthermore, when σ_{-i} is generated by $\mathcal{P}$ or equivalently $\mathcal{P}$ is a conditioning rule of σ_{-i}, $\sigma_{-i}(\alpha)(=(\sigma_j(\alpha))_{j\neq i})$ is well-defined for all $\alpha \in \mathcal{P}$: for all $j \neq i$, $\sigma_j(\alpha) := \sigma_j(h)$ for $h \in \alpha$. In addition, if $\mathcal{Q} \leq \mathcal{P}$ and σ_{-i} is generated by $\mathcal{Q}$, then σ_{-i} is also generated by $\mathcal{P}$; thus, for any $\beta \in \mathcal{Q}$, $\sigma_{-i}(\beta) = \sigma_{-i}(\alpha)$ for all $\alpha \in \mathcal{P}$ such that $\alpha \subset \beta$. We will make use of these ordering properties to a full extent for constructing smart prior beliefs.

3 Main Result

The main purpose of this paper is to provide a class of prior beliefs that almost surely lead to playing approximate Nash equilibrium in any infinitely repeated game with perfect monitoring. Furthermore, we show the result that our constructing prior beliefs are smart enough to approximately learn to predict as many opponents strategies as possible. To formalize our results, we introduce several concepts of learning: $\epsilon-$*weak merging* and *learnable set (correspondence).*

Definition 4 *We say that* $\mu_{(\sigma_i,\tilde{\rho}^i)}$ $\epsilon-$*weakly merges with* $\mu_{(\sigma_i,\sigma_{-i})}$ *or that* $\tilde{\rho}^i$ $\epsilon-$*learns to predict* σ_{-i} *with* σ_i *if, for all* $j \neq i$, $\limsup_{T\to\infty} \|\tilde{\rho}^i_j(h_T) - \sigma_j(h_T)\| \leq \epsilon$, $\mu_{(\sigma_i,\sigma_{-i})} - a.s.$ *Especially, when* $\epsilon = 0$*, we say that* $\mu_{(\sigma_i,\tilde{\rho}^i)}$ *weakly merges with* $\mu_{(\sigma_i,\sigma_{-i})}$ *or that* $\tilde{\rho}^i$ *learns to predict* σ_{-i} *with* σ_i*.*

Note that although the definition of $\epsilon-$weak merging only requires eventually making $\epsilon-$accurate predictions on *one period ahead* opponents actions, it implies eventually making $\epsilon-$accurate predictions on *any finite period future* opponents (mixed) actions. Next, learnable sets of strategies are defined as follows.

Definition 5 *Let $M_{-i}(\tilde{\rho}^i, \sigma_i)$ denote the set of all opponents strategies that $\tilde{\rho}^i$learns to predict with σ_i. $M_{-i}(\tilde{\rho}^i, \sigma_i)$ is called the $\tilde{\rho}^i-$learnable set with σ_i.*

Note that *set correspondence* $M_{-i}(\tilde{\rho}^i, \cdot) : \Sigma_i \to 2^{\Sigma_{-i}}$ completely represents the learning ability of prior belief $\tilde{\rho}^i$, where Σ_i is the set of all player i's strategies and $2^{\Sigma_{-i}}$ is the power set of all other players' strategy profiles. Conversely, we may define the concept of *learnable* set correspondence.

Definition 6 *Set correspondence $M_{-i} : \Sigma_i \to 2^{\Sigma_{-i}}$ is said to be learnable if there exists a prior belief $\tilde{\rho}^i$ such that, for all σ_i, $M_{-i}(\sigma_i) \subset M_{-i}(\tilde{\rho}^i, \sigma_i)$.*

Let us describe our main results. The first result is that, given any $\epsilon > 0$ and any prior beliefs $(\tilde{\rho}^i)_i$, we obtain prior beliefs $(\tilde{\rho}^i_*)_i$ such that each $\tilde{\rho}^i_*$ *not only* $\epsilon-$learns to predict all opponents strategies in $M_{-i}(\tilde{\rho}^i, \sigma_i)$ with all σ_i *but also* $\epsilon-$weakly merges with opponents' *true* strategies $(\sigma^*_j)_{j\neq i}$ (i.e., the smooth approximate optimal strategies to $(\tilde{\rho}^j_*)_{j\neq i}$): $\sigma^*_i := \arg\max_{\sigma_i} V_i^{v_i}(\sigma_i, \tilde{\rho}^i_*)$ for all i.

Theorem 1 *For any $\epsilon > 0$ and any prior beliefs $(\tilde{\rho}^i)_i$, there exist prior beliefs $(\tilde{\rho}^i_*)_i$ such that*

(1) for all i and all player i's strategies σ_i, $\tilde{\rho}^i_*$ $\epsilon-$learns to predict σ_{-i} with σ_i for all $\sigma_{-i} \in M_{-i}(\tilde{\rho}^i, \sigma_i)$,

(2) for any stage game payoffs $(u_i)_i$ and any discount factors $(\delta_i)_i$, there exists $\hat{v} > 0$ such that, for any (symmetric) payoff perturbations $(v_i)_i$ with $\mid v_i \mid \leq \hat{v}$ for all i, prior beliefs $(\tilde{\rho}^i_*)_i$ $\epsilon-$learn to predict players' true strategies $\sigma^* := (\sigma^*_i)_i$: with $\mu_{\sigma^*}-$probability one, there exists $\hat{T}$ such that, for all $T \geq \hat{T}$, all i, and all $j \neq i$,

$$\|\tilde{\rho}^i_{*,j}(h_T) - \sigma^*_j(h_T)\| \leq \epsilon.$$

Theorem 1 insists that even if Bayesian learners are (almost) as *smart* as possible, they can always (approximately) learn to predict each other *true* strategies as far as they are boundedly rational.

Remark 1 *Theorem 1 (1) implies that, for any learnable set correspondence* M_{-i}, $\tilde{\rho}^i_*$ $\epsilon-$*learns to predict* σ_{-i} *with* σ_i *for all* σ_i *and all* $\sigma_{-i} \in M_{-i}(\sigma_i)$.

The second result, i.e., convergence to approximate Nash equilibrium, is immediately obtained from Theorem 1 (2) by taking into considerations the variations of (maximum) payoffs by belief changes. Indeed, there is some bound on the variation rates of (maximum) payoffs by belief changes. Let $U := \max_i \max_a \mid u_i(a) \mid$ and $\bar{V}_i^{v_i}(\sigma_{-i}) := \max_{\sigma_i} V_i^{v_i}(\sigma_i, \sigma_{-i})$. Furthermore, let $\#A$ denote the number of all (pure) action profiles.

Lemma 1 *(1) For any* i *and any* $\sigma_{-i}, \sigma'_{-i}$,

$$\mid \bar{V}_i^{v_i}(\sigma_{-i}) - \bar{V}_i^{v_i}(\sigma'_{-i}) \mid \leq U\#A \sum_{T=0}^{\infty} \delta_i^T \max_{h \in H_T} \max_{j \neq i} \| \sigma_j(h) - \sigma'_j(h) \|,$$

(2) For any i *and any* σ_i *and any* $\sigma_{-i}, \sigma'_{-i}$,

$$\mid V_i^{v_i}(\sigma_i, \sigma_{-i}) - V_i^{v_i}(\sigma_i, \sigma'_{-i}) \mid \leq U\#A \sum_{T=0}^{\infty} \delta_i^T \sup_{h \in H_T} \max_{j \neq i} \| \sigma_j(h) - \sigma'_j(h) \|.$$

Proof. It is easily obtained from the intermediate value theorem and the recursive structure. ■

Furthermore, it is obvious that $\mid V_i^{v_i}(\sigma) - V_i(\sigma) \mid \leq \mid v_i \mid$ for all σ and all i. Then, take any $\varepsilon > 0$ and any $\bar{\delta} < 1$. Letting $\epsilon := \varepsilon(1-\bar{\delta}) \diagup 5U\#A$, $\bar{v} := \min[\#A, \frac{\varepsilon}{4}, \hat{v}]$, and $\mid v_i \mid \leq \bar{v}$ for all i, we obtain $\epsilon-$weak merging from Theorem 1 (2). From these it follows that, for

all $T \geq \hat{T}$ and for all i and all σ_i,

$$\begin{aligned}
V_i(\sigma_i, \sigma^*_{h_T,-i}) &\leq V_i^{v_i}(\sigma_i, \sigma^*_{h_T,-i}) + \mid v_i \mid \\
&\leq \bar{V}_i^{v_i}(\sigma^*_{h_T,-i}) + \mid v_i \mid \\
&< \bar{V}_i^{v_i}(\tilde{\rho}^i_{*,h_T}) + \frac{\varepsilon}{4} + \mid v_i \mid \\
&= V_i^{v_i}(\sigma^*_{h_T,i}, \tilde{\rho}^i_{*,h_T}) + \frac{\varepsilon}{4} + \mid v_i \mid \\
&< V_i^{v_i}(\sigma^*_{h_T,i}, \sigma^*_{h_T,-i}) + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} + \mid v_i \mid \\
&\leq V_i(\sigma^*_{h_T,i}, \sigma^*_{h_T,-i}) + \frac{\varepsilon}{2} + 2 \mid v_i \mid \leq V_i(\sigma^*_{h_T,i}, \sigma^*_{h_T,-i}) + \varepsilon.
\end{aligned}$$

The third and fifth inequalities are obtained from Theorem 1 (2) and Lemma 1. The fourth equality holds because $\sigma^*_i = \arg\max_{\sigma_i} V_i^{v_i}(\sigma_i, \tilde{\rho}^i_*)$. The other inequalities are obvious. Therefore, we have obtained Theorem 2 as a corollary of Theorem 1 (2).

Theorem 2 *For any $\varepsilon > 0$ and any $0 \leq \overline{\delta} < 1$, there exist prior beliefs $(\tilde{\rho}^i_*)_i$ such that, for any stage game payoffs $(u_i)_i$ and any discount factors $(\delta_i)_i$ with $\delta_i \leq \overline{\delta}$ for all i, there exists $\bar{v} > 0$ such that, for any (symmetric) payoff perturbations $(v_i)_i$ with $\mid v_i \mid \leq \bar{v}$ for all i, the smooth approximate optimal strategy profile σ^* to prior beliefs $\tilde{\rho}_* := (\tilde{\rho}^i_*)_i$ almost surely converges to $\varepsilon-$Nash equilibrium: with $\mu_{\sigma^*}-$probability one, there exists $\hat{T}$ such that, for all $T \geq \hat{T}$, $\sigma^*_{h_T}$ is an $\varepsilon-$Nash equilibrium.*

Note that each player does *not* need to know opponents payoff structures and discount factors. In other words, all any player has to do is to choose an appropriate prior belief and take a smooth approximate optimal behavior to his prior belief based on his own payoff structure and discount factor. Then, whatever stage game is repeatedly played, they learn to play approximate Nash equilibrium in the corresponding repeated game.

Finally, we remark how to construct *smart* prior beliefs, i.e., prior beliefs that $\epsilon-$learn to predict any learnable set correspondence. Noguchi (2005) shows that any learnable

set (correspondence) is completely characterized by a *countable* set of conditioning rules. Precisely, a set correspondence M_{-i} is learnable if and only if there exists a countable set $\{\mathcal{P}_s^i\}_s$ of conditioning rules such that for all σ_i, any σ_{-i} in $M_{-i}(\sigma_i)$ is *eventually* generated by $\{\mathcal{P}_s^i\}_s$ with playing σ_i (see Section 8 and Noguchi (2005) for details). The point is that we can make use of $\{\mathcal{P}_s^i\}_s$ to obtain Theorem 1. Precisely, for each i, we use a countable set $\{\mathcal{Q}_s^i\}_s$ of conditioning rules which are easily obtained from $\{\mathcal{P}_s^i\}_s$ such that (3.0) $\mathcal{P}_s^i \leq \mathcal{Q}_s^i$ for all s, (3.1) $\mathcal{Q}_s^i \leq \mathcal{Q}_{s+1}^i$ for all s, and (3.2) for all $j \neq i$, all s, and all T, there exists s' such that $\mathcal{F}_T\mathcal{Q}_s^j \leq \mathcal{Q}_{s'}^i$.[6] Property (3.0) allows us to replace $\{\mathcal{P}_s^i\}_s$ by $\{\mathcal{Q}_s^i\}_s$; for all i and all σ_i, any σ_{-i} in $M_{-i}(\sigma_i)$ is *eventually* generated by $\{\mathcal{Q}_s^i\}_s$ with playing σ_i. Properties (3.1) and (3.2) ensure that player i is able to learn any other player j's (smooth approximate) optimal strategy to any belief eventually generated by $\{\mathcal{Q}_s^j\}_s$, as will be shown. Therefore, in the remaining of this paper, without loss of generality, we may assume that $\{\mathcal{P}_s^i\}_{s,i}$ has Properties (3.1) and (3.2). We will construct prior beliefs (in Theorem 1) based on $\{\mathcal{P}_s^i\}_{s,i}$ in the next section and prove Theorem 1 in Sections 7 and 8.

[6] A finite partition $\mathcal{F}_T\mathcal{Q}_s^i$ of H is defined by the following equivalence relation on H:

$$h \sim_{\mathcal{F}_T\mathcal{Q}_s^i} h' \text{ if and only if } h \cdot \tilde{h} \sim_{\mathcal{Q}_s^i} h' \cdot \tilde{h} \text{ for all } \tilde{h} \in \bigcup_{t=0}^{T} H_t.$$

Note that, if opponents take behavior strategies generated by $\mathcal{Q}_s^i$, then just after any finite history in the same class of $\mathcal{F}_T\mathcal{Q}_s^i$, player i faces the *same* strategic situation up to the next T periods. Thus, taking a sufficiently large T, player i, who *discounts* future payoffs, plays almost the same optimal (mixed) actions just after any finite history belonging to the same class of $\mathcal{F}_T\mathcal{Q}_s^i$.

4 Prior Belief Formation

4.1 Phases in prior belief formation process

Each player i infinitely repeats four phases in a given repeated game. The first phase is that player i simply keeps his current *temporary* belief f^i,[7] which we call a *stationary phase*. The second one is called a *test phase* in which player i not only starts to perform new statistical tests against his belief f^i but also checks f^i. The third one is a *formation phase* in which, if player i's belief f^i is rejected in the previous test phase, player i observes realized opponents actions and forms a *new* temporary belief g^i; otherwise, player i continues employing his current belief f^i. The fourth one is a *transition phase* in which if player i's new belief has been formed in the previous formation phase, player i gradually switches from a rejected belief to a new belief. Then, the process proceeds to a new stationary phase. A time interval consisting of three subsequent phases, i.e., test, formation, and transition ones, is called an *active interval*. Furthermore, a time interval consisting of all four phases is called a *cycle*.

4.2 Test procedure

Given a conditioning rule $\mathcal{P}$, a toleration level ξ, and a least sample size $\hat{m}$, we define the statistical test procedure with $(\mathcal{P}, \xi, \hat{m})$ (in player i's prior belief formation process) as follows: suppose that player i has a temporary belief f^i (which is generated by $\mathcal{P}$) and employs $(\mathcal{P}, \xi, \hat{m})$ at the beginning of a given test phase. Then, player i collects realized actions in active periods of each $\alpha \in \mathcal{P}$ during the given test phase and obtains

[7] A temporary belief f^i of player i is formally a profile of opponents behavior strategies, i.e., $f^i := (f^i_j)_{j \neq i}$, where $f^i_j : H \to \Delta(A_j)$.

the empirical distribution $D_j^i(\alpha)$ of each player $j(\neq i)$'s realized actions in active periods of each α. Let $\tilde{m}^\alpha$ designate the number of times that α has been active *during* the test phase: $\tilde{m}^\alpha = \sum_{a_j} D_j^i(\alpha)[a_j]$ for all $j \neq i$. Furthermore, for all $\alpha \in \mathcal{P}$ and all $j \neq i$, define $f_j^i(\alpha) := f_j^i(h)$ for $h \in \alpha$.[8] Then, if $\|D_j^i(\alpha) - f_j^i(\alpha)\| > \xi$ for some $j \neq i$ and some α with $\tilde{m}^\alpha \geq \hat{m}$, we say that f^i is *rejected* (at the end of the given test phase). If a current belief f^i is rejected, player i gives up f^i and forms a new (temporary) belief g^i in the next formation phase. On the other hand, if $\|D_j^i(\alpha) - f_j^i(\alpha)\| \leq \xi$ for all $j(\neq i)$ and all $\alpha \in \mathcal{P}$ with $\tilde{m}^\alpha \geq \hat{m}$, we say that f^i is *not* rejected. In that case, player i continues employing f^i until the next test phase. We remark that player i will keep collecting samples (in *any* phases) for any class whose samples are *not* enough, i.e., any $\alpha \in \tilde{\mathcal{P}}$ with $\tilde{m}^\alpha < \hat{m}$, *until* either at least $\hat{m}$ samples are obtained, or f^i is rejected. In relation to this, f^i is also rejected at the end of the given test phase if, for some class α' that did *not* obtain enough samples in a *past* test phase of checking f^i, enough samples have been collected (up to the given test phase) and $\|D_j^i(\alpha') - f_j^i(\alpha')\| > \xi$ for some $j \neq i$.

4.3 Temporary belief formation

In each formation phase, if player i's temporary belief f^i was rejected in the previous test phase, player i forms a *new* (temporary) belief g^i. Specifically, player i employs a correspondence based on a conditioning rule $\tilde{\mathcal{P}}$ and an accuracy level $\underline{n}$. First of all, let $\Delta^{\underline{n}}(A_j) := \{\pi_j \in \Delta(A_j) \mid \text{for all } a_j,\ \pi_j[a_j] = \frac{n_j}{\underline{n}} \text{ for some nonnegative integer } n_j\}$ and let $\Delta_{-i}^{\underline{n}} := \prod_{j \neq i} \Delta^{\underline{n}}(A_j)$; note that for any $\pi_j \in \Delta(A_j)$, there exists $\pi_j' \in \Delta^{\underline{n}}(A_j)$ such that $\|\pi_j - \pi_j'\| \leq \frac{1}{\underline{n}}$.[9] Then, define a set of opponents strategies generated by

[8] Since f^i is generated by $\mathcal{P}$, for all $\alpha \in \mathcal{P}$ and all $h, h' \in \alpha$, $f^i(h) = f^i(h')$; see Section 2.7. Therefore, $f^i(\alpha)$ is well-defined for all $\alpha \in \mathcal{P}$.

[9] We use the maximum norm on $\Delta(A_j)$: $\|\pi_j\| := \max_{a_j} \mid \pi_j[a_j] \mid$.

$\tilde{\mathcal{P}}$, denoted by $\Sigma_{-i}(\tilde{\mathcal{P}}, \underline{n})$, as follows: $\Sigma_{-i}(\tilde{\mathcal{P}}, \underline{n}) := \{\sigma_{-i} \in G(\tilde{\mathcal{P}}) \text{ |for all } j \neq i \text{ and all } \alpha \in \tilde{\mathcal{P}}, \sigma_j(\alpha) \in \Delta^{\underline{n}}(A_j)\}$, where $\sigma_j(\alpha) := \sigma_j(h)$ for $h \in \alpha$.[10] It is the set of opponents strategies which are generated by $\tilde{\mathcal{P}}$ and whose (mixed) actions all belong to $\Delta^{\underline{n}}_{-i}$; for any σ_{-i} generated by $\tilde{\mathcal{P}}$, there exists σ'_{-i} in $\Sigma_{-i}(\tilde{\mathcal{P}}, \underline{n})$ such that $\|\sigma_j(h) - \sigma'_j(h)\| \leq \frac{1}{\underline{n}}$ for all h and all $j \neq i$. Note also that $\Sigma_{-i}(\tilde{\mathcal{P}}, \underline{n})$ is identified with $\underbrace{\Delta^{\underline{n}}_{-i} \times \cdots \times \Delta^{\underline{n}}_{-i}}_{\#\tilde{\mathcal{P}}}$. Let N^i denote the time length of the formation phase, and the set of all possible histories of *opponents* actions in the formation phase is denoted by H_{-i,N^i}.[11] Then, taking a sufficiently large N^i, consider any function from H_{-i,N^i} to $\Sigma_{-i}(\tilde{\mathcal{P}}, \underline{n})$ which is *surjective*, denoted by $\mathcal{B} : H_{-i,N^i} \to \Sigma_{-i}(\tilde{\mathcal{P}}, \underline{n})$; $\mathcal{B}$ is surjective if for any $\sigma_{-i} \in \Sigma_{-i}(\tilde{\mathcal{P}}, \underline{n})$, there exists $h_{-i} \in H_{-i,N^i}$ such that $\sigma_{-i} = \mathcal{B}(h_{-i})$. We call it player i's *belief correspondence*. Therefore, player i observes a history h_{-i} of opponents actions in the current formation phase and then forms a new belief $g^i = \mathcal{B}(h_{-i})$ at the end of the current formation phase.[12]

4.4 Belief transition

In each transition phase, if his belief was rejected in the previous test phase, each player i *gradually* switches from a rejected belief f^i to a new belief g^i (which has been formed in the previous formation phase). Specifically, let K^i denote the time length of the current transition phase, and let time $T+1$ be the first period of the current transition phase. Given a realized past history h_T, player i has the following *transition* belief during the

[10] Since σ_{-i} is generated by $\tilde{\mathcal{P}}$, $\sigma_j(\alpha)$ is well-defined for all $\alpha \in \tilde{\mathcal{P}}$ and all $j \neq i$.

[11] $H_{-i,N} := \{(a^1_{-i}, \cdots, a^N_{-i}) \mid a^t_{-i} \in \Pi_{j \neq i} A_j \text{ for all } t = 1, \cdots, N\}$.

[12] Note that player i ignores not only the past history before the formation phase but also his own actions in the formation phase to form his temporary belief.

transition phase: for all $1 \leq k \leq K^i$ and all $h'_{k-1} \in H_{k-1}$,

$$(1 - \frac{k}{K^i})f^i(h_T \cdot h'_{k-1}) + \frac{k}{K^i}g^i(h_T \cdot h'_{k-1}),$$

where $h_T \cdot h'_{k-1}$ is the concatenation of h_T and h'_{k-1}.[13]

4.5 Epochs

In order to complete the process, we determine $(\mathcal{P}, \xi, \hat{m})$ for each test phase and $(\tilde{\mathcal{P}}, \underline{n}, \mathcal{B})$ for each formation phase. Furthermore, we specify the *lengths* of all phases. For that purpose, we introduce a concept of time interval: *epochs*. An epoch (of player i) consists of subsequent cycles (of player i). Player i uses the *same* toleration level in all test phases during each epoch (of player i); thus, let ξ^i_s denote the player i's toleration level during the $s-$th epoch. Moreover, the same conditioning rule, accuracy level and belief correspondence are used in all formation phases during each epoch; let $\mathcal{P}^i_s$, $\underline{n}^i_s$ and $\mathcal{B}^i_s$ denote the conditioning rule, the accuracy level and the belief correspondence during the $s-$th epoch. Each epoch switches to the next epoch according to *the number of rejections*, denoted by R^i_s. Precisely, the $s-$th epoch (of player i) switches to the $(s+1)-$th epoch (of player i) if player i's rejections occur R^i_s-times in the $s-$th epoch. We assume that $R^i_s \leq R^i_{s+1}$ for all i and all s. Let us describe conditioning rules and parameters in each epoch.

• Test phases

(i) **Toleration levels:** Player i keeps using the same toleration level ξ^i_s in all test phases during each epoch s. We assume the decrease of $\{\xi^i_s\}_s$: $0 < \xi^i_{s+1} \leq \xi^i_s$ for all s.

(ii) **Conditioning rules:** The switching rule of conditioning rules is a little more com-

[13] $(1 - \frac{k}{K})f^i(h) + \frac{k}{K}g^i(h) := ((1 - \frac{k}{K})f^i_j(h) + \frac{k}{K}g^i_j(h))_{j \neq i}$.

plicated. When the process proceeds to the $s-$th epoch, player i starts to employ $\mathcal{P}_s^i$ in the first test phase (in the $s-$th epoch), and keeps switching to *finer* rules (than $\mathcal{P}_s^i$) until the first rejection occurs (in the $s-$th epoch): employing $\mathcal{P}_{s+1}^i$ in the second test phase (in the $s-$th epoch), employing $\mathcal{P}_{s+2}^i$ in the third test phase, and so on. Just after the first rejection has occurred, player i switches back to $\mathcal{P}_s^i$ in the next test phase and then, he again keeps switching to finer rules until the next rejection occurs; if the next rejection occurs, then he again switches back to $\mathcal{P}_s^i$. Player i repeats this switching behavior through the $s-$th epoch.

(iii) **Sample sizes:** For each $\mathcal{P}_s^i$, we define the *canonical* (least) sample size $\underline{m}_s^i$; we assume the increase of $\{\underline{m}_s^i\}_s$: $\underline{m}_s^i \leq \underline{m}_{s+1}^i$ for all s. Then, when player i employs $\mathcal{P}_{s+q}^i$ in a test phase during the $s-$th epoch, he uses $(\underline{m}_{s+q}^i + d - 1)$ as the (least) sample size in the test phase, where $(d-1)$ is the number of times that player i has employed $\mathcal{P}_{s+q}^i$ in *past* test phases (during the $s-$th epoch).

(iv) **Lengths:** Take a sufficiently large length for a test phase according to $\mathcal{P}_{s+q}^i$ and $(\underline{m}_{s+q}^i + d - 1)$ that are used in the test phase. For example, if player i employs $\mathcal{P}_{s+q}^i$ and $(\underline{m}_{s+q}^i + d - 1)$ in a test phase, let the length of the test phase be $6(2\bar{T}_s^i + 1)(\underline{m}_{s+q}^i + d - 1)(\#\mathcal{P}_{s+q}^i)^2$ periods; see Appendix D for details including $\bar{T}_s^i$.

• Formation phases

(i) **Conditioning rules:** Player i keeps employing $\mathcal{P}_s^i$ in all formation phases during the $s-$th epoch.

(ii) **Accuracy levels:** Player i keeps employing a positive integer $\underline{n}_s^i$ in all formation phases during the $s-$th epoch. We assume the increase of $\{\underline{n}_s^i\}_s$: $\underline{n}_s^i \leq \underline{n}_{s+1}^i$ for all s.

(iii) **Lengths:** We suppose that the lengths of all formation phases in each epoch are the

same. Thus, let N_s^i denote the length of any formation phase during the $s-$th epoch. We assume the increase of $\{N_s^i\}_s$: $N_s^i \leq N_{s+1}^i$ for all s. Furthermore, take a large N_s^i such that $(\#\Delta_{-i}^{\underline{n}_s^i})^{\#\mathcal{P}_s^i} \leq (\#A_{-i})^{N_s^i}$, where $A_{-i} := \Pi_{j\neq i} A_j$.

(iv) **Belief correspondences:** Player i keeps employing a belief correspondence $\mathcal{B}_s^i : H_{-i,N_s^i} \to \Sigma_{-i}(\mathcal{P}_s^i, \underline{n}_s^i)$ in all formation phases during the $s-$th epoch.

• **Transition phases**

(i) **Lengths:** The lengths of transition phases are increasing in time: letting $K^i(n)$ be the length of the $n-$th transition phase of player i (from the beginning of the repeated game), $K^i(n) \leq K^i(n+1)$ for all n, and $\lim_{n\to\infty} K^i(n) = \infty$.

More conditions will be imposed on $\{\xi_s^i\}$, $\{\underline{m}_s^i\}$, $\{\underline{n}_s^i\}$, $\{N_s^i\}$, $\{K^i(n)\}$, and $\{R_s^i\}$ to obtain convergence to $\varepsilon-$Nash equilibrium.

Remark 2 *We do not explicitly argue the lengths of stationary phases. However, we implicitly assume that the lengths of stationary phases grow much more rapidly than the lengths of the other three phases so that the lengths of active intervals become almost negligible compared with those of stationary phases.*

4.6 Constructing prior belief

Finally, we define the prior belief $\tilde{\rho}_*^i$ of each player i. According to player i's prior belief formation process, he keeps employing a temporary belief in the first three phases (i.e., stationary, test, and formation ones) of each cycle, and he may have a transition belief in each transition phase. Given a realized past history h_T, suppose that time $T+1$ is in one of the first three phases, and let f^i be the temporary belief of player i at time $T+1$. Then, define $\tilde{\rho}_*^i(h_T) := f^i(h_T)$. On the other hand, suppose that time $T+1$ is

in the $k-$th period of a transition phase. Let K^i be the length of the transition phase. Furthermore, let f^i denote the temporary belief that was employed until the previous formation phase and g^i denote the temporary belief that will be employed from the next stationary phase.[14] Then, let $\tilde{\rho}^i_*(h_T) := (1 - \frac{k}{K^i})f^i(h_T) + \frac{k}{K^i}g^i(h_T)$.

4.7 Optimal strategies to prior and temporary beliefs

Finally, we evaluate the difference between the smooth approximate optimal strategies to prior and temporary beliefs. We first provide the following lemma. Recall that $\sigma^\rho_i := \arg\max_{\sigma_i} V^{v_i}_i(\sigma_i, \rho_{-i})$ for opponents strategies ρ_{-i}. Let $D^2v_i(\pi_i)$ denote the second derivative of v_i at π_i and $\|(D^2v_i(\pi_i))^{-1}\|$ denote the standard norm of the inverse $(D^2v_i(\pi_i))^{-1}$. Furthermore, define $\|(D^2v_i)^{-1}\| := \sup\{\|(D^2v_i(\pi_i))^{-1}\| \mid \pi_i \in \Delta(A_i; \underline{l}_i)\}$, where $\Delta(A_i; \underline{l}_i) := \{\pi_i \mid \pi_i[a_i] \geq \underline{l}_i \text{ for all } a_i\}$. Then, we obtain the following lemma.

Lemma 2 *For any opponents strategies ρ_{-i}, ρ'_{-i},*

$$\|\sigma^\rho_i(h_0) - \sigma^{\rho'}_i(h_0)\| \leq \frac{U\#A\|(D^2v_i)^{-1}\|}{1-\delta_i} \sum_{T=0}^{\infty} \delta_i^T \max_{h\in H_T} \max_{j\neq i} \|\rho_j(h) - \rho'_j(h)\|.$$

Proof. It is easily obtained by applying the implicit function theorem to the first order condition. ■

Especially, from Lemma 2 it is derived that the difference between the smooth approximate optimal strategies to prior and temporary beliefs is *inversely* proportional to the length of the *next* transition phase. Indeed, let σ^*_i denote player i's *true* strategy, i.e., the smooth approximate optimal strategy against player i's prior belief $\tilde{\rho}^i_*$: $\sigma^*_i := \arg\max_{\sigma_i} V^{v_i}_i(\sigma_i, \tilde{\rho}^i_*)$. Similarly, for player i's temporary belief f^i, let $\sigma^f_i := \arg\max_{\sigma_i} V^{v_i}_i(\sigma_i, f^i)$. Define a subset of finite histories H_{f^i} as follows: $h_T \in H_{f^i}$ if and

[14] Of course, f^i and g^i may be the same.

only if player i employs f^i as his temporary belief (at time $T+1$) after h_T is realized. Then, the difference between σ_i^* and σ_i^f is inversely proportional to the length of the next transition phase as follows.

Lemma 3 *For all $h_T \in H_{f^i}$,*

$$\|\sigma_i^f(h_T) - \sigma_i^*(h_T)\| \leq \frac{\delta_i U \# A \|(D^2 v_i)^{-1}\|}{(1-\delta_i)^2} \frac{1}{K^i},$$

where K^i is the length of the next transition phase (of player i).

Proof. It is immediate from Lemma 2 and the definition of transition belief. ■

5 Conditions for Convergence

We impose conditions on players' prior belief formation processes to obtain convergence to approximate Nash equilibrium. Conditions 1-5 ensure that the probability of reaching an approximate equilibrium after a belief rejection within a certain time interval is bounded away from zero, which is discussed in Section 6. Then, Conditions 6 and 7 ensure convergence to approximate equilibrium, which is argued in Section 7.

The first condition requires that active intervals between players be completely *asyncronized*.

Condition 1. *Any active interval of any player does not overlap any active interval of any other player. In other words, any active interval of any player is included in an intersection of stationary phases of all other players.*

The second one is a bound condition which demands that two main parameters between players be *not* extremely different as time proceeds. Let $\bar{C}_{T+}^i$ denote the maximum among the lengths of past and *present* cycles of player i at time T: time T is in the player

i's present cycle. Let C_T^i designate the length of the most recent past cycle of player i at time T.

Condition 2. *There exists $\bar{c} \geq 1$ such that, for all i, all $j \neq i$ and all T, $\bar{C}_{T+}^i \diagup \bar{C}_T^j \leq \bar{c}$. Furthermore, there exists $\bar{n} \geq 1$ such that, for all i, all $j \neq i$ and all s, $N_s^i \diagup N_s^j \leq \bar{n}$.*

The third condition requires *rapid decrease* of toleration levels $\{\xi_s^i\}_s$ compared with $\sum_j \#\mathcal{P}_s^j$, and the fourth one demands *sufficiently high* accuracy levels $\{\underline{n}_s^i\}_s$ so that players are able to detect wrong beliefs and form accurate beliefs. Furthermore, combining Lemma 3, the fifth condition demands that the lengths of transition phases be *sufficiently large* that the (smooth approximate) optimal strategy σ_i^f against a current temporary belief f^i is eventually (statistically) the same as player i's true strategy σ_i^*.

Condition 3. *For all s and all i, $\xi_s^i \leq \min[\epsilon \diagup 3, 1 \diagup 8(I-1)(\#A+1)s\sum_j \#\mathcal{P}_s^j]$.*

Condition 4. *For all i, j and all s, $s \leq \underline{n}_s^i \xi_s^j$.*

Condition 5. *For all i, j and all s, $s \leq \underline{K}_s^i \xi_s^j$, where $\underline{K}_s^i$ is the minimum among the lengths of transition phases in the $s-$th epoch (of player i).*

The sixth condition demands *sufficiently many* rejections, i.e., a sufficiently large R_s^i, for switching epochs to obtain that approximate equilibrium is played a certain number of times in each epoch. Finally, the seventh condition requires *sufficiently large* (canonical least) sample sizes $\{\underline{m}_s^i\}$ to assure that our statistical tests rapidly become so powerful that those tests reject approximate equilibrium (i.e., almost correct beliefs) *at most finite times* (with probability one).

Condition 6. *For all i and s, $\sum_{m \geq w_s^i R_s^i} \exp(-\frac{1}{2} m (p_s^i)^{2s}) \leq \exp(-s)$, where $p_s^i := (\frac{1}{s})^{sN_s^i}$ and $w_s^i := \frac{1}{s}(\frac{1}{2}(p_s^i)^s)^I$.*

Condition 7. *For all i and all s, $R_s^i (\#\mathcal{P}_s^i) \sum_{m \geq \underline{m}_s^i} \exp(-\frac{1}{8} m (\xi_s^i)^2) \leq \exp(-s)$.*

In the remaining of this paper, we assume that our constructing prior beliefs $(\tilde{\rho}^i_*)_i$ satisfy Conditions 1-7.

Remark 3 *Although Condition 1 is rather restrictive, it is used through this paper only because it makes our argument simple. We remark that Condition 1 can be much weaken: it suffices to impose a regular condition which only demands that active intervals between players be not synchronized most of the time, that is, they be asynchronized in some proportion of the time.*

Remark 4 *Strictly speaking, Condition 1 is not needed for the case of two players.*

Remark 5 *The bound on $\{N^i_s\}$ in Condition 2 can be replaced by the bound on $\{R^i_s\}$: there exists $\bar{r} \geq 1$ such that, for all i, all $j \neq i$, and all s, $R^i_s / R^j_s \leq \bar{r}$.*

6 Equilibrium Reachable Interval

6.1 Approximate equilibrium state

First of all, we define an *approximate equilibrium state*. Let $\hat{\sigma} := (\hat{\sigma}_i)_i$ be an equilibrium of the repeated game with payoff perturbations: $\hat{\sigma}_i := \arg\max_{\sigma_i} V^{v_i}_i(\sigma_i, \hat{\sigma}_{-i})$ for all i.[15] Then, we say that time T is in an *approximate equilibrium state* (abbreviated to AES) if players have temporary beliefs $(f^i)_i$ at time T for which there exists an equilibrium $\hat{\sigma}$ such that, for all i,

$$\begin{aligned} \|f^i_j(h) - \sigma^f_j(h)\| &\leq \frac{\xi^i_{s^i}}{4} \text{ for all } j \neq i \text{ and all } h, \\ \|f^i_j(h) - \hat{\sigma}_j(h)\| &\leq \frac{\xi^i_{s^i}}{4} \text{ for all } j \neq i \text{ and all } h, \end{aligned}$$

[15] Clearly, $\hat{\sigma}$ is $2 \mid v \mid -$(subgame perfect) Nash equilibrium of the original repeated game, where $\mid v \mid := \max_i \mid v_i \mid$.

where s^i denotes the stage of player i's epoch at time T.

6.2 Equilibrium reachable interval

First of all, we introduce a concept of time interval: *equilibrium reachable intervals*. Let s_T^i denote the stage of player i's epoch (at time T), and s_T denote the maximum stage of epoch (at time T), i.e., $s_T := \max_i s_T^i$: we call s_T *the maximum epoch* (at time T). If player i is in the same stage as the maximum epoch, i.e., $s_T^i = s_T$, (at time T), we say that player i is a maximum epoch player (at time T). Furthermore, if $s_T = s$, we say that time T is in maximum epoch s. Suppose that in maximum epoch s, rejection by maximum epoch player has occurred for the *first* time; let player i be the first maximum epoch player who has made the rejection in maximum epoch s. Then, consider the shortest time interval such that (1) it starts from the next period to the rejection, that is, the first period of the next formation phase of player i, say, time T, (2) it includes at least one active interval of each of all other players, and (3) it ends with the last period of a transition phase of player i, and (4) all players' epochs are *always* no more than s through the interval, i.e., *whatever* history happens from time T on, all players' epochs are no more than s through the interval. The time interval is called the first *equilibrium reachable interval* in maximum epoch s; it is abbreviated to the first ER(s)−interval. Inductively, suppose that rejection by maximum epoch player has occurred for the first time *after* the the n−th ER(s)−interval. Then, the shortest interval satisfying (1), (2), (3), and (4) is called the $(n+1)$−st ER(s)−interval. Otherwise, i.e., there is *no* interval satisfying (1), (2), (3), and (4), then the procedure proceeds to ER($s+1$)−intervals.

6.3 Reaching AES

We give a brief explanation about how AES is reached in an ER(s)−interval with some positive probability; the detailed argument is given in Appendices A and B. Consider any ER(s)−interval; let *all* players epoch stages $(s^k)_k$ be *sufficiently large* at the beginning of the ER(s)−interval. Then, under Conditions 1 and 2, there is a positive probability that the learning procedure reaches an AES within the ER(s)−interval. Precisely, the probability of reaching an AES within the ER(s)−interval is at least

$$(\frac{1}{2})^I[(\Pi_k \underline{l}_k)^{\sum_k N_s^k}]^{2\bar{c}}.$$

Indeed, suppose that (maximum epoch) player i's rejection initiates the ER(s)−interval; thus, $s^i = s$. In the formation phase of player i (whose length is N_s^i) just after the player i's rejection, a finite history h^R with length N_s^i can always happen such that h^R, together with player i's belief correspondence $\mathcal{B}_s^i$, generates a temporary belief $f_R^i (= \mathcal{B}_s^i(h_{-i}^R))$ whose (smooth approximate) optimal strategy $\sigma_i^{f_R}$ leads all other players' tests to reject their current beliefs in their *first* test phases in the ER(s)−interval because $\sigma_i^{f_R}$ is statistically different from their beliefs and player i's *true* strategy σ_i^* is (almost) the same as $\sigma_i^{f_R}$ (by Lemma 3 and Condition 5); see Appendix B for how to construct f_R^i. The probability of forming f_R^i, i.e., the probability of h^R, is clearly (at least) $(\Pi_k \underline{l}_k)^{N_s^i}$. Note that player i can keep f_R^i until the *final* test phase (of player i) in the ER(s)−interval: even if f_R^i is rejected, it is possible to form f_R^i again in the next formation phase. Then, in the first test phase of any other player j, player j's powerful test rejects his belief with almost probability one (i.e., at least more than $\frac{1}{2}$) for the above reason. Then, player j forms a new belief in the next formation phase (whose length is $N_{s^j}^j$): that is, a finite history $\hat{h}$ with length $N_{s^j}^j$ can happen (in the next formation phase) such that $\hat{h}$, together with player j's belief correspondence $\mathcal{B}_{s^j}^j$, generates player j's new belief $\hat{g}^j (= \mathcal{B}_{s^j}^j(\hat{h}_{-j}))$

that corresponds to an AES and whose optimal strategy $\sigma_j^{\hat{g}}$ is *statistically different* from f_R^i; see Appendix B for details. The probability of $\hat{g}^j$, i.e., the probability of $\hat{h}$, is at least $(\Pi_k \underline{l}_k)^{N_{sj}^j}$. Thus, player j starts to take a behavior which is almost the same as an (approximate) equilibrium strategy $\sigma_j^{\hat{g}}$. As in the case of f_R^i, player j also can keep $\hat{g}^j$ in the remainder of the ER$(s)-$interval. Since it follows from Condition 2 that there are at most $2\bar{c}$ active intervals of each player in the ER$(s)-$interval,[16] the probability of keeping f_R^i and $(\hat{g}^j)_{j\neq i}$ until the *last* test phase (of player i) in the ER$(s)-$interval is at least $(\frac{1}{2})^{I-1}\Pi_{j\neq i}((\Pi_k \underline{l}_k)^{N_s^j})^{2\bar{c}}((\Pi_k \underline{l}_k)^{N_s^i})^{2\bar{c}-1}$. Finally, in the last test phase (of player i), the equilibrium strategies $(\sigma_j^{\hat{g}})_{j\neq i}$ played by all other players make player i's belief f_R^i rejected with almost probability one (i.e., more than $\frac{1}{2}$) because $(\sigma_j^{\hat{g}})_{j\neq i}$ is statistically different from f_R^i and all other players' true strategies $(\sigma_j^*)_{j\neq i}$ is (almost) the same as $(\sigma_j^{\hat{g}})_{j\neq i}$. Then, player i also can form a new belief $\hat{g}^i$ which corresponds to the AES played by all other players and its probability is at least $(\Pi_k \underline{l}_k)^{N_s^i}$; therefore, the AES is realized at the end of the ER$(s)-$interval. Thus, the learning procedure reaches AES within the ER$(s)-$interval with at least probability $(\frac{1}{2})^I[(\Pi_k \underline{l}_k)^{\sum_k N_s^k}]^{2\bar{c}}$.

7 Convergence to $\varepsilon-$Nash Equilibrium

7.1 Exponential inequality on conditional large deviation

A simple conditional extension of a basic fact of *large deviations* enables us to determine the (least) sample sizes for players' statistical tests in their prior belief formation processes and then obtain all results in this paper. Given a class α, let $\mathbf{S}_m^\alpha$ be the event that state

[16] Precisely, by Condition 2, there are at most $(\bar{c}+1)$ active intervals of player i during each ER$(s)-$interval (initiated by player i) and there are at most $2\bar{c}$ active intervals of any other player during each ER$(s)-$interval.

S occurs between the $m-$th $\alpha-$active period and the next $\alpha-$active period. We show that, if the probability that S occurs between an $\alpha-$active period and the next $\alpha-$active period has *common* upper and lower bounds, then the probability that the frequency of S after the first m $\alpha-$active periods is *not* between the bounds decreases *exponentially* in the sample size m. Let $\mathcal{T}_m^\alpha(h_\infty)$ denote the calendar time of the $m-$th $\alpha-$active period in h_∞; $\mathcal{T}_m^\alpha(h_\infty) < \infty$ means that α is active at least m times in h_∞. Let $\mathbf{d}_m^\alpha[S](h_\infty)$ designate the number of times that S has occurred between two (subsequent) $\alpha-$active periods after the first m $\alpha-$active periods in h_∞.

Lemma 4 *Take any history $h_T \in H$ and any class α such that, for all $h < h_T$, $h \notin \alpha$. Suppose that strategy profile σ and events $\{\mathbf{S}_m^\alpha\}_m$ satisfy the following condition: for all m and all $h_t \in \alpha$ such that $h_T \leq h_t$, $\mu_\sigma(h_t) > 0$, and α has been active exactly $(m-1)-$times in h_t, $l \leq \mu_\sigma(\mathbf{S}_m^\alpha \mid h_t) \leq L$, where l and L are nonnegative numbers. Then, for all $\varepsilon > 0$ and all $m = 1, 2, \cdots$,*

$$\mu_\sigma(\mathcal{T}_m^\alpha < \infty, \frac{\mathbf{d}_m^\alpha[S]}{m} \leq l - \varepsilon \text{ or } \frac{\mathbf{d}_m^\alpha[S]}{m} \geq L + \varepsilon \mid h_T) \leq 2\exp(-2m\varepsilon^2).$$

Proof. This lemma is a straightforward generalization of Lemma 1 in Noguchi (2005). The proof is just the same as that of Lemma 1 in Noguchi (2005). ■

7.2 AES occurs infinitely many times

The initial step to obtain convergence to approximate Nash equilibrium is to show that with probability one, if rejection occurs infinitely many times, AES occurs at least some fixed number of times in each maximum epoch. First of all, combining Conditions 3 to 7 with Lemma 4, we show that if rejection occurs infinitely many times, then *all* players make infinite rejections.

Lemma 5 *With $\mu_{\sigma^*}-$probability one, if rejection occurs infinitely many times, then all players make infinite rejections.*

Proof. See Appendix A. ■

Lemma 5 implies that every player's epoch stage goes to infinity as time proceeds: for all i, $s_T^i \to \infty$ as $T \to \infty$. Note also that, by Condition 2, there are at most $2\bar{c}$ test phases of any player in any ER$(s)-$interval so that, for each s, there are at least $(\underline{R}_s \diagup 2\bar{c})$ ER$(s)-$intervals in maximum epoch s, where $\underline{R}_s := \min_i R_s^i$. As discussed in the previous section, the probability of reaching AES in any ER$(s)-$interval in which all players' epoch stages are *sufficiently large* is at least $(\frac{1}{2})^I[(\Pi_k \underline{l}_k)^{\sum_k N_s^k}]^{2\bar{c}}$. From Condition 6, recall that $p_s^i = (\frac{1}{s})^{sN_s^i}$. Let $\underline{p}_s := \min_i p_s^i$ and $\bar{N}_s := \max_i N_s^i$; thus, $\underline{p}_s = (\frac{1}{s})^{s\bar{N}_s}$. Therefore, there exists $\bar{s}$ such that, for all $s \geq \bar{s}$, $(\frac{1}{2})^I[(\Pi_k \underline{l}_k)^{\sum_k N_s^k}]^{2\bar{c}} \geq \underline{p}_s$. Then, combining this lower bound $\underline{p}_s$ with Lemmas 4 and 5, we obtain the result that AES is reached (at least) in proportion $\frac{1}{2}\underline{p}_s$ of ER$(s)-$intervals.

Lemma 6 *With $\mu_{\sigma^*}-$probability one, if rejection occurs infinitely many times, then there exists s' such that, for each $s \geq s'$, AES is reached at least $\frac{1}{2}\underline{p}_s(\underline{R}_s \diagup 2\bar{c})$ times in the first $(\underline{R}_s \diagup 2\bar{c})$ ER$(s)-$intervals.*

Proof. See Appendix B. ■

From Lemmas 5 and 6 it follows that, for each $s \geq s'$, AES is reached in (at least) one of ER$(s)-$intervals. Therefore, we obtain the following corollary.

Corollary 1 *With $\mu_{\sigma^*}-$probability one, if there are infinitely many rejections, then AES occurs infinitely many times.*

7.3 No rejection from some period on

The second step is to prove that from some period on, *no* rejection occurs. For that purpose, it is convenient to introduce several concepts about test procedure. Suppose that in a current test phase of player i (in epoch s of player i), he employs $(\mathcal{P}^i_{s+q}, (\underline{m}^i_{s+q} + d - 1))$: player i has used $\mathcal{P}^i_{s+q}$ $(d-1)-$times in *past* test phases (during epoch s of player i). Then, as described in Section 4, for each class $\alpha \in \mathcal{P}^i_{s+q}$, player i starts to collect samples (i.e., opponents realized actions) in $\alpha-$active periods from the beginning of the current test phase, and keeps doing so until obtaining enough samples, i.e., (at least) $(\underline{m}^i_{s+q} + d - 1)$ samples, and then checks whether the empirical distribution $D^i(\alpha)$ of the collected (enough) samples is within ξ^i_s of current belief $f^i(\alpha)$ (at the end of the nearest test phase);[17] or if f^i is rejected (by *another* test), then player i stops collecting samples in $\alpha-$active periods and terminates the test. For convenience, the test procedure in $\alpha-$active periods will be called the *$(d-$th$)$ $\alpha-$test (in epoch s of player i)*; thus, by the definition of player i's prior belief formation process, for *all* $\alpha \in \mathcal{P}^i_{s+q}$, the $(d-$th$)$ $\alpha-$test (with the least sample size $(\underline{m}^i_{s+q} + d - 1)$) begins from the first period of the test phase onward. Furthermore, we say that the $(d-$th$)$ $\alpha-$test is *effective* at time T if the $(d-$th$)$ $\alpha-$test is collecting samples at time T.[18] Especially, letting $m^\alpha(h_\infty)$ denote the number of samples that the $\alpha-$test obtains in h_∞, we say that current belief f^i is rejected by the $(d-$th$)$ $\alpha-$test (at the end of a test phase of player i) if, for the $(d-$th$)$ $\alpha-$test, enough samples just have been obtained (up to the last period of the test phase), i.e.,

[17] Since $\mathcal{P}^i_s \leq \mathcal{P}^i_{s+1}$ for all s and f^i is generated by $\mathcal{P}^i_s$, $f^i(\alpha)$ is well-defined for each $\alpha \in \mathcal{P}^i_{s+q}$: $f^i(\alpha) := f^i(h)$ for $h \in \alpha$.

[18] We assume that even if player i has obtained enough samples, i.e., $(\underline{m}^i_s + d - 1)$ samples for the $\alpha-$test, he still keeps collecting samples in $\alpha-$active periods until reaching the last period of the nearest test phase.

$m^\alpha \geq \underline{m}_s^i + d - 1$,[19] but[20]

$$\|D_j^i(\alpha) - f_j^i(\alpha)\| > \xi_s^i \text{ for some } j \neq i.$$

Furthermore, we say that f^i is rejected with *type I error* if f^i is rejected by some $\alpha-$test but f^i is *statistically accurate* in $\alpha-$active periods, i.e., $\|f_j^i(h) - \sigma_j^*(h)\| \leq \xi_s^i/4$ for all $j \neq i$ in all $\alpha-$active periods (since the $\alpha-$test started) in which (enough) samples have been collected. In addition, if f^i is rejected with type I error, we say that the rejection is of type I error.

Condition 7 implies that players' tests rapidly become powerful as time proceeds so that their tests make type I error *at most finite times*. Indeed, we obtains the following result.

Lemma 7 *With $\mu_{\sigma^*}-$probability one, there are at most finite rejections of type I error.*

Proof. See Appendix B. ■

Note that even if rejection occurs in an AES, the probability of forming the same belief again (in the next formation phase) as one that corresponds to the AES is at least $\min_j(\Pi_k \underline{l}_k)^{N_s^j}$; for any sufficiently large s, $\min_j(\Pi_k \underline{l}_k)^{N_s^j} = (\Pi_k \underline{l}_k)^{\bar{N}_s} \geq (\frac{1}{s})^{s\bar{N}_s} = \underline{p}_s$. Thus, even when rejection occurs in an AES, the AES survives with at least probability $\underline{p}_s$. Therefore, we obtain the following lemma.

Lemma 8 *With $\mu_{\sigma^*}-$probability one, if rejection occurs infinitely many times, then there exists $\bar{s}$ such that, for each $s \geq \bar{s}$, the following event happens at least $(\frac{1}{2}\underline{p}_s)^I(\underline{R}_s/2\bar{c})-$times in maximum epoch s: an AES, which has been reached in an ER$(s)-$interval, survives through the first $(I-1)$ rejections after the ER$(s)-$interval.*

[19] Let m^α denote the number of samples that have been obtained until the last period of the test phase.

[20] Since f^i is generated by $\mathcal{P}_s^i$, $f^i(\beta)$ is well-defined for all $\beta \in \mathcal{P}_s^i$: $f^i(\beta) := f^i(h)$ for $h \in \beta$. Note that for all $\alpha \in \mathcal{P}_{s+q}^i$ there exists a unique $\beta \in \mathcal{P}_s^i$ such that $\beta \supset \alpha$; thus, $f^i(\alpha) = f^i(\beta)$.

Proof. See Appendix B. ■

We say that player i's (temporary) belief in an AES is *under correct testing* at time T if any *effective* test of player i at time T has started *after* reaching the AES. If all players' beliefs in an AES are under correct testing, we say that the AES is under correct testing. Note that if a player's belief under correct testing is rejected, then the rejection is of typer I error. Therefore, Lemmas 7 and 8 induce the following lemma.

Lemma 9 *With $\mu_{\sigma^*}-$probability one, if rejection occurs infinitely many times, AES under correct testing occurs infinitely many times.*

Proof. Suppose that there are infinitely many rejections. Then, from Lemmas 7 and 8, there exists $\bar{s}$ such that (1) for all $s \geq \bar{s}$, there is no rejection of type I error during the $s-$th epoch of any player, (2) for each $s \geq \bar{s}$, AES, which is realized in an ER$(s)-$interval, survives through the first $(I-1)$ test rejections after the ER$(s)-$interval at least $(\frac{1}{2}\underline{p}_s)^I(\underline{R}_s/2\bar{c})-$times in maximum epoch s. However, then, from (1) it follows that any rejection (from some period on) is *not* of type I error. Recall the way of reaching an AES in an ER$(s)-$interval: letting player i be the maximum epoch player who initiates the ER$(s)-$interval, in the final test and formation phases (of player i) in the ER$(s)-$interval, player i rejects his wrong belief and forms a belief that corresponds to an AES played by all other players, so that the process reaches the AES: the player i's rejection terminates all player i's tests (that have started before the rejection). Thus, player i's belief is under correct testing (just after the AES has been reached) so that if player i makes the first rejection, then it must be of type I error. Since any rejection is not of type I error, the first rejection is done by some other player, say, j_1. Note that after the first rejection, player j_1's belief is under correct testing because he rejected his previous belief, but formed the same belief as the previous one so that any effective test

of player j_1 has started after the first rejection. Therefore, the second rejection is done by a player other than i and j_1. We repeat this argument so that all players' beliefs are under correct testing after the $(I-1)-$st rejection: the AES is under correct testing. From this and (2) it follows that, for each $s \geq \bar{s}$, AES under correct testing occurs at least $(\frac{1}{2}\underline{p}_s)^I(\underline{R}_s \diagup 2\bar{c})-$times in maximum epoch s. Thus, the desired result easily follows. ■

No rejection from some period immediately follows from Lemmas 7 and 9.

Lemma 10 *With $\mu_{\sigma^*}-$probability one, there are at most finite rejections: no rejection occurs from some period on.*

Proof. By Lemma 9, with $\mu_{\sigma^*}-$probability one, if rejection occurs infinitely many times, AES under correct testing occurs infinitely many times. It means that AES under correct testing is rejected infinitely many times. However, then, it, in turn, implies that there are infinitely many rejections of type I error. This contradicts Lemma 7. Thus, there are at most finite rejections. ■

7.4 $\epsilon-$Learning to predict true strategies

Finally, we show that no rejection from some period implies convergence to approximate Nash equilibrium: the learning procedure has a kind of type II error free property. For that purpose, as shown in Section 3, it suffices to obtain that no rejection from some period implies that each player i's prior belief $\tilde{\rho}^i_*$ $\epsilon-$learns to predict his opponents *true* strategies σ^*_{-i}. The basic argument for obtaining the merging is as follows: note that all players keep some beliefs $(f^i)_i$ *forever* from some period, say, time T_0, because of no rejection from some period; thus, each player i also keeps being in some epoch, say, the s^i-th

epoch, forever from time T_0. Furthermore, from this, Condition 5, and Lemma 3, $(f^i)_i$ and $(\sigma_i^f)_i$ can be identified with $(\rho_*^i)_i$ and $(\sigma_i^*)_i$ respectively. Therefore, if some player's prior belief does *not* $\epsilon-$learn to predict his opponents true strategies, then, for some i_0 and some $j_0 \neq i$, $\|f_{j_0}^{i_0}(h_{t_k}) - \sigma_{j_0}^f(h_{t_k})\| > \epsilon$ for infinitely many t_k. Then, for all q, there exists $\alpha_q \in \mathcal{P}_{s^{i_0}+q}^{i_0}$ such that $h_{t_{k_n}} \in \alpha_q$ for all n; $\{t_{k_n}\}_n$ is an infinite subsequence of $\{t_k\}_k$; this means that the α_q-test obtains enough samples. Furthermore, from Properties (3.1) and (3.2) of $\{\mathcal{P}_s^i\}_{s,i}$ (see Section 3) it follows that $\mathcal{P}_{s^{i_0}+q}^{i_0}$ is a conditioning rule of $f_{j_0}^{i_0}$ for all q, and for any $\delta > 0$, there exists $\bar{q}$ such that, for all $q \geq \bar{q}$, $\mathcal{P}_{s^{i_0}+q}^{i_0}$ is a $\delta-$approximate conditioning rule of $\sigma_{j_0}^f$. From these it follows that, for any sufficiently large q, $\sigma_{j_0}^f(h)$'s are almost the same in all α_q-active periods while $\|f_{j_0}^{i_0}(\alpha_q) - \sigma_{j_0}^f(h)\| > \epsilon$ in all α_q-active periods. Since the α_q-test obtains enough samples, it, together with Lemma 4, implies that the empirical distribution $D_{j_0}^{i_0}(\alpha_q)$ of the collected (enough) samples is also far from $f_{j_0}^{i_0}$, i.e., $\|D_{j_0}^{i_0}(\alpha_q) - f_{j_0}^{i_0}(\alpha_q)\| > \epsilon/2 \geq 3\xi_{s^{i_0}}^{i_0}/2$ with almost probability one. Thus, the α_q-test rejects f^{i_0} with almost probability one. Therefore, for any sufficiently large q, there exists $\alpha_q \in \mathcal{P}_{s^{i_0}+q}^{i_0}$ such that the α_q-test (which starts after time T_0) rejects f^{i_0} with almost probability one: there are infinitely many $\alpha-$tests (which start after time T_0) that reject f^{i_0} with almost probability one. It, together with the Borel-Cantelli argument, implies that f^i is rejected in some test phase (after time T_0). It contradicts no rejection from time T_0. Therefore, we obtain the following proposition, which is exactly the second statement of Theorem1.

Proposition 1 *With $\mu_{\sigma^*}-$probability one, for all i, player i's prior belief ρ_*^i $\epsilon-$learns to predict opponents true strategies σ_{-i}^*: there exists $\hat{T}$ such that, for all $T \geq \hat{T}$, all i, and all $j \neq i$, $\|\tilde{\rho}_{*,j}^i(h_T) - \sigma_j^*(h_T)\| \leq \epsilon$.*

Proof. See Appendix B. ■

8 $\epsilon-$Merging with Any Learnable Set

Although each player's prior belief constructed in the previous section $\epsilon-$learns to predict most of strategies in any learnable set, it is still *not* sufficient for $\epsilon-$learning to predict *all* strategies in any learnable set. There are two reasons: one reason is that each player possibly takes a strategy whose mixed actions put all weight on some pure actions so that multiple such strategies may always generate the same history in any formation phase. It implies that some of those strategies may not be chosen as a temporary belief even if it is actually played. The second reason is that the (rapid) decrease of toleration levels $\{\xi^i_s\}_s$ may cause *infinitely repeated* rejection against some strategies in some learnable set. Accordingly, we need to modify each player's prior belief formation process. Specifically, each player i slightly changes the way of testing in each test phase and belief formation in each formation phase: in each test phase, player i chooses between some *constant* toleration level $\bar{\xi}^i$ and the toleration level ξ^i_s in the current epoch s (of player i) to employ for all $\alpha-$tests that start from (the beginning of) the test phase. Furthermore, in each formation phase, player chooses between "singular" beliefs and other ones.

Precisely, suppose that player i uses $(\mathcal{P}^i_{s+q}, (\underline{m}_{s+q} + d - 1))$ for the test phase: player i has used $\mathcal{P}^i_{s+q}$ $(d-1)-$times in past test phases during epoch s. Player i first collects samples during the test phase, and then obtains empirical distributions for those classes (in $\mathcal{P}^i_{s+q}$) which have obtained enough samples during the test phase. Let $\mathcal{C}$ denote the set of classess (in $\mathcal{P}^i_{s+q}$) which have obtained enough samples (i.e., $\tilde{m}^\alpha \geq \underline{m}_{s+q} + d - 1$) *during the test phase* and $(D^i(\alpha))_{\alpha\in\mathcal{C}}$ denote the family of emprical distributions for $\mathcal{C}$. Then, at the end of the test phase, player i performs a *preliminary* test to choose a toleration level, i.e., $\bar{\xi}^i$ or ξ^i_s by checking the differences between $(D^i(\alpha))_{\alpha\in\mathcal{C}}$. Precisely, if there exist $\alpha, \alpha' \in \mathcal{C}$ and $\beta \in \mathcal{P}^i_{s-1}$ such that $\alpha, \alpha' \subset \beta$ and $\|D^i(\alpha) - D^i(\alpha')\| > \bar{\xi}^i$,

then player i recognizes that opponents strategies are complicated, as he predicted, and employs ξ_s^i for all $\alpha-$tests (that have started from the test phase). Otherwise, player i recognizes opponents strategies as simpler than he predicted, and employs $\bar{\xi}^i$. Then, the remaining test procedure is quite the same as in the previous sections: when $\bar{\xi}^i$ (resp. ξ_s^i) is employed at the end of the test phase, it is used for any $\alpha-$test that has started from the test phase: once enough samples have been obtained for some $\alpha-$test, player i uses $\bar{\xi}^i$ (resp. ξ_s^i) to determine whether f^i is statistically different from the empirical distribution $D^i(\alpha)$ of those samples (at the end of the nearest test phase). In other words, if $\|f_j^i(\alpha) - D_j^i(\alpha)\| > \bar{\xi}^i$ (resp. $\|f_j^i(\alpha) - D_j^i(\alpha)\| > \xi_s^i$), player i rejects f^i (at the end of the test phase).[21]

To modify belief formation, let us first define *singular* beliefs in each epoch s. Define $\partial\Sigma_{-i}(\mathcal{P}_s^i, \underline{n}_s^i) := \{\sigma_{-i} \in \Sigma_{-i}(\mathcal{P}_s^i, \underline{n}_s^i) \mid \exists j \neq i, \exists \alpha \in \mathcal{P}_s^i, \exists a_j (\sigma_j(\alpha)[a_j] < \frac{1}{s})\}$.[22] Then, any opponents strategies in $\partial\Sigma_{-i}(\mathcal{P}_s^i, \underline{n}_s^i)$ is called a *singular* belief of player i (in epoch s): $\partial\Sigma_{-i}(\mathcal{P}_s^i, \underline{n}_s^i)$ is the set of all singular beliefs (in epoch s). Since $Z_s^i := \#\partial\Sigma_{-i}(\mathcal{P}_s^i, \underline{n}_s^i) < \infty$, we arbitrarily number strategy profiles in $\partial\Sigma_{-i}(\mathcal{P}_s^i, \underline{n}_s^i)$: $\partial\Sigma_{-i}(\mathcal{P}_s^i, \underline{n}_s^i) = \{\sigma_{-i}^1, \cdots, \sigma_{-i}^{Z_s^i}\}$. Let $\sigma_{-i}^z(h_{T-1})[a_{-i}] := \Pi_{j\neq i}\sigma_j^z(h_{T-1})[a_j]$. Furthermore, let $\underline{a}_{-i}^z(h_{T-1}) := \arg\min_{a_{-i}} \sigma_{-i}^z(h_{T-1})[a_{-i}]$ for $1 \leq z \leq Z_s^i - 1$, and $\underline{a}_{-i}^0(h_{T-1}) := \arg\min_{a_{-i}} \sigma_{-i}^{Z_s^i}(h_{T-1})[a_{-i}]$. Then, we slightly change temporary belief formation as follows: let the length of each formation phase in epoch s (of player i) be one period longer for each s: $(N_s^i + 1)$ periods. Then,

(1) suppose that time T is the next period to the first rejection by player i in epoch s (of player i): it is the first period of a formation phase. Letting h_{T-1} be a realized past history, if $\underline{a}_{-i}^1(h_{T-1})$ is played at time T, player i employs his belief correspondence

[21]Since $\mathcal{P}_s^i \leq \mathcal{P}_{s+q}^i$ and f^i is generated by $\mathcal{P}_s^i$, for each $\alpha \in \mathcal{P}_{s+q}^i$, $f^i(\alpha)$ is well-defined: $f^i(\alpha) := f^i(h)$ for $h \in \alpha$.

[22]Since σ_j is generated by $\mathcal{P}_s^i$, $\sigma_j(\alpha)$ is well-defined for each $\alpha \in \mathcal{P}_s^i$: $\sigma_j(\sigma) := \sigma_j(h)$ for $h \in \alpha$.

$\mathcal{B}_s^i : H_{-i,N_s^i} \to \Sigma_{-i}(\mathcal{P}_s^i, \underline{n}_s^i)$: then player i obtains a realized history h_{-i} (of opponents actions with length N_s^i) in the remaining periods of the formation phase and then forms a new belief $g^i = \mathcal{B}_s^i(h_{-i})$ at the end of the formation phase, as before. Otherwise, i.e., any other opponents action profile is played at time T, player i chooses σ_{-i}^1 as a new belief at the end of the formation phase.

(2) In general, suppose that time T is the next period to the $(nZ_s^i + z)-$th rejection by player i in epoch s of player i, where n is a nonnegative integer and z is a nonnegative integer less than Z_s^i. Then, if $\underline{a}_{-i}^z(h_{T-1})$ is played at time T, then player i employs $\mathcal{B}_s^i$, obtains a realized history h_{-i} (of opponents actions) in the remaining periods of the formation phase and forms a new belief $g^i = \mathcal{B}_s^i(h_{-i})$. Otherwise, player i chooses σ_{-i}^z as a new belief.

The following condition is imposed on $\{\bar{\xi}^i\}_i$ to obtain the property of merging with any learnable set. It demands that constant toleration levels have certain upper bounds.

Condition 8. *For all i and all s, $\xi_s^i \leq \bar{\xi}^i \leq \min[\epsilon/3, 1/8(I-1)(\#A+1)]$.*

We show the second result, i.e., Proposition 2, which implies the first statement of Theorem 1. First of all, in order to characterize a learnable set correspondence, we need to slightly extend the generation of strategies by conditioning rules. The following definition simply says that, for any $\varepsilon > 0$, the regularity of σ_{-i} is (almost surely) $\varepsilon-$approximated by one of conditioning rules $\{\mathcal{P}_s^i\}_s$ *from some period on.*

Definition 7 *We say that opponents strategies σ_{-i} are eventually generated by a (countable) set of conditioning rules $\{\mathcal{P}_s^i\}_s$ with a player's strategy σ_i if, for all $\varepsilon > 0$, there exist an index s_0, a $\mu_{(\sigma_i,\sigma_{-i})}-$probability one set $\mathbf{Z}_0$, and a time function $T_0 : \mathbf{Z}_0 \to \mathbb{N}$ such that, for all $\alpha \in \mathcal{P}_{s_0}^i$ and all $h_T, h'_{T'} \in \alpha$, if there exist $h_\infty, h'_\infty \in \mathbf{Z}_0$ such that $h_T < h_\infty$*

and $T \geq T_0(h_\infty)$ *and* $h'_{T'} < h'_\infty$ *and* $T' \geq T_0(h'_\infty)$*, then* $\|\sigma_j(h_T) - \sigma_j(h'_{T'})\| < \varepsilon$ *for all* $j \neq i$*.*

Let $EG(\{\mathcal{P}^i_s\}_s, \sigma_i)$ denote the set of all opponents strategies eventually generated by $\{\mathcal{P}^i_s\}_s$ with σ_i. Then, Noguchi (2005) characterizes a learnable set correspondence by using the following result:

Proposition *(Noguchi (2005)) For any prior belief* $\tilde{\rho}^i$ *(of player* i*), there exists a countable set* $\{\mathcal{P}^i_s\}_s$ *of conditioning rules such that, for all* σ_i*,* $M_{-i}(\tilde{\rho}^i, \sigma_i) \subset EG(\{\mathcal{P}^i_s\}_s, \sigma_i)$*.*

From this characterization result and the conditioning rule argument in Section 3 it suffices to show that, for all σ_i, prior belief $\tilde{\rho}^i_*$ $\epsilon-$learns to predict all opponents strategies in $EG(\{\mathcal{P}^i_s\}_s; \sigma_i)$. The proof is almost similar to that of convergence in the previous section. Thus, we obtain the following proposition.

Proposition 2 *For all* i*, player* i*'s prior belief* $\tilde{\rho}^i_*$ $\epsilon-$*learns to predict* σ_{-i} *with* σ_i *for all* $\sigma_{-i} \in EG(\{\mathcal{P}^i_s\}_s; \sigma_i)$ *and all* σ_i*.*

Proof. See Appendix C. ■

Finally, we remark that the modification on $(\tilde{\rho}^i_*)_i$ in this section does *not* change the convergence result in the previous section: the modified prior beliefs $(\tilde{\rho}^i_*)_i$ also (almost surely) converges to $\varepsilon-$Nash equilibrium; the proof is quite the same as in the previous sections except the argument about time interval in which approximate equilibrium is reached with some positive probability; see Appendix D for details.

9 Implication to Impossibility Results

We argue the implication of our positive result to impossibility results in Foster and Young (2001) and Nachbar ((1997), (2005)). Foster and Young (2001) show that under

perfect rationality (i.e., exact optimal behavior), given any prior beliefs, it is impossible to learn to play Nash equilibrium in almost all (stage) games near the matching pennies one. On the other hand, our positive result shows that under *bounded* rationality (i.e., smooth approximate optimal behavior), it is fairly possible to learn to play *approximate* Nash equilibrium for any stage game and any discount factors. From these we conclude that Foster and Young's impossibility crucially depends on perfect rationality. In other words, their impossibility is *not* robust to bounded rationality (i.e., approximate optimal behavior).

Our positive result also has the implication to Nachbar's impossibility. Following Nachbar ((1997), (2005)), let us consider any infinitely repeated game of two players with a certain weak condition.[23] First of all, we slightly extend the evil twin property in Nachbar (2005) for our purpose: we say that a (player j's) pure strategy $\mathbf{s}_j$ is an $(\varepsilon, \epsilon)-$evil twin of a (player i's) pure strategy $\mathbf{s}_i$ if $\mathbf{s}_i$ is *not* $\varepsilon-$uniformly optimal against *any* prior belief $\tilde{\rho}^i$ which $\epsilon-$learns to predict $\mathbf{s}_j$ with $\mathbf{s}_i$;[24] note that the original definition in Nachbar (2005) corresponds to the case that $\epsilon = 0$. Accordingly, it is easy to see that Nachbar's impossibility result still holds for any sufficiently small $\varepsilon, \epsilon > 0$: letting $\hat{\Sigma}_i$ denote a set of player i's strategies, for the given infinitely repeated game, there exists $\eta > 0$ such that, for all $0 \leq \epsilon, \varepsilon \leq \eta$, if any prior beliefs $\tilde{\rho}^1$ and $\tilde{\rho}^2$ have the property of $\epsilon-$learnability[25] on any

[23] Nachbar (1997) and (2005) requires that a stage game of each player satisfy the NWD condition or the MM condition.

[24] σ_i is $\varepsilon-$uniformly optimal against $\tilde{\rho}^i$ if for any finite history h, $V_i(\sigma_i, \tilde{\rho}^i \mid h) + \varepsilon \geq V_i(\sigma_i', \tilde{\rho}^i \mid h)$ for all σ_i'.

[25] We say that prior beliefs $(\tilde{\rho}^1, \tilde{\rho}^2)$ satisfy $\epsilon-$learnability on $\hat{\Sigma}_1 \times \hat{\Sigma}_2$ if, for all $i = 1, 2$ and all $j \neq i$, $\tilde{\rho}^i$ $\epsilon-$learns to predict *all* opponent strategies in $\hat{\Sigma}_j$ with *all* player i's own strategies in $\hat{\Sigma}_i$.

product $\hat{\Sigma}_1 \times \hat{\Sigma}_2$ of strategy sets that[26] satisfies Condition CSP,[27] then they cannot satisfy $\varepsilon-$consistency on $\hat{\Sigma}_1 \times \hat{\Sigma}_2$, i.e., $(\sigma_1, \sigma_2) \notin \hat{\Sigma}_1 \times \hat{\Sigma}_2$ for *any* $\varepsilon-$uniformly optimal strategies σ_1 and σ_2 to $\tilde{\rho}^1$ and $\tilde{\rho}^2$; Nachbar original result corresponds to the case that $\epsilon = 0$. Therefore, as opposed to Foster and Young's impossibility, Nachbar's impossibility result is robust to bounded rationality (i.e., $\varepsilon-$optimal behavior) and approximate learning (i.e., $\epsilon-$weak merging). On the other hand, according to the argument on learnable set in Noguchi (2005) and the previous argument in this paper, we can easily see that, for any $\epsilon > 0$ and any $\hat{\Sigma}_1 \times \hat{\Sigma}_2$ that satisfies $0-$learnability[28] (and Condition CSP), there exist $(\tilde{\rho}^1_*, \tilde{\rho}^2_*)$ in our constructing class of prior beliefs such that $(\tilde{\rho}^1_*, \tilde{\rho}^2_*)$ satisfy $\epsilon-$learnability on $\hat{\Sigma}_1 \times \hat{\Sigma}_2$. Therefore, Nachbar's impossibility result applies to $(\tilde{\rho}^1_*, \tilde{\rho}^2_*)$, i.e., $(\sigma^*_1, \sigma^*_2) \notin \hat{\Sigma}_1 \times \hat{\Sigma}_2$ for any smooth $\varepsilon-$optimal strategies (σ^*_1, σ^*_2) against $(\tilde{\rho}^1_*, \tilde{\rho}^2_*)$. Roughly speaking, this means that some player j's true strategy σ^*_j cannot be learnable *uniformly in* $\hat{\Sigma}_i$ in the sense that $\tilde{\rho}^i_*$ cannot $\epsilon-$learn to predict σ^*_j with *all* player i's strategies in $\hat{\Sigma}_i$. The point is that this negative fact does *not* exclude the possibility that $\tilde{\rho}^i_*$ $\epsilon-$learns to predict σ^*_j with σ^*_i, that is, (σ^*_1, σ^*_2) converges to (approximate) Nash equilibrium. Indeed, as we have shown in Theorem 1, (σ^*_1, σ^*_2) (almost surely) converges to $\varepsilon-$Nash equilibrium. Therefore, we have obtained the following *possibility* result.

Theorem 3 *For any* $\varepsilon, \epsilon > 0$ *and any product* $\hat{\Sigma}_1 \times \hat{\Sigma}_2$ *of strategy sets that satisfies*

[26]Strictly speaking, Nachbar (2005) only requires the learnability on *pure* strategies in $\hat{\Sigma}_1 \times \hat{\Sigma}_2$, which is called *weak* learnability.

[27]Roughly speaking, Condition CS requires the diversity of strategies in $\hat{\Sigma}_1$ and those in $\hat{\Sigma}_2$ and the symmetry between strategies in $\hat{\Sigma}_1$ and those in $\hat{\Sigma}_2$. Condition P demands that for each $i = 1, 2$, any (mixed) behavior strategy in $\hat{\Sigma}_i$ can be approximated by some pure behavior strategy in $\hat{\Sigma}_i$ in some sense. See Nachbar (2005) for details.

[28]$\hat{\Sigma}_1 \times \hat{\Sigma}_2$ is said to satisfy $\epsilon-$learnability if there exist prior beliefs $(\tilde{\rho}^1, \tilde{\rho}^2)$ such that $(\tilde{\rho}^1, \tilde{\rho}^2)$ satisfy $\epsilon-$learnability on $\hat{\Sigma}_1 \times \hat{\Sigma}_2$.

0−learnability and Condition CSP, there exist prior beliefs $(\tilde{\rho}_*^1, \tilde{\rho}_*^2)$ *and those smooth* $\varepsilon-$*optimal strategies* (σ_1^*, σ_2^*) *such that although* $(\tilde{\rho}_*^1, \tilde{\rho}_*^2)$ *satisfy* $\epsilon-$*learnability on* $\hat{\Sigma}_1 \times \hat{\Sigma}_2$ *and Nachbar's impossibility holds for* (σ_1^*, σ_2^*) *on* $\hat{\Sigma}_1 \times \hat{\Sigma}_2$*,* (σ_1^*, σ_2^*) *(almost surely) converges to* $\varepsilon-$*Nash equilibrium.*

Theorem 3 insists that *under bounded rationality (i.e.,* $\varepsilon-$*optimal behavior) and approximate learning (i.e.,* $\epsilon-$*weak merging)*, although Nachbar's impossibility still holds, it is fairly possible for players to learn to play approximate Nash equilibrium. In other words, our positive result clarifies that Nachbar's impossibility is different from the impossibility of learning to play approximate Nash equilibrium in a general sense, and that the richness (and symmetry) of players' learnable sets does *not* necessarily prevent Bayesian learning from converging to approximate Nash equilibrium.

10 Appendix A

10.1 Belief leading to opponents rejections

First of all, we have to show that player i can always form a new belief that leads opponents to *reject* their current beliefs. In this subsection, we provide several classes of beliefs and strategies that can lead to such opponents rejections, according to players' payoffs and discount factors. Let v_i^* denote the player i's maximun payoff, i.e., $v_i^* := \max_a u_i(a)$, and $\underline{v}_i$ denote the player i's minimax payoff, that is, $\underline{v}_i := \min_{\pi_{-i}} \max_{a_i} u_i(a_i, \pi_{-i})$; let $a^* \in \arg\max_a u_i(a)$, $\underline{\pi}_{-i} \in \arg\min_{\pi_{-i}} \max_{a_i} u_i(a_i, \pi_{-i})$, and $\underline{a}_i \in \arg\max_{a_i} u(a_i, \underline{\pi}_{-i})$.

• The case in which there is no weakly dominant action

Take $\tilde{a}_{-i}$ such that $\max_{a_i} u_i(a_i, \tilde{a}_{-i}) > u_i(a_i^*, \tilde{a}_{-i})$ and let $\tilde{a}_i \in \arg\max_{a_i} u_i(a_i, \tilde{a}_{-i})$. Furthermore, given opponents actions $a_{-i} := (a_j)_{j \neq i}$, let π_j^a denote the mixed action of

player j such that $\pi_j^a[a_j] = 1$, and let $\pi_{-i}^a := (\pi_j^a)_{j \neq i}$. Then, define $\pi_{-i}[t] := t\pi_{-i}^{a^*} + (1-t)\pi_{-i}^{\tilde{a}}$. Consider the following player i's belief $\tilde{\rho}_t^i$: for all $j \neq i$, player j always plays $\pi_j[t]$. The smooth approximate optimal strategy σ_i^t to $\tilde{\rho}_t^i$ is to always play $BR_i^{v_i}(\pi_{-i}[t]) := \arg\max_{\pi_i} u_i(\pi_i, \pi_{-i}[t]) + v_i(\pi_i)$: $\sigma_i^t(h) := BR_i^{v_i}(\pi_{-i}[t])$ for all h. Therefore, given any small (symmetric) payoff perturbation, for all $h \in H$, $\sigma_i^0(h)[a_i^*] \approx 0$ and $\sigma_i^1(h)[a_i^*] > \frac{1}{\#A_i+1}$. Since $\sigma_i^t(h)$ is Lipschitz continuous in t, for any $0 < c \leq \frac{1}{\#A_i+1}$, there exists $0 \leq t_c \leq 1$ such that $\sigma_i^{t_c}(h)[a_i^*] = c$ for all $h \in H$.

• The case in which weakly dominant action exists

Let A_i^* denote the set of weakly dominant actions, and fix any $a_i^* \in A_i^*$ and $\bar{a}_{-i} \in \arg\min_{a_{-i}}[u_i(a_i^*, a_{-i}) - \max_{a_i \notin A_i^*} u_i(a_i, a_{-i})]$. Furthermore, define $\bar{u}_i := \max_{a_i \notin A_i^*} u_i(a_i, \bar{a}_{-i})$ and $u_i^* := \max_{a_i} u_i(a_i, \bar{a}_{-i}) (= u_i(a_i^*, \bar{a}_{-i}))$; set $\bar{a}_i \in \arg\max_{a_i \notin A_i^*} u_i(a_i, \bar{a}_{-i})$.

(1) $\delta_i = 0$, or $\delta_i > 0$ and $v_i^* = \underline{v}_i$

We consider two subcases according to the values of u_i^* and $\bar{u}_i$.

(1.1) $u_i^* > \bar{u}_i$

It means that weakly dominant actions always give player i more payoffs than any other actions in a stage game. Thus, given any sufficiently small payoff perturbation, player i always play an (almost) *fixed* mixed action (which puts almost all weight on weakly dominant actions) through a repeated game, which enables us to ignore player i through our argument.

(1.2) $u_i^* = \bar{u}_i$

Since $\bar{a}_i \notin A_i^*$, there exists $\tilde{a}_{-i}$ such that $u_i(\bar{a}_i, \tilde{a}_{-i}) < u_i(a_i^*, \tilde{a}_{-i})$; clearly, $\tilde{a}_{-i} \neq \bar{a}_{-i}$ because $u_i^* = \bar{u}_i$. Let $\pi_{-i}[t] := t\pi_{-i}^{\bar{a}} + (1-t)\pi_{-i}^{\tilde{a}}$. The remaining is the same as the case of no weakly dominant action: given any small (symmetric) payoff perturbation, for all

$h \in H$, $\sigma_i^0(h)[\bar{a}_i] \approx 0$ and $\sigma_i^1(h)[\bar{a}_i] > \frac{1}{\#A_i+1}$. Since $\sigma_i^t(h)$ is Lipschitz continuous in t (for all h), for any $0 < c \leq \frac{1}{\#A_i+1}$, there exists $0 \leq t_c \leq 1$ such that $\sigma_i^{t_c}(h)[a_i^*] = c$ for all $h \in H$.

(**2**) $\delta_i > 0$ **and** $v_i^* > \underline{v}_i$

(**2.1**) **Multiple weakly dominant actions**

Fix any two weakly dominant actions, denoted by a_i^* and b_i^*, and take a sufficiently large integer $\bar{T}$ such that $(1-\delta_i)\underline{v}_i + (\delta_i - \delta_i^{\bar{T}})v_i^* > (1-\delta_i)\underline{v}_i + (\delta_i - \delta_i^{\bar{T}})\underline{v}_i$: for example, $\bar{T} = 2$. Then, letting $m(T)$ be the largest integer such that $m(T)\bar{T} + 1 \leq T$, consider the following player i's belief $\tilde{\rho}_x^i$: for all $j \neq i$, player j takes a minimax action $\underline{\pi}_j$ (against player i) at time $m\bar{T}+1$ for all $m = 0, 1, 2, \cdots$, and player j takes a maximum action a_j^* at any other time $T (\neq m\bar{T}+1$ for all $m)$ if player i plays a (pure) action x at time $m(T)\bar{T}+1$, and he takes $\underline{\pi}_j$ at any other time T if player i plays any other action than x at time $m(T)\bar{T}+1$. Given any sufficiently small payoff perturbation, if x is a weakly dominant action, player i's smooth approximate optimal strategy to $\tilde{\rho}_x^i$ is to play a fixed mixed action (which puts almost probability one on x) at time $m\bar{T}+1$ for all m and play a fixed (mixed) action (that puts almost all weight on weakly dominant actions) at all other times. Thus, the smooth approximate optimal strategy σ_i^t to $\tilde{\rho}_t^i := t\tilde{\rho}_{a_i^*}^i + (1-t)\tilde{\rho}_{b_i^*}^i$ has the following properties: $\sigma_i^0(h)[a_i^*] \approx 0$ and $\sigma_i^1(h)[a_i^*] \approx 1$ for all $h \in \bigcup_m H_{m\bar{T}}$. Furthermore, since $\sigma_i^t(h)$ is Lipschitz continuous in t, for any $0 < c < 1$, there exists $0 \leq t_c \leq 1$ such that $\sigma_i^{t_c}(h)[a_i^*] = c$ for all $h \in \bigcup_m H_{m\bar{T}}$.

(**2.2**) **Unique weakly dominant action**

Let a_i^* denote the unique weakly dominant action. We consider three subcases according to the relation between the player i's discount factor and stage game payoffs.

(**2.2.1**) $(1-\delta_i)\bar{u}_i + \delta_i v_i^* > (1-\delta_i)u_i^* + \delta_i \underline{v}_i$

Take a sufficiently large integer $\bar{T}$ such that $(1-\delta_i)\bar{u}_i + (\delta_i - \delta_i^{\bar{T}})v_i^* > (1-\delta_i)u_i^* + (\delta_i - \delta_i^{\bar{T}})\underline{v}_i$. Then, consider the following player i's belief $\tilde{\rho}_x^i$: for all $j \neq i$, player j plays $\bar{a}_{-i}$ at time $m\bar{T}+1$ for all $m = 0, 1, 2, \cdots$, and player j plays a_j^* at any other time $T(\neq m\bar{T}+1$ for all $m)$ if player i takes x at time $m(T)\bar{T}+1$, and player j plays $\underline{\pi}_j$ at any other time $T(\neq m\bar{T}+1$ for all $m)$ if player i takes any other action than x at time $m(T)\bar{T}+1$. Given any sufficiently small payoff perturbation, the smooth approximate optimal strategy σ_i^t to $\tilde{\rho}_t^i := t\tilde{\rho}_{a_i^*}^i + (1-t)\tilde{\rho}_{\bar{a}_i}^i$ has the following properties: $\sigma_i^0(h)[a_i^*] \approx 0$ and $\sigma_i^1(h)[a_i^*] \approx 1$ for all $h \in \bigcup_m H_{m\bar{T}}$. Since $\sigma_i^t(h)$ is Lipschitz continuous in t, for any $0 < c < 1$, there exists $0 \leq t_c \leq 1$ such that $\sigma_i^{t_c}(h)[a_i^*] = c$ for all $h \in \bigcup_m H_{m\bar{T}}$.

(**2.2.2**) $(1-\delta_i)\bar{u}_i + \delta_i v_i^* = (1-\delta_i)u_i^* + \delta_i \underline{v}_i$

In this case, the order of taking the limits is reverse. We first take a sufficiently small (symmetric) payoff perturbation, and then, take a sufficiently large integer $\bar{T}$. Consider the following player i's belief $\tilde{\rho}_x^i$: for all $j \neq i$, player j takes $\bar{a}_j$ at time $m\bar{T}+1$ for all $m = 0, 1, 2, \cdots$, and he takes a_j^* at any other time $T(\neq m\bar{T}+1$ for all $m)$ if player i plays x at time $m(T)\bar{T}+1$, and takes $\underline{\pi}_j$ at any other time $T(\neq m\bar{T}+1$ for all $m)$ if player i takes any other action than x at time $m(T)\bar{T}+1$. Then, letting $\tilde{\rho}_t^i := t\tilde{\rho}_{a_i^*}^i + (1-t)\tilde{\rho}_{\bar{a}_i}^i$, the smooth approximate optimal strategy σ_i^t to $\tilde{\rho}_t^i$ has the following properties: $\sigma_i^0(h)[a_i^*] \approx \frac{1}{2}$ and $\sigma_i^1(h)[a_i^*] \approx 1$ for all $h \in \bigcup_m H_{m\bar{T}}$. Since $\sigma_i^t(h)$ is Lipschitz continuous in t, for any $\frac{1}{2} < c < 1$, there exists $0 \leq t_c \leq 1$ such that $\sigma_i^{t_c}(h)[a_i^*] = c$ for all $h \in \bigcup_m H_{m\bar{T}}$.

(**2.2.3**) $(1-\delta_i)\bar{u}_i + \delta_i v_i^* < (1-\delta_i)u_i^* + \delta_i \underline{v}_i$

In this case, given any sufficiently small payoff perturbation, player i always plays an (almost) fixed mixed action (that puts almost all weight on a_i^*) through a repeated game, which allows us to ignore player i through our argument.

Finally, let us consider the conditioning rules of the above beliefs. For all i, all $a_i, a_i' \in A_i$, and all $\bar{T} = 1, 2, \cdots$, define a partition $\mathcal{P}(\bar{T}, a_i, a_i') := \{\alpha_{\bar{T}}, \alpha_{a_i}, \alpha_{a_i'}, \alpha_-\}$, where $\alpha_{\bar{T}} := \bigcup_m H_{m\bar{T}}$, $\alpha_x := \{h_T \mid T \neq m\bar{T}$ for all m, x is realized at time $m(T)\bar{T}+1$ in $h_T\}$ for $x = a_i, a_i'$, and $\alpha_- := \{h_\infty \mid T \neq m\bar{T}$ for all m, any other action than a_i and a_i' is realized at time $m(T)\bar{T}+1$ in $h_T\}$. As shown above, $\mathcal{P}(\bar{T}, a_i, a_i')$ is the conditioning rule of a belief $\tilde{\rho}_T^i$ (that leads to opponents rejections): $\mathcal{P}_{\tilde{\rho}_T^i} = \mathcal{P}(\bar{T}, a_i, a_i')$. Then, in additon to (3.1) and (3.2), we assume through the remainder of this paper that (3.3) for all i, i', all $a_i, a_i' \in A_i$, and all $\bar{T} = 1, 2, \cdots$, there exists s' such that $\mathcal{P}(\bar{T}, a_i, a_i') \leq \mathcal{P}_{s'}^{i'}$. It means that each player's belief that leads to his opponents rejections are eventually learnable for all players including his opponents.

10.2 All players make infinite rejections

10.2.1 Rejecting opponents beliefs in their initial epochs

In the remainder of Appendix A, we argue how opponents beliefs are rejected in their first test phases in a given ER(s)−interval (initiated by maximum epoch player i); it is important to assume that (maximum epoch) player i's *epoch stage* s is *sufficiently large* and any other player (i.e., opponent)'s *index* $s^j + q^j$ of her conditioning rule $\mathcal{P}_{s^j+q^j}^j$ in the (first) test phase is also *sufficiently large* at the beginning of the ER(s)−interval. Note that player $j(\neq i)$ may be in a very *early* epoch stage: s^j may be quite *small*. Suppose that (maximum epoch) player i's belief was rejected in the previous test phase. Then, player i forms a temporary belief such that it makes his opponents reject their beliefs. Since s and $s^j + q^j$ are sufficiently large, by Properties (3.1), (3.2), and (3.3), without loss of generality we may assume that $\mathcal{P}_{\tilde{\rho}_t^i} \leq \mathcal{P}_s^i$ and $\mathcal{P}_{\tilde{\rho}_t^i} \leq \mathcal{P}_{s^j+q^j}^j$ for all $j \neq i$. Note also that each player $j(\neq i)$ temporary belief g_0^j (at the beginning of the ER(s)−interval) is

generated by $\mathcal{P}^j_{s^j}$.

In the remainder of this paper we only consider the case of multiple weakly dominant actions in which $\delta_i > 0$ and $v^*_i > \underline{v}_i$; all other cases are quite similar and we omit them.

• **The case of multiple weakly dominant actions:** $\delta_i > 0$ **and** $v^*_i > \underline{v}_i$

In the previous argument we have shown that in this case, σ^t_i has the following properties: there exist $a^*_i \in A_i$ and $\bar{T}$ such that, for any $0 < c < 1$, there exists $0 \leq t_c \leq 1$ such that $\sigma^{t_c}_i(h)[a^*_i] = c$ for all $h \in \bigcup_m H_{m\bar{T}}$. Furthermore, notice that the conditioning rule $\mathcal{P}_{\tilde{\rho}^i_t}$ of $\tilde{\rho}^i_t$ is the following partition: $\mathcal{P}_{\tilde{\rho}^i_t} = \mathcal{P}(\bar{T}, a_i, b_i) = \{\alpha_{\bar{T}}, \alpha_{a_i}, \alpha_{b_i}, \alpha_-\}$; especially, recall that $\alpha_{\bar{T}} = \bigcup_m H_{m\bar{T}}$ and $\alpha_{a^*_i} = \{h_T \mid T \neq m\bar{T}$ for all m, and a^*_i is realized at time $m(T)\bar{T} + 1$ in $h_T\}$. Since player j's current belief g^j_0 is generated by $\mathcal{P}^j_{s^j}$ for all $j \neq i$, $\#\{g^j_{0,i}(h) \in \Delta(A_i) \mid h \in H, j \neq i\} \leq \sum_{j\neq i} \#\mathcal{P}^j_{s^j}$. Then, since $\sigma^t_i(h)[a^*_i]$ is Lipschitz continuous in t, from Condition 3 it is not difficult to show that (for any sufficiently large $s, (s^j)_{j\neq i},$) there exist $0 \leq t_0 \leq 1$ and $0 < c_0 < \frac{1}{\#A_i+1}$ such that (for any sufficiently large $s, (s^j + q^j)_{j\neq i},$) for all $h \in \alpha_{\bar{T}}$ and all $j \neq i$, $(A.1)$ $c_0 - \frac{1}{6}\xi^j_s \leq \sigma^{t_0}_i(h)[a^*_i] \leq c_0 + \frac{1}{6}\xi^j_s$ and $(A.2)$ $\mid \sigma^{t_0}_i(h)[a^*_i] - g^j_{0,i}(h)[a^*_i] \mid > 2\xi^j_{s^j}$. Indeed, by Condition 3, $4\xi^j_{s^j}\#\mathcal{P}^j_{s^j} \leq \frac{4\#\mathcal{P}^j_{s^j}}{8(I-1)(\#A+1)s^j\sum_k \#\mathcal{P}^k_{s^j}} \leq \frac{1}{2(I-1)(\#A_i+1)}$ for all $j \neq i$. Thus, $\sum_{j\neq i} 4\xi^j_{s^j}\#\mathcal{P}^j_{s^j} \leq \frac{1}{2(\#A_i+1)}$. Then, letting $J_g(\alpha, 2\xi^j_{s^j}) := \{x \mid g^j_{0,i}(\alpha)[a^*_i] - 2\xi^j_{s^j} \leq x \leq g^j_{0,i}(\alpha)[a^*_i] + 2\xi^j_{s^j}\}$,

$$\mu_L(\bigcup_{j\neq i}\bigcup_{\alpha\in\mathcal{P}^j_{s^j}} J_g(\alpha, 2\xi^j_{s^j})) \leq \sum_{j\neq i} 4\xi^j_{s^j}\#\mathcal{P}^j_{s^j} \leq \frac{1}{2(\#A_i+1)},$$

where μ_L is the Lebesgue measure on the real line. Note that $\bigcup_{j\neq i}\bigcup_{\alpha\in\mathcal{P}^j_{s^j}} J_g(\alpha, 2\xi^j_{s^j})$ consists of (at most) $\sum_{j\neq i}\#\mathcal{P}^j_{s^j}$ intervals. Therefore, $[0, \frac{1}{\#A_i+1}] \setminus \bigcup_{j\neq i}\bigcup_{\alpha\in\mathcal{P}^j_{s^j}} J_g(\alpha, 2\xi^j_{s^j})$ consists of (at most) $(\sum_{j\neq i}\#\mathcal{P}^j_{s^j} + 1)$ intervals. Especially, the length of one of them must be at least $\frac{1}{2(\#A_i+1)(\sum_{j\neq i}\#\mathcal{P}^j_{s^j}+1)}$. However, then, by Condition 3, $4\xi^j_s \leq \frac{4}{8(I-1)(\#A+1)s\sum_k \#\mathcal{P}^k_s} \leq$

$\frac{1}{2(\#A_i+1)(\sum_{j\neq i}\#\mathcal{P}^j_{s^j}+1)}$ for all j. Then, define $I(c,2\xi^j_s) := \{x \mid c-2\xi^j_s \leq x \leq c+2\xi^j_s\}$. From these and Condition 3, it is easily derived that there exists c_0 such that (for any sufficiently large s,) (1) for all $j \neq i$, $I(c_0,2\xi^j_s)\bigcap\bigcup_{j\neq i}\bigcup_{\alpha\in\mathcal{P}^j_{s^j}} J_g(\alpha,2\xi^j_{s^j}) = \emptyset$ and (2) $I(c_0,2\xi^j_s)\subset[0,\frac{1}{\#A_i+1}]$. Then, it is easy to see that there exists $0\leq t_0\leq 1$ such that, for all $h\in\alpha_{\bar{T}}$, $\sigma^{t_0}_i(h)[a^*_i]=c_0$. We have shown $(A.1)$. From this and (1), it follows that, for all $h\in\alpha_{\bar{T}}$ and all $j\neq i$, $\mid \sigma^{t_0}_i(h)[a^*_i]-g^j_{0,i}(h)[a^*_i]\mid \geq 2\xi^j_{s^j}+(2\xi^j_s-\frac{1}{6}\xi^j_s) > 2\xi^j_{s^j}$. We have show $(A.2)$.

On the other hand, since $\tilde{\rho}^i_{t_0}$ is generated by $\mathcal{P}^i_s$, there always exists a finite history h^R_{-i} in any formation phase (in epoch s) such that $\|f^i_{R,j}(h)-\tilde{\rho}^i_{t_0,j}(h)\|\leq\frac{1}{\underline{n}^i_s}$ for all h and all $j\neq i$, where $f^i_R := \mathcal{B}^i_s(h^R_{-i})$. Therefore, from Condition 4 it follows that (for any sufficiently large s,) for all h, all $j\neq i$, and all k, $\|f^i_{R,j}(h)-\tilde{\rho}^i_{t_0,j}(h)\|\leq\frac{1}{12B_i}\xi^k_s$, where $B_i := \max[1,U\#A\|(D^2v_i)^{-1}\|/(1-\delta_i)]$. From this and Lemma 2 we obtain that, for all h and all $j(\neq i)$, $\|\sigma^{f_R}_i(h)-\sigma^{t_0}_i(h)\|\leq\frac{1}{12}\xi^j_s(\leq\frac{1}{12}\xi^j_{s^j})$. Furthermore, as for player i's true strategy σ^*_i, Condition 5 and Lemma 3 imply that (for any sufficiently large s,) $\|\sigma^*_i(h)-\sigma^{f_R}_i(h)\|\leq\frac{1}{12}\xi^j_s(\leq\frac{1}{12}\xi^j_{s^j})$ for all $h\in H_{f^i_R}$ and all $j\neq i$. Therefore, it, together with the above argument, induces that $c_0-\frac{1}{3}\xi^j_s\leq\sigma^*_i(h)[a^*_i]\leq c_0+\frac{1}{3}\xi^j_s$ for all $h\in H_{f^i_R}\bigcap\alpha_{\bar{T}}$ and all $j\neq i$. Furthermore, recall that $\mathcal{P}_{\tilde{\rho}^i_t}\leq\mathcal{P}^j_{s^j}$ for all $j\neq i$, and $\alpha_{\bar{T}}\in\mathcal{P}_{\tilde{\rho}^i_t}$. Since each test phase is sufficiently long (see Section 4.5), (for all $j\neq i$,) in the first test phase of player j, there exists $\alpha'\in\mathcal{P}^j_{s^j+q^j}$ such that (3) $\alpha'\subset\alpha_{\bar{T}}$ and (4) α' has obtained enough samples during the first test phase, i.e., $\tilde{m}^{\alpha'}\geq\underline{m}^j_{s^j+q^j}+d-1$. From these, Condition 7 and Lemma 4, it follows that for all $h\in\alpha_{\bar{T}}$, $\sigma^{t_0}_i(h)[a^*_i]-\frac{1}{6}\xi^j_s-\frac{1}{3}\xi^j_{s^j}-\frac{1}{2}\xi^j_{s^j}\leq D^j_i(\alpha')[a^*_i]\leq\sigma^{t_0}_i(h)[a^*_i]+\frac{1}{6}\xi^j_s+\frac{1}{3}\xi^j_{s^j}+\frac{1}{2}\xi^j_{s^j}$ with almost probability one. From this and $(A.2)$, for any $h\in\alpha'$, $\|D^j_i(\alpha')-g^j_{0,i}(\alpha')\|\geq\mid D^j_i(\alpha')[a^*_i]-g^j_{0,i}(\alpha')[a^*_i]\mid\geq\mid\sigma^{t_0}_i(h)[a^*_i]-g^j_{0,i}(h)[a^*_i]\mid -\xi^j_{s^j} > 2\xi^j_{s^j}-\xi^j_{s^j}=\xi^j_{s^j}$ with almost probability one (i.e., at

least probability $\frac{1}{2}$): for all $j \neq i$, player j's belief g_0^j is rejected in the first test phase of player j with almost probability one. Therefore, the probability that all opponents reject their beliefs in their first test phases is at least $(\prod_k \underline{l}_k)^{(\bar{c}+1)\bar{N}_s}(\frac{1}{2})^{I-1}$. Indeed, the probability of forming f_R^i in the first formation phase is at least $(\prod_k \underline{l}_k)^{N_s^i}$. Furthermore, player i can continue to employ f_R^i until a given ER(s)−interval, whose probability is at least $(\prod_k \underline{l}_k)^{\bar{c}N_s^i}$ because there are at most $\bar{c}$ active intervals of player i during the ER(s)−interval. Finally, if f_R^i keeps employed, then it makes any other player $j(\neq i)$ reject her belief in her first test phase with almost probability one (i.e., at least $\frac{1}{2}$), as argued above. Therefore, the probability that all opponents reject their beliefs in their first test phases is at least $(\prod_k \underline{l}_k)^{N_s^i}(\prod_k \underline{l}_k)^{\bar{c}N_s^i}(\frac{1}{2})^{I-1} = (\prod_k \underline{l}_k)^{(\bar{c}+1)N_s^i}(\frac{1}{2})^{I-1} \geq (\prod_k \underline{l}_k)^{(\bar{c}+1)\bar{N}_s}(\frac{1}{2})^{I-1}$, where $\bar{N}_s := \max_k N_s^k$.

10.2.2 All players make infinite rejections

We first define class γ_s as follows: $h_T \in \gamma_s$ if and only if (1) time $T+1$ is the first period of an ER(s)−interval, (2) maximum epoch s is no less than $\bar{s}_0$: $s \geq \bar{s}_0$, and (3) the index $s^j + q^j$ of each opponent's conditioning rule employed in her first test phase during the ER(s)−interval is no less than $\bar{s}_0$: $s^j + q^j \geq \bar{s}_0$. Let $\mathbf{d}_m^{\gamma_s}$ denote the number of times that all opponents reject their beliefs during an ER(s)−interval that satisfies (2) and (3) after the first m ER(s)−intervals that satisfy (2) and (3). As shown in the previous subsection, taking a sufficiently large $\bar{s}_0$, the probability that all opponents reject their beliefs in their first test phases is at least $\hat{p}_s := (\frac{1}{2})^{I-1}(\prod_k \underline{l}_k)^{(\bar{c}+1)\bar{N}_s}$. Then, define

$$\mathbf{A}_m^s := \{h_\infty \mid \mathcal{T}_m^{\gamma_s} < \infty, \frac{\mathbf{d}_m^{\gamma_s}}{m} < \hat{p}_s - \frac{1}{2}\underline{p}_s\}.$$

Applying Lemma 4, we obtain that $\mu_{\sigma^*}(\mathbf{A}_m^s) \leq \exp(-\frac{1}{2}m(\underline{p}_s)^2)$. Recall that there are at least $\underline{R}_s/2\bar{c}$ ER(s)−intervals in each maximum epoch s. From this and Condition

2 it follows that $\underline{p}_s = (\frac{1}{s})^{s\bar{N}_s} \geq (\frac{1}{s})^{s\bar{n}N_s^k} = (p_s^k)^{\underline{n}} \geq (p_s^k)^s$ for any sufficiently large s. Furthermore, $\hat{p}_s \geq \underline{p}_s$ for any sufficiently large s. Note also that, for any sufficiently large s, there exists k_s such that

$$\frac{\underline{R}_s}{2\bar{c}} = \frac{R_s^{k_s}}{2\bar{c}} \geq \frac{R_s^{k_s}}{s}.$$

From these and Condition 6, it follows that, for any sufficiently large s',

$$\begin{aligned}
\mu_{\sigma^*}(\bigcap_{s'\geq 1}\bigcup_{s\geq s'}\bigcup_{m\geq \frac{\underline{R}_s}{2\bar{c}}} \mathbf{A}_m^s) &\leq \mu_{\sigma^*}(\bigcup_{s\geq s'}\bigcup_{m\geq \frac{\underline{R}_s}{2\bar{c}}} \mathbf{A}_m^s) \\
&\leq \sum_{s\geq s'}\sum_{m\geq \frac{\underline{R}_s}{2\bar{c}}} \exp(-\frac{1}{2}m(\underline{p}_s)^2) \\
&\leq \sum_{s\geq s'}\sum_{m\geq w_s^{k_s}R_s^{k_s}} \exp(-\frac{1}{2}m((p_s^{k_s})^{2s}) \\
&\leq \sum_{s\geq s'} \exp(-s) = (1-\exp(-1))^{-1}\exp(-s').
\end{aligned}$$

Therefore, $\mu_{\sigma^*}(\bigcap_{s'\geq 1}\bigcup_{s\geq s'}\bigcup_{m\geq \frac{\underline{R}_s}{2\bar{c}}} \mathbf{A}_m^s) = 0$. Let $\mathbf{A} := \bigcup_{s'\geq 1}\bigcap_{s\geq s'}\bigcap_{m\geq \frac{\underline{R}_s}{2\bar{c}}} (\mathbf{A}_m^s)^c$, where $(\mathbf{A}_m^s)^c$ is the complement of $\mathbf{A}_m^s$. Then, $\mu_{\sigma^*}(\mathbf{A}) = 1$. From this we obtain Lemma 5.

Proof of Lemma 5: Suppose that there are infinitely many rejections in $h_\infty \in \mathbf{A}$. Then, maximum epoch goes to infinity as time proceeds: $s_T \to \infty$ as $T \to \infty$. Furthermore, by the definition of player's prior belief formation process, even if some player j only makes finite rejections, the index s^j+q^j of her conditioning rule employed in her test phase goes to infinity. From these it follows that there exists $\bar{s}_1(\geq \bar{s}_0)$ such that, for all $s \geq \bar{s}_1$, the index of any player's conditioning rule employed in any test phase is no less than $\bar{s}_0$ in maximum epoch s. Therefore, for all $s \geq \bar{s}_1$, $\mathbf{d}_m^{\gamma_s}$ =the number of times that all opponents reject their beliefs during an ER(s)−interval after the first m ER(s)−intervals. Furthermore, since $h_\infty \in \mathbf{A}$, there exists $\bar{s}_2(\geq \bar{s}_1)$ such that, for all $s \geq \bar{s}_2$, $\mathbf{d}_m^{\gamma_s} \geq (\hat{p}_s - \frac{1}{2}\underline{p}_s)m$ for all

$m \geq \underline{R}_s \diagup 2\bar{c}$. Note also that, for any sufficiently large s, $\hat{p}_s \geq \underline{p}_s$. Therefore, since there are at least $\underline{R}_s \diagup 2\bar{c}$ ER$(s)-$intervals for all s, for any sufficiently large s, all opponents reject their beliefs at least $\frac{1}{2}\underline{p}_s(\underline{R}_s \diagup 2\bar{c})$ times in maximum epoch s; in addition, any (maximum epoch) player who initiates an ER$(s)-$interval has rejected his belief just before the ER$(s)-$interval. Therefore, all players reject their beliefs at least $\frac{1}{2}\underline{p}_s(\underline{R}_s \diagup 2\bar{c})$ times in maximum epoch s. Notice that $\underline{p}_s\underline{R}_s \to \infty$ as $s \to \infty$ by Condition 6. Therefore, it means that all players make infinite rejections. ■

11 Appendix B

11.1 Rejecting belief and forming equilibrium

11.1.1 Rejecting opponents beliefs

In this subsetion, we provide the detailed arguement of how the procedure reaches an AES in a given ER$(s)-$interval (initiated by maximum epoch player i); we assume that (maximum epoch) player i's epoch stage s and all other players' epoch stages $(s^j)_{j\neq i}$ are *sufficiently large* at the beginning of the ER$(s)-$interval. Suppose that (maximum epoch) player i's belief was rejected in the previous test phase. Fix any Nash equilibrium $\hat{\sigma}$ of the repeated game with payoff perturbations: $\hat{\sigma}$ is a $2 \mid v \mid -$(subgame perfect) Nash equilibrium of the original repeated game. Without loss of generaity we may asssume that $\hat{\sigma}$ has its conditioning rule, denoted by $\mathcal{P}_{\hat{\sigma}}$: $\#\mathcal{P}_{\hat{\sigma}} < \infty$.[29] First of all, player i forms a temporary belief such that the temporary belief *not only* makes the opponents reject their beliefs *but also* will be rejected (with almost probability one) when the opponents

[29]Clearly, we may define a conditioning rule for a strategy profile in the same way as we did for a belief (i.e., an opponents strategy profile) in Section 2.7.

play an AES near $\hat{\sigma}$. Since s, s^j are sufficiently large, from Properties (3.1), (3.2), and (3.3), without loss of generality we may assume that $\mathcal{P}_{\hat{\sigma}} \leq \mathcal{P}^i_s, \mathcal{P}^j_{s^j}$ for all $j(\neq i)$, and that $\mathcal{P}_{\tilde{\rho}^i_t} \leq \mathcal{P}^i_s, \mathcal{P}^j_{s^j}$ for all $j \neq i$. Note also that each player $j(\neq i)$ temporary belief g^j_0 is generated by $\mathcal{P}^j_{s^j}$.

• **The case of multiple weakly dominant actions:** $\delta_i > 0$ **and** $v^*_i > \underline{v}_i$

From the previous argument, in this case, $\tilde{\rho}^i_t := t\tilde{\rho}^i_{a^*_i} + (1-t)\tilde{\rho}^i_{b^*_i}$: for all $j \neq i$, player j takes a minimax action $\underline{\pi}_j$ (against player i) at time $m\bar{T}+1$ for all $m = 0, 1, 2, \cdots$, and player j takes $t\pi^{a^*}_j + (1-t)\underline{\pi}_j$ at any other time $T(\neq m\bar{T}+1$ for all $m)$ if player i plays a dominant action a^*_i at time $m(T)\bar{T}+1$, and player j takes $t\underline{\pi}_j + (1-t)\pi^{a^*}_j$ at any other time $T(\neq m\bar{T}+1$ for all $m)$ if player i plays a dominant action b^*_i at time $m(T)\bar{T}+1$, and player j takes $\underline{\pi}_j$ at any other time T if player i plays any other action than a^*_i, b^*_i at time $m(T)\bar{T}+1$. Thus, the conditioning rule $\mathcal{P}_{\tilde{\rho}^i_t}$ of $\tilde{\rho}^i_t$ is the following partition: $\mathcal{P}_{\tilde{\rho}^i_t} = \mathcal{P}(\bar{T}, a^*_i, b^*_i) = \{\alpha_{\bar{T}}, \alpha_{a^*_i}, \alpha_{b^*_i}, \alpha_-\}$; especially, recall that $\alpha_{\bar{T}} = \bigcup_m H_{m\bar{T}}$ and $\alpha_{a^*_i} = \alpha_{a^*_i} = \{h_T \mid T \neq m\bar{T}$ for all m, and a^*_i is realized at time $m(T)\bar{T}+1$ in $h_T\}$. In other words, $\tilde{\rho}^i_t(h) = \underline{\pi}_{-i}$ for all $h \in \alpha_{\bar{T}}$, $\tilde{\rho}^i_t(h) = t\pi^{a^*}_{-i} + (1-t)\underline{\pi}_{-i}$ for all $h \in \alpha_{a^*_i}$, $\tilde{\rho}^i_t(h) = t\underline{\pi}_{-i} + (1-t)\pi^{a^*}_{-i}$ for all $h \in \alpha_{b^*_i}$, and $\tilde{\rho}^i_t(h) = \underline{\pi}_{-i}$ for all $h \in \alpha_-$. Given any sufficiently small payoff perturbation, the smooth approximate optimal strategy σ^t_i to $\tilde{\rho}^i_t$ has the following properties: there exists $\bar{T}$ such that, for any $0 < c < 1$, there exists $0 \leq t_c \leq 1$ such that $\sigma^{t_c}_i(h)[a^*_i] = c$ for all $h \in \alpha_{\bar{T}}$. Also, for any $h \notin \alpha_{\bar{T}}$, $\sigma^t_i(h) = BR^{v_i}(\tilde{\rho}^i_t(h)) = \arg\max_{\pi_i} u_i(\pi_i, \tilde{\rho}^i_t(h)) + v_i(\pi_i)$. Furthermore, since $\sigma^t_i(h)$ is Lipschitz continuous in t, for any $0 < c < 1$, there exists $0 \leq t_c \leq 1$ such that $\sigma^{t_c}_i(h)[a^*_i] = c$ for all $h \in \alpha_{\bar{T}}$.

Since player j's current belief g^j_0 is generated by $\mathcal{P}^j_{s^j}$ for all $j \neq i$, $\#\{g^j_{0,i}(h) \in \Delta(A_i) \mid h \in H, j \neq i\} \leq \sum_{j\neq i} \#\mathcal{P}^j_s$. Also, it is obvious that $\#\{\hat{\sigma}_j(h) \in \Delta(A_j) \mid h \in H\} \leq \#\mathcal{P}_{\hat{\sigma}}$ for all $j \neq i$. Then, since $\sigma^t_i(h)[a^*_i]$ is Lipschitz continuous in t, from Condition 3 it is

not difficult to show that (for any sufficiently large $s, (s_j)_{j\neq i},$) there exist $0 \leq t_0 \leq 1$ and $\frac{1}{2} < c_0 < 1$ such that $(B.1)$ there exists $j_0 \neq i$ such that, for all $h \in \alpha_{a_i^*}$, $\|\tilde{\rho}^i_{t_0,j_0}(h) - \hat{\sigma}_{j_0}(h)\| > 2\xi^i_{s^{j_0}}$, $(B.2)$ $c_0 - \frac{1}{6}\xi^j_s \leq \sigma^{t_0}_i(h)[a_i^*] \leq c_0 + \frac{1}{6}\xi^j_s$ for all $j \neq i$ and all $h \in \alpha_{\bar{T}}$, and $(B.3)$ $\mid \sigma^{t_0}_i(h)[a_i^*] - g^j_{0,i}(h)[a_i^*] \mid > 2\xi^j_{s_j}$ for all $j \neq i$ and all $h \in \alpha_{\bar{T}}$. Indeed, since $v_i^* > \underline{v}_i$, $\pi^{a^*}_{-i} \neq \underline{\pi}_{-i}$. It means that there exists $j_0 (\neq i)$ such that $\underline{\pi}_{j_0}[a^*_{j_0}] < 1 = \pi^{a^*}_{j_0}[a^*_{j_0}]$. Then, define a (closed) interval $J^0_{\hat{\sigma}}(\alpha, 2\xi^i_{s^{j_0}}) := \{x \mid \hat{\sigma}_{j_0}(\alpha)[a^*_{j_0}] + 2\xi^i_{s^{j_0}} \leq x \leq \hat{\sigma}_{j_0}(\alpha)[a^*_{j_0}] + 2\xi^i_{s^{j_0}}\}$. Furthermore, define $I^\sigma_i(\alpha, 2\xi^i_{s^{j_0}}) := \{\sigma^t_i(h)[a_i^*] \mid t + (1-t)\underline{\pi}_{j_0}[a^*_{j_0}] \in J_{\hat{\sigma}}(\alpha, 2\xi^i_{s^{j_0}})\}$ for $h \in \alpha_{\bar{T}}$; notice that for $h \in \alpha_{a_i^*}$, $\tilde{\rho}^i_{t,j_0}(h)[a^*_{j_0}] = t\pi^{a^*}_{j_0}[a^*_{j_0}] + (1-t)\underline{\pi}_{j_0}[a^*_{j_0}] = t + (1-t)\underline{\pi}_{j_0}[a^*_{j_0}]$. Clearly, $I^\sigma_i(\alpha, 2\xi^i_{s^{j_0}})$ is either a (closed) interval, or an empty set. Also, since $\mu_L(J^0_{\hat{\sigma}}(\alpha, 2\xi^i_{s^{j_0}})) \to 0$ as $s^{j_0} \to \infty$, from Lemma 2 it follows that $\mu_L(I^\sigma_i(\alpha, 2\xi^i_{s^{j_0}})) \to 0$ as $s^{j_0} \to \infty$. From this and $\#\mathcal{P}_{\hat{\sigma}} < \infty$, it follows that $\mu_L(\bigcup_{\alpha\in\mathcal{P}_{\hat{\sigma}}} I^\sigma_i(\alpha, 2\xi^i_{s^{j_0}})) \leq \sum_{\alpha\in\mathcal{P}_{\hat{\sigma}}} \mu_L(I^\sigma_i(\alpha, 2\xi^i_{s^{j_0}})) \to 0$ as $s^{j_0} \to \infty$.

Next, by Condition 3, $4\xi^j_{s^j}\#\mathcal{P}^j_{s^j} \leq \frac{4\#\mathcal{P}^j_{s^j}}{8(I-1)(\#A+1)s^j\sum_k \#\mathcal{P}^k_{s^j}} \leq \frac{1}{2(I-1)(\#A_i+1)s^j}$ for all $j \neq i$. Then, letting $J_g(\alpha, 2\xi^j_{s^j}) := \{x \mid g^j_{0,i}(\alpha)[a_i^*] - 2\xi^j_{s^j} \leq x \leq g^j_{0,i}(\alpha)[a_i^*] + 2\xi^j_{s^j}\}$,

$$\begin{aligned}
\mu_L(\bigcup_{j\neq i}\bigcup_{\alpha\in\mathcal{P}^j_{s^j}} J_g(\alpha, 2\xi^j_{s^j})) &\leq \sum_{j\neq i} 4\xi^j_{s^j}\#\mathcal{P}^j_{s^j} \\
&\leq \sum_{j\neq i} \frac{4\#\mathcal{P}^j_{s^j}}{8(I-1)(\#A+1)s^j\sum_k \#\mathcal{P}^k_{s^j}} \\
&\leq \sum_{j\neq i} \frac{1}{2(I-1)(\#A+1)s^j} \to 0 \text{ as } s^j \to \infty \text{ for all } j \neq i.
\end{aligned}$$

Consider the union of $\bigcup_{\alpha\in\mathcal{P}_{\hat{\sigma}}} I^\sigma_i(\alpha, 2\xi^i_{s^{j_0}})$ and $\bigcup_{j\neq i}\bigcup_{\alpha\in\mathcal{P}^j_{s^j}} J_g(\alpha, 2\xi^j_{s^j})$, denoted by $U(2\xi^i_{s^{j_0}}, (2\xi^j_{s^j})_{j\neq i})$: $U(2\xi^i_{s^{j_0}}, (2\xi^j_{s^j})_{j\neq i}) := (\bigcup_{\alpha\in\mathcal{P}_{\hat{\sigma}}} I^\sigma_i(\alpha, 2\xi^i_{s^{j_0}})) \bigcup (\bigcup_{j\neq i}\bigcup_{\alpha\in\mathcal{P}^j_{s^j}} J_g(\alpha, 2\xi^j_{s^j}))$. Clearly, $U(2\xi^i_{s^{j_0}}, (2\xi^j_{s^j})_{j\neq i})$ consists of at most $(\sum_{j\neq i}\#\mathcal{P}^j_{s^j} + \#\mathcal{P}_{\hat{\sigma}})$ intervals. Therefore, $[\frac{1}{2}, 1] \setminus U(2\xi^i_{s^{j_0}}, (2\xi^j_{s^j})_{j\neq i})$ consists of at most $(\sum_{j\neq i}\#\mathcal{P}^j_{s^j} + \#\mathcal{P}_{\hat{\sigma}} + 1)$ intervals; $\mu_L([\frac{1}{2}, 1] \setminus U(2\xi^i_{s^{j_0}}, (2\xi^j_{s^j})_{j\neq i})) \to \frac{1}{2}$, as $s^j \to \infty$ for all $j \neq i$. Then, the length of one of them must be at least $\frac{1}{3(\sum_{j\neq i}\#\mathcal{P}^j_{s^j} + \#\mathcal{P}_{\hat{\sigma}} + 1)}$ (for sufficiently large s^j's). However, then, from Condition 3 it follows that (for any suf-

ficiently large s,) $4\xi^j_s \leq \frac{4}{8(I-1)(\#A+1)s\sum_k \#\mathcal{P}^k_s} < \frac{1}{3(\sum_{j\neq i}\#\mathcal{P}^j_{s^j}+\#\mathcal{P}_{\hat{\sigma}}+1)}$ for all $j \neq i$. Then, define $I(c, 2\xi^j_s) := \{x \mid c - 2\xi^j_s \leq x \leq c + 2\xi^j_s\}$. From these, it is easily derived that there exists $\frac{1}{2} < c_0 < 1$ such that (1) for all $j \neq i$, $I(c_0, 2\xi^j_s) \bigcap U(2\xi^i_{s^{j_0}}, (2\xi^j_{s^j})_{j\neq i}) = \emptyset$ and (2) for all $j \neq i$, $I(c_0, 2\xi^j_s) \subset [0, 1]$. Then, it is easy to see that there exists $0 \leq t_0 \leq 1$ such that (3) for all $h \in \alpha_{\bar{T}}$, $\sigma^{t_0}_i(h)[a^*_i] = c_0$ and (4) for all $h \in \alpha_{\bar{T}}$, $\sigma^{t_0}_i(h)[a^*_i] \notin \bigcup_{\alpha\in\mathcal{P}_{\hat{\sigma}}} I^{\sigma}_i(\alpha, 2\xi^i_{s^{j_0}})$. Then, from (4) and the definition of $\tilde{\rho}^i_{t_0,j_0}$ it follows that, for all $h \in \alpha_{a^*_i}$, $\tilde{\rho}^i_{t_0,j_0}(h)[a^*_{j_0}] = t_0 + (1 - t_0)\underline{\pi}_{j_0}[a^*_{j_0}] \notin \bigcup_{\alpha\in\mathcal{P}_{\hat{\sigma}}} J_{\hat{\sigma}}(\alpha, 2\xi^i_{s^{j_0}})$. It implies $(B.1)$. $(B.2)$ is immediate from (2) and (3). From (1) and (3), it follows that, for all $h \in \alpha_{\bar{T}}$ and all $j \neq i$, $\mid \sigma^{t_0}_i(h)[a^*_i] - g^j_{0,i}(h)[a^*_i] \mid \geq 2\xi^j_{s^j} + 2\xi^j_s > 2\xi^j_{s^j}$. We have show $(B.3)$.

On the other hand, since $\tilde{\rho}^i_{t_0}$ is generated by $\mathcal{P}^i_s$, there always exists a finite history h^R_{-i} in any formation phase (in epoch s) such that $\|f^i_{R,j}(h) - \tilde{\rho}^i_{t_0,j}(h)\| \leq \frac{1}{\underline{n}^i_s}$ for all h and all $j \neq i$, where $f^i_R := \mathcal{B}^i_s(h^R_{-i})$. Therefore, from Condition 4 it follows that (for any sufficiently large s,) for all h, all $j \neq i$, and all k, $\|f^i_{R,j}(h) - \tilde{\rho}^i_{t_0,j}(h)\| \leq \frac{1}{12B_i}\xi^k_s$. From this and Lemma 2 we obtain that, for all h and all $j(\neq i)$, $\|\sigma^{f_R}_i(h) - \sigma^{t_0}_i(h)\| \leq \frac{1}{12}\xi^j_s (\leq \frac{1}{12}\xi^j_{s_j})$. Furthermore, as for player i's true strategy σ^*_i, Condition 5 and Lemma 3 imply that (for any sufficiently large s,) $\|\sigma^*_i(h) - \sigma^{f_R}_i(h)\| \leq \frac{1}{12}\xi^j_s (\leq \frac{1}{12}\xi^j_{s_j})$ for all $h \in H_{f^i_R}$ and all $j \neq i$. Therefore, it, together with the above argument, induces that $c_0 - \frac{1}{3}\xi^j_s \leq \sigma^*_i(h)[a^*_i] \leq c_0 + \frac{1}{3}\xi^j_s$ for all $h \in H_{f^i_R} \bigcap \alpha_{\bar{T}}$ and all $j \neq i$. This, together with Condition 7 and Lemma 4, implies that in the first test phase of player j, player j's test rejects g^j_0 with almost probability one (for all $j \neq i$). Indeed, since $\mathcal{P}_{\tilde{\rho}^i_t} \leq \mathcal{P}^j_{s_j}$ for all $j \neq i$ and $\alpha_{\bar{T}} \in \mathcal{P}_{\tilde{\rho}^i_t}$ and the first test phase of player j is sufficiently long (see Section 4.5), there exists a class $\alpha' \in \mathcal{P}^j_{s^j+q^j}$ such that $\alpha' \subset \alpha_{\bar{T}}$ and α' has obtained enough samples during the first test phase, i.e., $\tilde{m}^{\alpha'} \geq \underline{m}^j_{s^j+q^j} + d - 1$. Furthermore, from these, $(B.2)$, Condition 7, and Lemma 4, $\sigma^{t_0}_i(h)[a^*_i] - \frac{1}{6}\xi^j_s - \frac{1}{3}\xi^j_s - \frac{1}{2}\xi^j_{s^j} \leq D^j_i(\alpha')[a^*_i] \leq \sigma^{t_0}_i(h)[a^*_i] + \frac{1}{6}\xi^j_s + \frac{1}{3}\xi^j_s + \frac{1}{2}\xi^j_{s^j}$

for all $h \in \alpha_{\bar{T}}$ with almost probability one. From this and $(B.3)$ it follows that, for all $h \in \alpha'$, $\|D_i^j(\alpha') - g_{0,i}^j(\alpha')\| \geq |\ D_i^j(\alpha')[a_i^*] - g_{0,i}^j(\alpha')[a_i^*]\ | \geq |\ \sigma_i^{t_0}(h)[a_i^*] - g_{0,i}^j(h)[a_i^*]\ | - \xi_{s^j}^j > 2\xi_{s^j}^j - \xi_{s^j}^j = \xi_{s^j}^j$ with almost probability one (i.e., at least probability $\frac{1}{2}$).

11.1.2 Forming equilibrium beliefs

We have shown that maximum epoch player j forms a belief f_R^i (in the first formation phase) with at least probability $(\prod_k \underline{l}_k)^{N_s^i}$ and then it makes all other players reject their beliefs in their first test phases with almost probability one (i.e., at least probability $\frac{1}{2}$) (during the given ER$(s)-$interval). In this subsection we argue that after the rejections, all other players form approximate equilibrium beliefs as new ones in their next formation phases and those beliefs, in turn, make player i reject f_R^i in the final test phase (of player i) during the given ER$(s)-$interval.

• **The case of multiple weakly dominant actions:** $\delta_i > 0$ and $v_i^* > \underline{v}_i$

Since $\mathcal{P}_{\hat{\sigma}} \leq \mathcal{P}_{s^j}^j$, there always exists a finite history $\hat{h}_{-j}$ (in the next formation phase) which, together with player j's belief correspondence $\mathcal{B}_{s^j}^j$, generates a new belief $\hat{g}^j = \mathcal{B}_{s^j}^j(\hat{h}_{-j})$ such that $\|\hat{g}_k^j(h) - \hat{\sigma}_k(h)\| \leq \frac{1}{\underline{n}_{s^j}^j}$ for all h and all $k \neq j$. Thus, it follows from Condition 4 that (for any sufficiently large s_j,) $\|\hat{g}_k^j(h) - \hat{\sigma}_k(h)\| \leq \frac{1}{12B_j}\xi_{s^j}^l$ for all h, all $j \neq i$, all $k \neq j$, and all l, where $B_j := \max[1, U\#A\|(D^2 v_j)^{-1}\|/(1-\delta_j)]$. From this and Lemma 2 we obtain that, for each $j \neq i$, the smooth approximate optimal strategy $\sigma_j^{\hat{g}}$ to $\hat{g}^j$ satisfies that $\|\sigma_j^{\hat{g}}(h) - \hat{\sigma}_j(h)\| \leq \frac{1}{12}\xi_{s^j}^l$ for all h and all l. Thus, $\|\sigma_j^{\hat{g}}(h) - \hat{g}_j^k(h)\| \leq \frac{1}{6}\xi_{s^j}^l$ for all h, all $k \neq i$, all $j \neq k, i$, and all l. Also, Condition 5 and Lemma 3 imply that (for any sufficiently large s_j,) $\|\sigma_j^*(h) - \sigma_j^{\hat{g}}(h)\| \leq \frac{1}{12}\xi_{s^j}^l$ for all $j \neq i$, all l, and all $h \in \bigcap_{j \neq i} H_{\hat{g}^j}$. Therefore, all players other than i have (approximate) equilibrium beliefs $(\hat{g}^j)_{j \neq i}$. Note that even if player j $(\neq i)$ faces another test phase and rejects $\hat{g}^j$ before the final test phase

of player i (in the given ER$(s)-$interval), $\hat{g}^j$ can be formed again in the next formation phase (combined with player j's belief correspondence); by Condition 2, the number of test phases of player j during any given ER$(s)-$interval is at most $2\bar{c}$. Therefore, maximum epoch player i forms a rejecting belief f_R^i, which makes all players other than i reject his current beliefs $(g_0^j)_{j\neq i}$ and then they form new beliefs $(\hat{g}^j)_{j\neq i}$ and keep employing them until the last test phase of player i (in the given ER$(s)-$interval) with at least probability $(\frac{1}{2})^{I-1}(\prod_k \underline{l}_k)^{N_s^i+2\bar{c}\sum_{j\neq i}N_s^j}$.

Finally, consider player i. After player i has had a belief f_R^i which leads to opponent rejections, he keeps employing f_R^i (until the last test phase of player i in the given ER$(s)-$interval) in the sense that even when player i has an interim test phase and rejects f_R^i, f_R^i can be formed again (combined with player i's belief correspondence) in the next formation phase; the number of player i's test phases during any given ER$(s)-$interval is at most $\bar{c}$. Thus, player i can keep employing f_R^i until the last test phase (of player i) in the given ER$(s)-$interval with at least probability $(\prod_k \underline{l}_k)^{\bar{c}N_s^i}$. Recall $(A.4)$ in the previous section: for all $h \in \alpha_{a_i^*}$, $\|\tilde{\rho}^i_{t_0,j_0}(h) - \hat{\sigma}_{j_0}(h)\| > 2\xi^i_{s^{j_0}}$. Recall also that, for all h and all $j \neq i$ $\|f^i_{R,j}(h) - \tilde{\rho}^i_{t_0,j}(h)\| \leq \frac{1}{12B_i}\xi^i_s (\leq \frac{1}{12}\xi^i_{s^j})$. Therefore, f_R^i is different from approximate equilibrium $\hat{\sigma}$: $\|f^i_{R,j_0}(h) - \hat{\sigma}_{j_0}(h)\| > \frac{23}{12}\xi^i_{s^{j_0}} (\geq \frac{23}{12}\xi^i_s)$ for all $h \in \alpha_{a_i^*}$. However, then, from Condition 5 and Lemma 3 it follows that (for any sufficiently large s_j,) $\|\sigma_j^*(h) - \sigma_j^{\hat{g}}(h)\| \leq \frac{1}{12}\xi^i_{s^j}$ for all $j \neq i$ and all $h \in \bigcap_{j\neq i} H_{\hat{g}^j}$. From this and the above argument, it is derived that $\|\sigma_j^*(h) - \hat{\sigma}_j(h)\| \leq \frac{1}{6}\xi^i_{s_j}$ for all $j \neq i$ and all $h \in \bigcap_{j\neq i} H_{\hat{g}^j}$. Thus, $\|f^i_{R,j_0}(h) - \sigma^*_{j_0}(h)\| > \frac{21}{12}\xi^i_{s^{j_0}} (\geq \frac{21}{12}\xi^i_s)$ for all $h \in (\bigcap_{j\neq i} H_{\hat{g}^j}) \bigcap \alpha_{a_i^*}$. Note that $\mathcal{P}_{\tilde{\rho}^i_t} \leq \mathcal{P}^i_{s+q}$ and that $\sigma_i^*(h)[a_i^*] \geq \frac{1}{2} - \frac{1}{3}\xi^j_s$ for all $j \neq i$ and all $h \in \alpha_{\bar{T}} \bigcap H_{f_R^i}$. Then, since the last test phase of player i is sufficiently long, in the last test phase of player i, there exists $\alpha'' \in \mathcal{P}^i_{s+q}$ such that $\alpha'' \subset \alpha_{a_i^*}$ and α'' has obtained enough sam-

ples during the last test phase, i.e., $\tilde{m}^{\alpha''} \geq \underline{m}^i_{s+q} + d - 1$ *with almost probability one.* Furthermore, since $\mathcal{P}_{\hat{\sigma}} \leq \mathcal{P}^i_{s+q}$, $\sigma_{j_0}(h)$ is constant in $\alpha''-$active periods; note also that $\|\sigma^*_{j_0}(h) - \hat{\sigma}_{j_0}(h)\| \leq \frac{1}{6}\xi^i_{s^{j_0}}$ for all $h \in H_{\hat{g}^{j_0}}$. From these, Condition 7, and Lemma 4, it follows that $\hat{\sigma}_{j_0}(\alpha'')[a^*_{j_0}] - \frac{1}{6}\xi^i_{s^{j_0}} - \frac{1}{4}\xi^i_s \leq D^i_{j_0}(\alpha'')[a^*_{j_0}] \leq \hat{\sigma}_{j_0}(\alpha'')[a^*_{j_0}] + \frac{1}{6}\xi^i_{s^{j_0}} + \frac{1}{4}\xi^i_s$ with almost probability one. It, together with $(A.3)$, implies that for $h \in \alpha''$, $\|D^i_{j_0}(\alpha'') - f^i_{R,j_0}(\alpha'')\| \geq \|f^i_{R,j_0}(h) - \hat{\sigma}_{j_0}(h)\| - \frac{1}{6}\xi^i_{s^{j_0}} - \frac{1}{4}\xi^i_{s^{j_0}} \geq \frac{23}{12}\xi^i_{s^{j_0}} - \frac{5}{12}\xi^i_{s^{j_0}} = \frac{3}{2}\xi^i_{s^{j_0}} > \xi^i_s$ with almost probability one: with almost probability one (i.e., at least probability $\frac{1}{2}$), player i rejects f^i_R. Then, player i forms a new belief in the last formation phase: since $\mathcal{P}_{\hat{\sigma}} \leq \mathcal{P}^i_s$, there always exists a finite history $\hat{h}_{-i}$ (in the formation phase) which, together with $\mathcal{B}^i_s$, generates a temporary belief $\hat{g}^i = \mathcal{B}^i_s(\hat{h}_{-i})$ such that, for all h, all $j \neq i$ and all l, $\|\hat{g}^i_j(h) - \hat{\sigma}_j(h)\| \leq \frac{1}{12B_i}\xi^l_s$. Furthermore, it, together with Lemma 2, implies that, for all h and all l, $\|\sigma^{\hat{g}}_i(h) - \hat{\sigma}_i(h)\| \leq \frac{1}{12}\xi^l_s$. Finally, the probability of $\hat{g}^i$, i.e., the probability of $\hat{h}_{-i}$, is at least $(\prod_k \underline{l}_k)^{N^i_s}$. From this and the above argument about $(\hat{g}^j)_{j\neq i}$, it follows that, for all i, j with $i \neq j$ and all h, $\|\hat{g}^i_j(h) - \sigma^{\hat{g}}_j(h)\| \leq \frac{1}{4}\xi^i_s$ and $\|\hat{g}^i_j(h) - \hat{\sigma}_j(h)\| \leq \frac{1}{6}\xi^i_s$. Therefore, all players have (approximate equilibrium) beliefs $(\hat{g}^k)_k$ at the end of the given ER$(s)-$interval. In other words, an AES $(\sigma^{\hat{g}}_k)_k$ is reached in the given ER$(s)-$interval with at least probability $(\frac{1}{2})^I(\prod_k \underline{l}_k)^{(\bar{c}+1)N^i_s + 2\bar{c}\sum_{j\neq i} N^j_s} \geq (\frac{1}{2})^I(\prod_k \underline{l}_k)^{2\bar{c}\sum_k N^k_s} \geq (\frac{1}{2})^I(\prod_k \underline{l}_k)^{2\bar{c}I\bar{N}_s}$.

11.2 AES is reached infinitely many times

For all s, define class α_s such that $h_T \in \alpha_s$ if and only if (1) time $T+1$ is the first period of an ER$(s)-$interval and (2) all players' epoch stages are no less than $\tilde{s}_0$ at the beginning of the ER$(s)-$interval: $s, s^j \geq \tilde{s}_0$ for all $j \neq i$. From the previous argument in Appendix A, it follows that, taking a sufficiently large $\tilde{s}_0$, the probability that AES is reached in an ER$(s)-$interval is at least $(\frac{1}{2})^I(\prod_k \underline{l}_k)^{2\bar{c}\sum_k N^k_s} \geq (\frac{1}{2})^I(\Pi_k \underline{l}_k)^{2\bar{c}I\bar{N}_s})$. Then,

define $\tilde{p}_s := (\frac{1}{2})^I(\Pi_k \underline{l}_k)^{2\bar{c}I\bar{N}_s}$. Let $\mathbf{d}_m^{\alpha_s}(h_\infty)$ denote the number of times that AES has been reached in the first m ER$(s)-$intervals in which all players' epoch stages are no less than $\tilde{s}_0$. Let$\mathcal{T}_m^{\alpha_s}(h_\infty)$ denote the calendar time when α_s is active the $m-$th time in h_∞. Define

$$\mathbf{B}_m^s := \{h_\infty \mid \mathcal{T}_m^{\alpha_s} < \infty, \frac{\mathbf{d}_m^{\alpha_s}}{m} < \tilde{p}_s - \frac{1}{2}\underline{p}_s\}.$$

Applying Lemma 4, we obtain that $\mu_{\sigma^*}(\mathbf{B}_m^s) \leq \exp(-2m(\frac{1}{2}\underline{p}_s)^2) = \exp(-\frac{1}{2}m(\underline{p}_s)^2)$. Recall that there are at least $(\underline{R}_s/2\bar{c})$ ER$(s)-$intervals in each maximum epoch s. From this and Condition 2 it follows that for all k, $\underline{p}_s = (\frac{1}{s})^{s\bar{N}_s} \geq (\frac{1}{s})^{s\bar{n}N_s^k} = (p_s^k)^{\underline{n}} \geq (p_s^k)^s$ for any sufficiently large s. Furthermore, $\tilde{p}_s \geq \underline{p}_s$ for any sufficiently large s. Note also that, for any sufficiently large s, there exists k_s such that

$$\frac{\underline{R}_s}{2\bar{c}} = \frac{R_s^{k_s}}{2\bar{c}} \geq \frac{R_s^{k_s}}{s}.$$

From these and Condition 6, it follows that, for any sufficiently large s',

$$\begin{aligned}
\mu_{\sigma^*}(\bigcap_{s'\geq 1}\bigcup_{s\geq s'}\bigcup_{m\geq \frac{R_s}{2\bar{c}}} \mathbf{B}_m^s) &\leq \mu_{\sigma^*}(\bigcup_{s\geq s'}\bigcup_{m\geq \frac{R_s}{2\bar{c}}} \mathbf{B}_m^s) \\
&\leq \sum_{s\geq s'}\sum_{m\geq \frac{R_s}{2\bar{c}}} \exp(-\frac{1}{2}m(\underline{p}_s)^2) \\
&\leq \sum_{s\geq s'}\sum_{m\geq w_s^{k_s}R_s^{k_s}} \exp(-\frac{1}{2}m((p_s^{k_s})^{2s}) \\
&\leq \sum_{s\geq s'} \exp(-s) = (1-\exp(-1))^{-1}\exp(-s').
\end{aligned}$$

Therefore, $\mu_{\sigma^*}(\bigcap_{s'\geq 1}\bigcup_{s\geq s'}\bigcup_{m\geq \frac{R_s}{2\bar{c}}} \mathbf{B}_m^s) = 0$. Let $\mathbf{B} := \bigcup_{s'\geq 1}\bigcap_{s\geq s'}\bigcap_{m\geq \frac{R_s}{2\bar{c}}}(\mathbf{B}_m^s)^c$, where $(\mathbf{B}_m^s)^c$ is the complement of $\mathbf{B}_m^s$. Then, $\mu_{\sigma^*}(\mathbf{B}) = 1$. From this we obtain Lemma 6.

Proof of Lemma 6: Consider $\mathbf{A}\bigcap\mathbf{B}$; $\mu_{\sigma^*}(\mathbf{A}\bigcap\mathbf{B}) = 1$. Suppose that there are infinitely many rejections in $h_\infty \in \mathbf{A}\bigcap\mathbf{B}$. Since $h_\infty \in \mathbf{A}$, from Lemma 5 it follows that all players

make infinite rejections. It means that there exists $\tilde{s}_1(\geq \tilde{s}_0)$ such that for all $s \geq \tilde{s}_1$, all players' epoch stages are no less than $\tilde{s}_0$ through maximum epoch s. Therefore, for all $s \geq \tilde{s}_1$, $\mathbf{d}_m^{\alpha_s}$ =the number of times that AES has been reached in the first m ER$(s)-$intervals. In addition, since $h_\infty \in \mathbf{B}$, there exists $\tilde{s}_2(\geq \tilde{s}_1)$ such that, for all $s \geq \tilde{s}_2$, $\mathbf{d}_m^{\alpha_s} \geq (\tilde{p}_s - \frac{1}{2}\underline{p}_s)m$ for all $m \geq \frac{R_s}{2\bar{c}}$. Since $\tilde{p}_s \geq \underline{p}_s$ for any sufficiently large s and there are at least $\frac{R_s}{2\bar{c}}$ ER$(s)-$intervals (in maximum epoch) for all s, it implies that there exists $s'(\geq \tilde{s}_2)$ such that for all $s \geq s'$, $\mathbf{d}_m^{\alpha_s} \geq \frac{1}{2}\underline{p}_s m$ for $m = \frac{R_s}{2\bar{c}}$. It means that AES is reached at least $\frac{1}{2}\underline{p}_s(\underline{R}_s/2\bar{c})$ times in the first $\underline{R}_s/2\bar{c}$ ER$(s)-$intervals (in maximum epoch s). This completes the proof. ■

11.3 No rejection from some period

• Let h_{T-1} be a realized past history such that (1) time T is the first period of a cycle (of player i) and (2) the $n-$th rejection (of player i from the beginning of the repeated game) occurred in the previous test phase. Note that player i has formed a new belief denoted by f^i in the previous formation phase. Furthermore, player i keeps being in the same epoch, say, the $s-$th epoch, at least until the $(n+1)-$st rejection occurs; f^i is generated by $\mathcal{P}_s^i$. Then, $\{\mathcal{P}_{s+q}^i\}_{q=0}^\infty$ will be employed in test phases until the $(n+1)-$st rejection occurs: for each $q = 0, 1, 2, \cdots$ and each $\alpha \in \mathcal{P}_{s+q}^i$, the $\alpha-$test starts from the $(q+1)-$st test phase *after* h_{T-1} (unless the $(n+1)-$st rejection occurs). Recall that the $\alpha-$test is said to be *effective* at time T if the $\alpha-$test is collecting samples at time T. Then, for each $q = 0, 1, 2, \cdots$ and each $\alpha \in \mathcal{P}_{s+q}^i$, define the corresponding class $\alpha(s, q)$ such that $h_{\bar{T}} \in \alpha(s, q)$ if and only if (1) $h_{T-1} \leq h_{\bar{T}}$, (2) $h_{\bar{T}} \in \alpha$, (3) the $\alpha-$test is effective at time $\bar{T}+1$, and (4) for all $h_{T-1} \leq h_t \leq h_{\bar{T}}$ such that $h_t \in \alpha$ and the $\alpha-$test is effective at time

$t+1$,

$$\|f_j^i(h_t) - \sigma_j^*(h_t)\| \leq \frac{\xi_s^i}{4} \text{ for all } j \neq i.$$

Then, let $\mathbf{d}_{j,m}^{\alpha(s,q)}[a_j]$ denote the number of times that a_j has been realized in the first m $\alpha(s,q)-$active periods in which the $\alpha-$test is effective and let $\mathbf{d}_{j,m}^{\alpha(s,q)} := (\mathbf{d}_{j,m}^{\alpha(s,q)}[a_j])_{a_j}$. Define $\mathbf{C}_m^{\alpha(s,q)} := \{h_\infty \mid \mathcal{T}_m^{\alpha(s,q)} < \infty, \exists j \neq i(\|\mathbf{d}_{j,m}^{\alpha(s,q)} / m - f_j^i(\alpha)\| > \xi_s^i)\}$. Furthermore, let $\bar{f}_j^i(\alpha)[a_j] := f_j^i(\alpha)[a_j] + \frac{\xi_s^i}{4}$ and $\underline{f}_j^i(\alpha)[a_j] := f_j^i(\alpha)[a_j] - \frac{\xi_s^i}{4}$. Let

$$\mathbf{D}_m^{\alpha(s,q)}(j,a_j) := \{h_\infty \mid \mathcal{T}_m^{\alpha(s,q)} < \infty, \frac{\mathbf{d}_{j,m}^{\alpha(s,q)}[a_j]}{m} > \bar{f}_j^i(\alpha)[a_j] + \frac{\xi_s^i}{2} \text{ or } \frac{\mathbf{d}_{j,m}^{\alpha(s,q)}[a_j]}{m} < \underline{f}_j^i(\alpha)[a_j] - \frac{\xi_s^i}{2}\}.$$

Then, we easily obtain that $\mathbf{C}_m^{\alpha(s,q)} \subset \bigcup_{j\neq i}\bigcup_{a_j} \mathbf{D}_m^{\alpha(s,q)}(j,a_j)$. From this and Lemma 4 it follows that $\mu_{\sigma^*}(\mathbf{C}_m^{\alpha(s,q)} \mid h_{T-1}) \leq \mu_{\sigma^*}(\bigcup_{j\neq i}\bigcup_{a_j} \mathbf{D}_m^{\alpha(s,q)}(j,a_j) \mid h_{T-1}) \leq (\sum_{j\neq i} \#A_j) 2\exp(-\frac{1}{2}m(\xi_s^i)^2)$.

For each i, we define a stochastic process $\{X_n^i\}_n$ and a filtration $\{\mathcal{H}_n^i\}_n$ as follows: $X_n^i := 1$ if the $n-$th rejection (of player i) is of type I error, and $X_n^i = 0$ otherwise; see Section 7.2 for type I error. Moreover, let $\mathcal{H}_n^i := \sigma(X_1^i, \cdots, X_n^i)$, i.e., the $\sigma-$algebra generated by $(X_1^i, \cdots, X_n^i)$. By the definition, $E[X_{n+1}^i \mid \mathcal{H}_n^i] \leq$the probability that the $(n+1)-$st rejection is of type I error conditional on $\mathcal{H}_n^i$. Notice that, for all $h_\infty > h_{T-1}$, if $X_{n+1}^i(h_\infty) = 1$, then $h_\infty \in \bigcup_{q=0}^\infty \bigcup_{\alpha\in\mathcal{P}_{s+q}^i} \bigcup_{m=\underline{m}_{s+q}^i}^\infty \mathbf{C}_m^{\alpha(s,q)}$. Therefore,

$$\begin{aligned}
E[X_{n+1}^i \mid h_{T-1}] &\leq \mu_{\sigma^*}(\bigcup_{q=0}^{\infty} \bigcup_{\alpha\in\mathcal{P}_{s+q}^i} \bigcup_{m=\underline{m}_{s+q}^i}^{\infty} \mathbf{C}_m^{\alpha(s,q)} \mid h_{T-1}) \\
&\leq \sum_{q=0}^{\infty} (\#\mathcal{P}_{s+q}^i) \sum_{m\geq \underline{m}_{s+q}^i} \mu_{\sigma^*}(\mathbf{C}_m^{\alpha(s,q)} \mid h_{T-1}) \\
&\leq (\sum_{j\neq i} \#A_j) \sum_{q=0}^{\infty} (\#\mathcal{P}_{s+q}^i) \sum_{m\geq \underline{m}_{s+q}^i} 2\exp(-\frac{1}{2}m(\xi_s^i)^2).
\end{aligned}$$

This inequality holds for any realized history h_{T-1} satisfying (1) and (2). Note also that by the definition of epochs, the $(n+1)-$th rejection (of player i) must always occur

in the *same* epoch, say, the $s-$th epoch (of player i), since the number of rejections completely determines the switching of epochs for each player. From these it follows that, for all h_∞,

$$E[X^i_{n+1} \mid \mathcal{H}^i_n] \leq (\sum_{j\neq i} \#A_j) \sum_{q=0}^{\infty} (\#\mathcal{P}^i_{s+q}) \sum_{m\geq \underline{m}^i_{s+q}} 2\exp(-\frac{1}{2}m(\xi^i_s)^2).$$

Therefore, letting $J^i_0 := 0$ and $J^i_s := \sum_{v=1}^{s} R^i_v$ for $s \geq 1$, for *all* h_∞,

$$\begin{aligned}
\sum_{n=1}^{\infty} E[X^i_{n+1} \mid \mathcal{H}^i_n] &= \sum_{s=1}^{\infty} \sum_{n=J^i_{s-1}}^{J^i_s-1} E[X^i_{n+1} \mid \mathcal{H}^i_n] \\
&\leq \sum_{s=1}^{\infty} \sum_{n=J^i_{s-1}}^{J^i_s-1} (\sum_{j\neq i} \#A_j) \sum_{q=0}^{\infty} (\#\mathcal{P}^i_{s+q}) \sum_{m\geq \underline{m}^i_{s+q}} 2\exp(-\frac{1}{2}m(\xi^i_s)^2) \\
&\leq 2\sum_{s=1}^{\infty} R^i_s (\sum_{j\neq i} \#A_j) \sum_{q=0}^{\infty} (\#\mathcal{P}^i_{s+q}) \sum_{m\geq \underline{m}^i_{s+q}} \exp(-\frac{1}{2}m(\xi^i_s)^2) \\
&\leq 2(\sum_{j\neq i} \#A_j) \sum_{s=1}^{\infty}\sum_{q=0}^{\infty} R^i_s (\#\mathcal{P}^i_{s+q}) \sum_{m\geq \underline{m}^i_{s+q}} \exp(-\frac{1}{2}m(\xi^i_{s+q})^2) \\
&\leq 2(\sum_{j\neq i} \#A_j) \sum_{s=1}^{\infty}\sum_{q=0}^{\infty} R^i_{s+q} (\#\mathcal{P}^i_{s+q}) \sum_{m\geq \underline{m}^i_{s+q}} \exp(-\frac{1}{2}m(\xi^i_{s+q})^2) \\
&= 2(\sum_{j\neq i} \#A_j) \sum_{s=1}^{\infty} \exp(-s)(1-\exp(-1))^{-1} \\
&= 2(\sum_{j\neq i} \#A_j) \exp(-1)(1-\exp(-1))^{-2} < \infty.
\end{aligned}$$

The fifth inequality obtains because $R^i_s \leq R^i_{s+1}$ for all s, and the sixth equality holds because of Condition 7. The other inequalities are obvious. Therefore, $\sum_{n=1}^{\infty} E[X^i_{n+1} \mid \mathcal{H}^i_n] < \infty$ for all h_∞. This result leads us to obtain Lemma 7.

Proof of Lemma 7: Let $\mathbf{E}_i := \{h_\infty \mid$there are infinitely many player i's test rejections of type I error$\}$. Notice that, by the definition of $\{X^i_n\}_n$, $\mathbf{E}_i = \{h_\infty \mid \sum_{n=1}^{\infty} X^i_n = \infty\}$. Furthermore, by a generalized argument of the Borel-Cantelli Lemma (see Section 5 in

Chapter 7 of Shiryaev (1984)),[30]

$$\{h_\infty \mid \sum_{n=1}^{\infty} X_n^i = \infty\} = \{h_\infty \mid \sum_{n=1}^{\infty} E[X_n^i \mid \mathcal{H}_n^i] = \infty\},\ \mu_{\sigma^*} - a.s.$$

However, then, $\mu_{\sigma^*}(\{h_\infty \mid \sum_{n=1}^{\infty} E[X_n^i \mid \mathcal{H}_n^i] = \infty\}) = \mu_{\sigma^*}(\emptyset) = 0$. Therefore,

$$\mu_{\sigma^*}(\mathbf{D}_i) = \mu_{\sigma^*}(\{h_\infty \mid \sum_{n=1}^{\infty} E[X_n^i \mid \mathcal{H}_n^i] = \infty\}) = 0.$$

Thus, $\mu_{\sigma^*}(\bigcup_i \mathbf{E}_i) = 0$. From this the desired result immediately follows. ■

• Define a class γ_s^1 such that $h_T \in \gamma_s^1$ if and only if at time T, rejection occurs for the *first* time *after* an ER$(s)-$interval in which AES has been reached. Define a class $\gamma_s^2 (\subset \gamma_s^1)$ such that $h_T \in \gamma_s^2$ if and only if (1) the player who made the first rejection, say, player j, has formed the *same* belief as the previous one, i.e., an approximate equilibrium belief and (2) at time T, rejection occurs the *second* time *after* the ER$(s)-$interval; thus, the AES survives (at least) until time T. For $3 \leq y \leq I-1$, define γ_s^y inductively as follows: $h_T \in \gamma_s^y$ if and only if (1) the same belief as theprevious one, i.e., an approximate equilibrium belief, has been formed after any of the last $(y-1)$ rejections and (2) at time T, rejection occurs the $y-$th time after the ER$(s)-$interval; thus, the AES survives (through the $(y-1)$ rejections) until time T. Note also that $\gamma_s^{y+1} \subset \gamma_s^y$ for all $1 \leq y \leq I-2$ and all s.

As in the case of reaching an AES, the probability of forming the same belief again just after the $y-$th rejection is (at least) $\min_j(\Pi_k \underline{l}_k)^{N_s^j} = (\Pi_k \underline{l}_k)^{\bar{N}_s}$. Let $\check{p}_s := (\Pi_k \underline{l}_k)^{\bar{N}_s}$. Let $\mathbf{d}_m^{\gamma_s^y}(h_\infty)$ denote the number of times that the same beliefs as approximate equilibrium beliefs have been formed after the first m γ_s^y-active periods; in other words, $\mathbf{d}_m^{\gamma_s^y}(h_\infty)$ is the number of times that AES has survived after the first m γ_s^y-active periods. Then, define $\mathbf{F}_m^{s,y} := \{h_\infty \mid \mathcal{T}_m^{\gamma_s^y} < \infty, \mathbf{d}_m^{\gamma_s^y} / m < \check{p}_s - \frac{1}{2}\underline{p}_s\}$. By Lemma 4, $\mu_{\sigma^*}(\mathbf{F}_m^{s,y}) \leq \exp(-\frac{1}{2}m(\underline{p}_s)^2)$.

[30] For any measurable sets $\mathbf{X}$ and $\mathbf{Y}$, $\mathbf{X} = \mathbf{Y}$, $\mu_{\sigma^*} - a.s.$ if and only if $\mu_{\sigma^*}((\mathbf{X} \setminus \mathbf{Y}) \cup (\mathbf{Y} \setminus \mathbf{X})) = 0$.

From Condition 2 it follows that for all k, $\underline{p}_s = (\frac{1}{s})^{s\bar{N}_s} \geq (\frac{1}{s})^{s\bar{n}N_s^k} = (p_s^k)^{\underline{n}} \geq (p_s^k)^s$ for any sufficiently large s. Thus, for all $1 \leq y \leq I-1$ and all k, $\frac{1}{2\bar{c}}(\frac{1}{2}\underline{p}_s)^y \geq \frac{1}{s}(\frac{1}{2}(p_s^k)^s)^I = w_s^k$ for any sufficiently large s. Furthermore, $\check{p}_s \geq \underline{p}_s$ for any sufficiently large s. Note also that, for any sufficiently large s, there exists k_s such that

$$\frac{\underline{R}_s}{2\bar{c}} = \frac{R_s^{k_s}}{2\bar{c}} \geq \frac{R_s^{k_s}}{s}.$$

From these we obtain that $(\frac{1}{2}\underline{p}_s)^y(\underline{R}_s/2\bar{c}) \geq w_s^{k_s}R_s^{k_s}$. From this and Condition 6 it follows that, for any sufficiently large s',

$$\begin{aligned}
\mu_{\sigma^*}(\bigcap_{s'\geq 1}\bigcup_{s\geq s'}\bigcup_{m\geq(\frac{1}{2}\underline{p}_s)^y(\underline{R}_s/2\bar{c})}\mathbf{F}_m^{s,y}) &\leq \mu_{\sigma^*}(\bigcup_{s\geq s'}\bigcup_{m\geq(\frac{1}{2}\underline{p}_s)^y(\underline{R}_s/2\bar{c})}\mathbf{F}_m^{s,y}) \\
&\leq \sum_{s\geq s'}\sum_{m\geq(\frac{1}{2}\underline{p}_s)^y(\underline{R}_s/2\bar{c})}\mu_{\sigma^*}(\mathbf{F}_m^{s,y}) \\
&\leq \sum_{s\geq s'}\sum_{m\geq(\frac{1}{2}\underline{p}_s)^y(\underline{R}_s/2\bar{c})}\exp(-\frac{1}{2}m(\underline{p}_s)^2) \\
&\leq \sum_{s\geq s'}\sum_{m\geq w_s^{k_s}R_s^{k_s}}\exp(-\frac{1}{2}m((p_s^{k_s})^s)^2) \\
&\leq \sum_{s\geq s'}\exp(-s) = (1-\exp(-1))^{-1}\exp(-s').
\end{aligned}$$

Therefore, $\mu_{\sigma^*}(\bigcap_{s'\geq 1}\bigcup_{s\geq s'}\bigcup_{m\geq(\frac{1}{2}\underline{p}_s)^y(\underline{R}_s/2\bar{c})}\mathbf{F}_m^{s,y}) = 0$ for all $1 \leq y \leq I-1$. Define

$$\mathbf{F} := \bigcap_{y=1}^{I-1}\bigcup_{s'\geq 1}\bigcap_{s\geq s'}\bigcap_{m\geq(\frac{1}{2}\underline{p}_s)^y(\underline{R}_s/2\bar{c})}(\mathbf{F}_m^{s,y})^c.$$

Then, $\mu_{\sigma^*}(\mathbf{F}) = 1$. From this we obtain Lemma 8.

Proof of Lemma 8: Let $\mathbf{E} := \bigcap_i(\mathbf{E}_i)^c$; by the proof of Lemma 7, $\mu_{\sigma^*}(\mathbf{E}) = 1$. Then, $\mu_{\sigma^*}(\mathbf{A}\bigcap\mathbf{B}\bigcap\mathbf{E}\bigcap\mathbf{F}) = 1$. Suppose that there are infinitely many rejections in $h_\infty \in \mathbf{A}\bigcap\mathbf{B}\bigcap\mathbf{E}\bigcap\mathbf{F}$. Then, there exists s'' such that, for each $s \geq s''$, for all $1 \leq y \leq I-1$, $\mathbf{d}_m^{\gamma_s^y} \geq (\check{p}_s - \frac{1}{2}\underline{p}_s)m \geq \frac{1}{2}\underline{p}_s m$ for all $m \geq (\frac{1}{2}\underline{p}_s)^y(\underline{R}_s/2\bar{c})$. However, then, from Lemma

6 it follows that, for all $s \geq s''$, AES is reached at least $\frac{1}{2}\underline{p}_s(\underline{R}_s/2\bar{c})$ times in the first $(\underline{R}_s/2\bar{c})$ ER$(s)-$intervals. It means that γ_s^1 is active at least $\frac{1}{2}\underline{p}_s(\underline{R}_s/2\bar{c})$ times. Then, the corresponding player forms the same belief as the previous one and then AES survives at least $\frac{1}{2}\underline{p}_s(\frac{1}{2}\underline{p}_s\underline{R}_s/2\bar{c})$ times after $\frac{1}{2}\underline{p}_s(\underline{R}_s/2\bar{c})$ γ_s^1-active periods. It means that γ_s^2 is active at least $(\frac{1}{2}\underline{p}_s)^2(\underline{R}_s/2\bar{c})$ times. Again, AES survives at least $\frac{1}{2}\underline{p}_s(\frac{1}{2}\underline{p}_s)^2(\underline{R}_s/2\bar{c})$ times after $(\frac{1}{2}\underline{p}_s)^2(\underline{R}_s/2\bar{c})$ γ_s^2-active periods. We can repeat this argument so that AES survives through the first $(I-1)$ rejections after an ER$(s)-$interval (in which AES has been reached) at least $(\frac{1}{2}\underline{p}_s)^I(\underline{R}_s/2\bar{c})$ times. ■

Remark 6 *From Condition 6 it follows that* $(\frac{1}{2}\underline{p}_s)^I\underline{R}_s \to \infty$ *as* $s \to \infty$.

11.4 Proof of Proposition 1

For any positive integer L, let $\Delta_L^i := \{\pi_i \mid$for all $a_i \in A_i$, there exists a nonnegative integer l such that $\pi_i[a_i] = \frac{l}{L}\}$ and $S_L^i(\pi_i) := \{\pi_i' \mid \|\pi_i' - \pi_i\| \leq \frac{2}{L}\}$. Notice that $\bigcup_{\pi_i \in \Delta_L^i} S_L^i(\pi_i) = \Delta(A_i)$, and that for any subset Δ of $\Delta(A_i)$ with its diameter no more than $\frac{1}{2L}$, i.e., $diam(\Delta) \leq \frac{1}{2L}$,[31] there exists $\pi_i \in \Delta_L^i$ such that $\Delta \subset S_L^i(\pi_i)$. Next, let $L_\epsilon := \min\{L \mid \frac{2}{L} \leq \frac{\epsilon}{6}\}$ and $\mathcal{S}^i(L_\epsilon) := \{S_{L_\epsilon}^i(\pi_j) \mid \pi_i \in \Delta_{L_\epsilon}^i\}$. Then, for all i, all $j \neq i$, all s, all q, all $\alpha \in \mathcal{P}_{s+q}^i$, and all $S^j \in \mathcal{S}^j(L_\epsilon)$, we define a class $\alpha(S^j)$ as follows: $h_T \in \alpha(S^j)$ if and only if (1) $h_T \in \alpha$, (2) the $\alpha-$test has started from the *first* test phase of employing $\mathcal{P}_{s+q}^i$ (in epoch s of player i) in h_T, i.e., the *first* $\alpha-$test (with the least sample size $\underline{m}_s^i$) has started in h_T, (3) the first $\alpha-$test is effective at time $T+1$, and (4) for all $h_t \leq h_T$ such that $h_t \in \alpha$ and the first $\alpha-$test was effective at time $t+1$, $\pi_j[a_j] - \frac{\epsilon}{6} \leq \sigma_j^*(h_t)[a_j] \leq \pi_j[a_j] + \frac{\epsilon}{6}$ for all $a_j \in A_j$, where π_j is the center of $S^j(= S_{L_\epsilon}^j(\pi_j))$. Let $\mathbf{d}_{j,m}^{\alpha(S^j)}[a_j]$ denote the number of times that a_j has been realized in the first m $\alpha(S^j)-$active periods. Then, for all i, all

[31] The diameter of Δ is defined by $diam(\Delta) := \sup\{\|\pi_j - \pi_j'\| \mid \pi_j, \pi_j' \in \Delta\}$.

$j \neq i$, all s, all q, all $\alpha \in \mathcal{P}^i_{s+q}$, all $S^j \in \mathcal{S}^j(L_\epsilon)$, and all m, define

$$\mathbf{P}^i_j(s,q,\alpha,S^j,m) := \{h_\infty \mid \mathcal{T}^{\alpha(S^j)}_m < \infty, \exists a_j(\frac{\mathbf{d}^{\alpha(S^j)}_{j,m}[a_j]}{m} > \pi_j[a_j] + \frac{\epsilon}{3} \text{ or } \frac{\mathbf{d}^{\alpha(S^j)}_{j,m}[a_j]}{m} < \pi_j[a_j] - \frac{\epsilon}{3})\}.$$

Let $\mathbf{U} := \{h_\infty \mid$there are at most finite rejections in $h_\infty\}$. We say that players' (temporary) beliefs are significantly different from players' true strategies infinitely many times in h_∞ if, for infinitely many $h(< h_\infty)$, there exists i and $j(\neq i)$ such that player i's belief f^i and player j's true strategy σ^*_j are significantly different, i.e., $\|f^i_j(h) - \sigma^*_j(h)\| > \epsilon$. Let $\mathbf{V} := \{h_\infty \mid$players' beliefs are significantly different from players' true strategies infinitely many times in $h_\infty\}$. Then, we obtain that

$$\begin{aligned}
\mathbf{U} \bigcap \mathbf{V} &\subset \bigcup_{i=1}^{I} \bigcup_{j\neq i} \bigcup_{s=1}^{\infty} \bigcup_{\bar{q}=1} \bigcap_{q\geq\bar{q}} \bigcup_{\alpha\in\mathcal{P}^i_{s+q}} \bigcup_{S^j\in\mathcal{S}^j(L_\epsilon)} \bigcup_{m\geq \underline{m}^i_{s+q}} \mathbf{P}^i_j(s,q,\alpha,S^j,m) \\
&\subset \bigcup_{i=1}^{I} \bigcup_{j\neq i} \bigcup_{s=1}^{\infty} \bigcap_{\bar{q}=1}^{\infty} \bigcup_{q=\bar{q}}^{\infty} \bigcup_{\alpha\in\mathcal{P}^i_{s+q}} \bigcup_{S^j\in\mathcal{S}^j(L_\epsilon)} \bigcup_{m\geq \underline{m}^i_{s+q}} \mathbf{P}^i_j(s,q,\alpha,S^j,m).
\end{aligned}$$

The second inclusion is obvious. We show the first inclusion. Suppose that $h_\infty \in \mathbf{U} \bigcap \mathbf{V}$. Since $h_\infty \in \mathbf{U}$, there exists a (temporary) belief profile $(f^i)_i$ such that each player i keeps f^i *forever* from some period, say, time $\tilde{T}$; each player i also keeps being in some epoch, say, the s^i-th epoch, forever from time $\tilde{T}$. On the other hand, since $h_\infty \in \mathbf{V}$, there exist i_0 and $j_0(\neq i_0)$ such that $\|f^{i_0}_{j_0}(h_{t_k}) - \sigma^*_{j_0}(h_{t_k})\| > \epsilon$ for infinitely many $h_{t_k} < h_\infty$. However, then, from Properties (3.1) and (3.2) of $\{\mathcal{P}^i_s\}_s$ it follows that, for all i, all $j(\neq i)$ and all s, there exists $\hat{q}(i,j,s)$ such that, for all $q \geq \hat{q}(i,j,s)$, $\mathcal{P}^i_{s+q}$ is a conditioning rule of f^i and $\mathcal{P}^i_{s+q}$ is also a $\frac{\epsilon}{96}-$approximate conditioning rule of σ^f_j: for all $\alpha \in \mathcal{P}^i_{s+q}$ and all $h, h' \in \alpha$, $f^i_j(h) = f^i_j(h')$ and $\|\sigma^f_j(h) - \sigma^f_j(h')\| \leq \frac{\epsilon}{96}$. Moreover, it follows from Condition 5 and Lemma 3 that, for all i, $\|\sigma^f_i(h_T) - \sigma^*_i(h_T)\| \to 0$ as $T \to \infty$. Thus, there exists $\hat{T} \geq \tilde{T}$ such that, for all i, all $j \neq i$, all s and all $q \geq \hat{q}(i,j,s)$, if $h_t, h_{t'} \in \alpha$

for some $\alpha \in \mathcal{P}^i_{s+q}$ and $t, t' \geq \hat{T}$, then $\|\sigma^*_j(h_t) - \sigma^*_j(h_{t'})\| \leq \frac{\epsilon}{48}$: that is, for all i, all $j \neq i$, all s, all $q \geq \hat{q}(i,j,s)$ and all $\alpha \in \mathcal{P}^i_{s+q}$, $diam(\{\sigma^*_j(h_t) \mid t \geq \hat{T}, h_t \in \alpha, h_t < h_\infty\}) \leq \frac{\varepsilon}{48} \leq \frac{1}{2L_\epsilon}$. Then, for all i, all $j \neq i$, all s, all $q \geq \hat{q}(i,j,s)$ and all $\alpha \in \mathcal{P}^i_{s+q}$, there exists $S^j \in \mathcal{S}^j(L_\epsilon)$ such that $\{\sigma^*_j(h_t) \mid t \geq \hat{T}, h_t \in \alpha, h_t < h_\infty\} \subset S^j$, as noted above. Also, since $\|f^{i_0}_{j_0}(h_{t_k}) - \sigma^*_{j_0}(h_{t_k})\| > \epsilon$ for infinitely many $h_{t_k} < h_\infty$, it is obvious that there exists $\bar{q} \geq \hat{q}(i_0, j_0, s^{i_0})$ such that, for all $q \geq \bar{q}$, there exists $\bar{\alpha} \in \mathcal{P}^{i_0}_{s^{i_0}+q}$ such that (I) the first $\bar{\alpha}-$test starts after time $\hat{T}$ (in epoch s^{i_0}) and (II) $h_{t_{k_n}} \in \bar{\alpha}$ for an *infinite* subsequence $\{h_{t_{k_n}}\}_n$ of $\{h_{t_k}\}_k$. Therefore, it follows from (II) that $\|f^{i_0}_{j_0}(h_t) - \sigma^*_{j_0}(h_t)\| > \epsilon - \frac{\epsilon}{48} = \frac{47}{48}\epsilon$ for all $h_t \in \bar{\alpha}$ such that $h_t < h_\infty$ and $t \geq \hat{T}$. Thus, from these it is derived that, for all $q \geq \bar{q}$, there exist $\bar{\alpha} \in \mathcal{P}^{i_0}_{s^{i_0}+q}$ and $\bar{S}^{j_0} \in \mathcal{S}^{j_0}(L_\epsilon)$ such that (i) the first $\bar{\alpha}-$test starts after time $\hat{T}$ (in epoch s^{i_0}), (ii) $\{\sigma^*_{j_0}(h_t) \mid t \geq \hat{T}, h_t \in \bar{\alpha}, h_t < h_\infty\} \subset \bar{S}^{j_0}$, (iii) $\|f^{i_0}_{j_0}(h) - \sigma^*_{j_0}(h)\| > \frac{47}{48}\epsilon$ for all $h < h_\infty$ such that $h \in \bar{\alpha}(\bar{S}^{j_0})$, and (iv) $\#\{h_t \mid t \geq \hat{T}, h_t \in \bar{\alpha}, h_t < h_\infty\} = \infty$. From (i), (ii), (iv) and the definition of the learning procedure we obtain that $\#\{h \mid h \in \bar{\alpha}(\bar{S}^{j_0}), h < h_\infty\} = m^{\bar{\alpha}} \geq \underline{m}^{i_0}_{s^{i_0}+q}$, which implies that the first $\bar{\alpha}-$test obtains enough samples but does not reject f^i in h_∞. In addition, it implies that, in any $\bar{\alpha}-$effective period in which the first $\bar{\alpha}-$test is effective, $\bar{\pi}_{j_0}[a_{j_0}] - \frac{\varepsilon}{6} \leq \sigma^*_{j_0}(h_t)[a_{j_0}] \leq \bar{\pi}_{j_0}[a_{j_0}] + \frac{\varepsilon}{6}$ for all $a_{j_0} \in A_{j_0}$, where $\bar{\pi}_{j_0} (\in \Delta^{j_0}_{L_\epsilon})$ is the center of $\bar{S}^{j_0} (= S^{j_0}_{L_\epsilon}(\bar{\pi}_{j_0}))$. Therefore, $h_\infty \in \mathbf{P}^{i_0}_{j_0}(s^{i_0}, q, \bar{\alpha}, \bar{S}^{j_0}, m^{\bar{\alpha}})$. Otherwise, $\|\mathbf{d}^{\bar{\alpha}(\bar{S}^{j_0})}_{j_0, m^{\bar{\alpha}}} / m^{\bar{\alpha}} - \pi_{j_0}\| \leq \frac{\epsilon}{3}$. However, then, since there is no rejection from time $\tilde{T}$, the first $\bar{\alpha}-$test must be passed, which means that $\|f^{i_0}_{j_0}(h) - \mathbf{d}^{\bar{\alpha}(\bar{S}^{j_0})}_{j_0, m^{\bar{\alpha}}} / m^{\bar{\alpha}}\| \leq \xi^{i_0}_{s^{i_0}} \leq \frac{\epsilon}{3}$ (or $\leq \bar{\xi}^{i_0} \leq \frac{\epsilon}{3}$) for all $h \in \bar{\alpha}$. Furthermore, recall that $\pi_{j_0}[a_{j_0}] - \frac{\epsilon}{6} \leq \sigma^*_{j_0}(h)[a_{j_0}] \leq \pi_{j_0}[a_{j_0}] + \frac{\epsilon}{6}$ for all $a_{j_0} \in A_{j_0}$ and all $h < h_\infty$ such that $h \in \bar{\alpha}(\bar{S}^{j_0})$. Therefore, $\|f^{i_0}_{j_0}(h) - \sigma^*_{j_0}(h)\| \leq \|f^{i_0}_{j_0}(h) - \mathbf{d}^{\bar{\alpha}(\bar{S}^{j_0})}_{j_0, m^{\bar{\alpha}}} / m^{\bar{\alpha}}\| + \|\mathbf{d}^{\bar{\alpha}(\bar{S}^{j_0})}_{j_0, m^{\bar{\alpha}}} / m^{\bar{\alpha}} - \pi_{j_0}\| + \|\pi_{j_0} - \sigma^*_{j_0}(h)\| \leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{6} = \frac{5}{6}\epsilon = \frac{40}{48}\epsilon$ for all $h < h_\infty$ such that $h \in \bar{\alpha}(\bar{S}^{j_0})$. This is a contradiction to (iii), i.e., $\|f^{i_0}_{j_0}(h) - \sigma^*_{j_0}(h)\| > \frac{47}{48}\epsilon$ for all $h < h_\infty$ such that $h \in \bar{\alpha}(\bar{S}^{j_0})$.

Therefore, $h_\infty \in \bigcup_{i=1}^{I}\bigcup_{j\neq i}\bigcup_{s=1}^{\infty}\bigcup_{\bar{q}\geq 1}\bigcap_{q\geq\bar{q}}\bigcup_{\alpha\in\mathcal{P}^i_{s+q}}\bigcup_{S^j\in\mathcal{S}^j(L_\epsilon)}\bigcup_{m\geq\underline{m}^i_{s+q}}\mathbf{P}^i_j(s,q,\alpha,S^j,m)$.

Finally, from Lemma 4 and Conditions 3 and 8 it follows that, for all i, all $j\neq i$, all s, and all $\bar{q}$,

$$\begin{aligned}
&\mu_{\sigma^*}(\bigcup_{q=\bar{q}}^{\infty}\bigcup_{\alpha\in\mathcal{P}^i_{s+q}}\bigcup_{S^j\in\mathcal{S}^j(L_\epsilon)}\bigcup_{m\geq\underline{m}^i_{s+q}}\mathbf{P}^i_j(s,q,\alpha,S^j,m))\\
\leq\ & \sum_{q\geq\bar{q}}(\#\mathcal{P}^i_{s+q})(\#\mathcal{S}^j(L_\epsilon))\sum_{m\geq\underline{m}^i_{s+q}}(\#A_j)2\exp(-\frac{1}{2}m(\frac{\epsilon}{3})^2)\\
=\ & 2(\#\mathcal{S}^j(L_\epsilon))(\#A_j)(1-\exp(-\frac{1}{2}(\frac{\epsilon}{3})^2))^{-1}\sum_{q\geq\bar{q}}(\#\mathcal{P}^i_{s+q})\exp(-\frac{1}{2}\underline{m}^i_{s+q}(\frac{\epsilon}{3})^2)\\
\leq\ & 2(\#\mathcal{S}^j(L_\epsilon))(\#A_j)(1-\exp(-\frac{1}{2}(\frac{\epsilon}{3})^2))^{-1}\sum_{q\geq\bar{q}}(\#\mathcal{P}^i_{s+q})\exp(-\frac{1}{2}\underline{m}^i_{s+q}(\xi^i_{s+q})^2)\\
\leq\ & 2(\#\mathcal{S}^j(L_\epsilon))(\#A_j)(1-\exp(-1))^{-1}\sum_{q\geq\bar{q}}\exp(-s-q)\\
\leq\ & 2(\#\mathcal{S}^j(L_\epsilon))(\#A_j)(1-\exp(-1))^{-2}\exp(-s-\bar{q}).
\end{aligned}$$

Thus, for all $\bar{q}$,

$$\begin{aligned}
&\mu_{\sigma^*}(\bigcap_{\bar{q}=1}^{\infty}\bigcup_{q=\bar{q}}^{\infty}\bigcup_{\alpha\in\mathcal{P}^i_{s+q}}\bigcup_{S^j\in\mathcal{S}^j(L_\epsilon)}\bigcup_{m\geq\underline{m}^i_{s+q}}\mathbf{P}^i_j(s,q,\alpha,S^j,m))\\
\leq\ & 2(\#\mathcal{S}^j(L_\epsilon))(\#A_j)(1-\exp(-1))^{-2}\exp(-s-\bar{q})).
\end{aligned}$$

Therefore, letting $\bar{q}\to\infty$, $\mu_{\sigma^*}(\bigcap_{\bar{q}=1}^{\infty}\bigcup_{q=\bar{q}}^{\infty}\bigcup_{\alpha\in\mathcal{P}^i_{s+q}}\bigcup_{S^j\in\mathcal{S}^j(L_\epsilon)}\bigcup_{m\geq\underline{m}^i_{s+q}}\mathbf{P}^i_j(s,q,\alpha,S^j,m)) = 0$ for all i, all $j(\neq i)$ and all s. Thus,

$$\mu_{\sigma^*}(\bigcup_{i=1}^{I}\bigcup_{j\neq i}\bigcup_{s=1}^{\infty}\bigcap_{\bar{q}=1}^{\infty}\bigcup_{q=\bar{q}}^{\infty}\bigcup_{\alpha\in\mathcal{P}^i_{s+q}}\bigcup_{S^j\in\mathcal{S}^j(L_\epsilon)}\bigcup_{m\geq\underline{m}^i_{s+q}}\mathbf{P}^i_j(s,q,\alpha,S^j,m))=0.$$

Therefore, $\mu_{\sigma^*}(\mathbf{U}\bigcap\mathbf{V})=0$. Since $\mu_{\sigma^*}(\mathbf{U})=1$, $\mu_{\sigma^*}(\mathbf{U}\bigcap(\mathbf{V})^c)=1$. Finally, for all $h_\infty\in\mathbf{U}$, some temporary beliefs $(f^i)_i$ keep employed forever from some time, say, $\tilde{T}$, which means that, for all i, $f^i(h_T)=\tilde{\rho}^i_*(h_T)$ for all $T\geq\tilde{T}$. In addition, for all $h_\infty\in\mathbf{V}^c$, from some period on, say, time $\check{T}$, players' beliefs $(f^i)_i$ are *not* significantly different from players' true strategies $(\sigma^*_i)_i$: for all i and all $j\neq i$, $\|f^i_j(h_T)-\sigma^*_j(h_T)\|\leq\epsilon$ for all

$T \geq \check{T}$. From these it is easily derived that, for all $h_\infty \in \mathbf{U} \bigcap \mathbf{V}^c$, for all i and all $j(\neq i)$, $\limsup_{T\to\infty} \|\tilde{\rho}^i_{*,j}(h_T) - \sigma^*_j(h_T)\| \leq \epsilon$. ■

12 Appendix C

12.1 Proof of Proposition 2

Take any σ_i and any $\sigma_{-i} \in EG(\{\mathcal{P}^i_s\}_s; \sigma_i)$; let $\sigma := (\sigma_i, \sigma_{-i})$. Then, there exist an index s_0, a $\mu_\sigma-$probability one set $\mathbf{Z}_0$, and a time function $T_0 : \mathbf{Z}_0 \to \mathbb{N}$ such that, for all $\alpha \in \mathcal{P}^i_{s_0}$ and all $h_T, h'_{T'} \in \alpha$, if there exist $h_\infty, h'_\infty \in \mathbf{Z}_0$ such that $h_T < h_\infty$ and $T \geq T_0(h_\infty)$ and $h'_{T'} < h'_\infty$ and $T' \geq T_0(h'_\infty)$, then $\|\sigma_j(h_T) - \sigma_j(h'_{T'})\| < \frac{1}{4}\bar{\xi}^i$ for all $j \neq i$. Accordingly, we provide several definitions. For any $\alpha \in \mathcal{P}^i_{s_0}$, class $\tilde{\alpha}$ is defined as follows: $h_T \in \tilde{\alpha}$ if and only if (1) $h_T \in \alpha$ and (2) there exists $h_\infty \in \mathbf{Z}_0$ such that $h_T < h_\infty$ and $T \geq T_0(h_\infty)$. Next, for all $j \neq i$, let $L^\alpha_j[a_j] := \sup_{h\in\tilde{\alpha}} \sigma_j(h)[a_j]$ and $l^\alpha_j[a_j] := \inf_{h\in\tilde{\alpha}} \sigma_j(h)[a_j]$; obviously, for all $j \neq i$, $L^\alpha_j[a_j] - l^\alpha_j[a_j] \leq \frac{1}{4}\bar{\xi}^i$ for all $\alpha \in \mathcal{P}^i_{s_0}$ and all a_j. Furthermore, for all $j \neq i$, all $s \geq s_0$, all $\beta \in \mathcal{P}^i_s$, and all a_j, let $L^\beta_j[a_j] := L^\alpha_j[a_j]$ and $l^\beta_j[a_j] := l^\alpha_j[a_j]$, where $\alpha \supset \beta$ and $\alpha \in \mathcal{P}^i_{s_0}$.[32] We say that a temporary belief f^i is $\frac{1}{4}\bar{\xi}^i-$close to opponents strategies σ_{-i} or that f^i is $\frac{1}{4}\bar{\xi}^i-$accurate against σ_{-i} if for all $\alpha \in \mathcal{P}^i_{s_0}$ and all $h \in \tilde{\alpha}$, $\|f^i_j(h) - \sigma_j(h)\| \leq \frac{1}{4}\bar{\xi}^i$ for all $j \neq i$. Then, we obtain the following lemma.

Lemma C.1 *With $\mu_\sigma-$probability one, if rejection occurs infinitely many times, then there exists $\bar{s}(\geq s_0 + 1)$ such that, for all $s \geq \bar{s}$, player i chooses a temporary belief that is $\frac{1}{4}\bar{\xi}^i-$close to σ_{-i} (at least) $\frac{1}{2}p^i_s(R^i_s/Z^i_s)$ times in epoch s (of player i): for all $\alpha \in \mathcal{P}^i_{s_0}$ and all $h \in \tilde{\alpha}$, $\|f^i_j(h) - \sigma_j(h)\| \leq \frac{1}{4}\bar{\xi}^i$ for all $j \neq i$.*

Proof. From Conditions 4 and 8 and the modification of player i's prior belief forma-

[32] Since $\mathcal{P}^i_{s_0} \leq \mathcal{P}^i_s$, for all $\beta \in \mathcal{P}^i_s$ there exists a *unique* $\alpha \in \mathcal{P}^i_{s_0}$ such that $\beta \subset \alpha$.

tion process in Section 8, it follows that after the process has proceeded to some epoch $s_1(\geq s_0+1)$, there exists an integer $0 \leq z_0 \leq Z_s^i - 1$ such that the probability of choosing histories that correspond to $\frac{1}{4}\bar{\xi}^i-$accurate beliefs (in any formation phase after *the* $(nZ_s^i + z_0)-$th *rejection for all* $n \in \mathbb{N}$) is always at least $\check{p}_s^i := (\prod_k \underline{l}_k)^{(N_s^i+1)} (\geq p_s^i)$ in each epoch $s(\geq s_1)$.[33] Let $\mathcal{T}_{R_s^i}$ denote the calendar time in which the R_s^i-th rejection in epoch s occurs and $\mathbf{d}_{(R_s^i / Z_s^i)}$ denote the number of times that $\frac{1}{4}\bar{\xi}^i-$accurate belief has been chosen after every Z_s^i rejections (in epoch s). Define $\mathbf{G}_s := \{h_\infty \mid \mathcal{T}_{R_s^i} < \infty, \mathbf{d}_{(R_s^i / Z_s^i)} / (R_s^i / Z_s^i) < \check{p}_s^i - \frac{1}{2}p_s^i\}$. Then, from Lemma 4 it follows that $\mu_\sigma(\mathbf{F}_s) \leq \exp(-\frac{1}{2}(R_s^i / Z_s^i)(p_s^i)^2)$. From this and Condition 6 it is easily derived that for any sufficiently large s',

$$\begin{aligned}
\mu_\sigma(\bigcup_{s \geq s'} \mathbf{G}_s) &\leq \sum_{s \geq s'} \exp(-\frac{1}{2}(R_s^i / Z_s^i)(p_s^i)^2) \\
&\leq \sum_{s \geq s'} \sum_{m \geq w_s^i R_s^i} \exp(-\frac{1}{2} m (p_s^i)^2) \\
&\leq \sum_{s \geq s'} \exp(-s) = (1 - \exp(-1))^{-1} \exp(-s').
\end{aligned}$$

See Remark 7 for the second inequality. Thus, for any sufficiently large s', $\mu(\bigcap_{s' \geq 1} \bigcup_{s \geq s'} \mathbf{G}_s) \leq \mu(\bigcup_{s \geq s'} \mathbf{G}_s) \leq (1 - \exp(-1))^{-1} \exp(-s')$. Therefore, $\mu(\bigcap_{s' \geq 1} \bigcup_{s \geq s'} \mathbf{G}_s) = 0$. Finally, letting $\mathbf{G} := \bigcup_{s' \geq 1} \bigcap_{s \geq s'} (\mathbf{G}_s)^c$, $\mu(\mathbf{G} \bigcap \mathbf{Z}_0) = 1$. Furthermore, from the definitions of $\mathbf{G}$ and $\mathbf{Z}_0$, it is obvious that if there are infinitely many rejections along any $h_\infty \in \mathbf{G} \bigcap \mathbf{Z}_0$, then there exists $\bar{s}(\geq s_1)$ such that, for all $s \geq \bar{s}$, $\mathbf{d}_{(R_s^i / Z_s^i)} \geq (\check{p}_s^i - \frac{1}{2}p_s^i)m \geq \frac{1}{2}p_s^i m$ for all $m \geq R_s^i / Z_s^i$. It means that player i chooses a $\frac{1}{4}\bar{\xi}^i-$accurate belief (against σ_{-i}) at least $\frac{1}{2}p_s^i(R_s^i / Z_s^i)$ times in each epoch $s(\geq \bar{s})$. This completes the proof. ■

Remark 7 *By the definitions in Sections 4.5 and 8,* $Z_s^i = \#\partial\Sigma_{-i}(\mathcal{P}_s^i, \underline{n}_s^i) \leq \#\Sigma_{-i}(\mathcal{P}_s^i, \underline{n}_s^i) = (\#\Delta_{-i}^{\underline{n}_s^i})^{\#\mathcal{P}_s^i}$. *Recall that* $(\#\Delta_{-i}^{\underline{n}_s^i})^{\#\mathcal{P}_s^i} \leq (\#A_{-i})^{N_s^i}$. *Thus,* $Z_s^i \leq s^{N_s^i}$ *for any sufficiently*

[33] Let $\mathbb{N}$ denote the set of all natural numbers.

large s. Then, from $w^i_s := \frac{1}{s}(\frac{1}{2}(p^i_s)^s)^I$ and $p^i_s := (\frac{1}{s})^{sN^i_s}$, and Condition 6 it is derived that $w^i_s R^i_s \leq (\frac{1}{2}(p^i_s)^s)^I R^i_s \leq (\frac{1}{2}p^i_s)(\frac{1}{s})^{sN^i_s} R^i_s \leq \frac{1}{2}(p^i_s)(R^i_s \diagup Z^i_s)$. However, then, by Condition 6, $w^i_s R^i_s \to \infty$ as $s \to \infty$. Therefore, $\frac{1}{2}(p^i_s)(R^i_s \diagup Z^i_s) \to \infty$ as $s \to \infty$.

• For all $s = s_0 + 1, s_0 + 2, \cdots$, all $q = 0, 1, 2, \cdots$, all $\alpha \in \mathcal{P}^i_{s+q}$ and all $d = 1, 2, \cdots$, define the corresponding class $\alpha(s, q, d)$ such that $h_T \in \alpha(s, q, d)$ if and only if (1) time $T + 1$ is in epoch s (of player i), (2) $h_T \in \alpha$, (3) the $d-$th $\alpha-$test (in epoch s) is effective at time $T + 1$, (4) for some $h_\infty > h_T$, $T_0(h_\infty) \leq T$ and the $d-$th $\alpha-$test (in epoch s) obtains the *first* sample between time $T_0(h_\infty) + 1$ and time $T + 1$ in h_T. Let $\mathbf{d}^{\alpha(s,q,d)}_{j,m}[a_j]$ denote the number of times that a_j has been realized in the first m $\alpha(s, q, d)-$active periods, i.e., the number of a_j in the first m samples obtained for the $d-$th $\alpha-$test. Then, define

$$\mathbf{I}^{\alpha(s,q,d)}_m(j, a_j) := \{h_\infty \mid \mathcal{T}^{\alpha(s,q,d)}_m < \infty, \frac{\mathbf{d}^{\alpha(s,q,d)}_{j,m}[a_j]}{m} > L^\alpha_j[a_j] + \frac{\bar{\xi}^i}{4} \text{ or } \frac{\mathbf{d}^{\alpha(s,q,d)}_{j,m}[a_j]}{m} < l^\alpha_j[a_j] - \frac{\bar{\xi}^i}{4}\}$$

and $\mathbf{I}^{\alpha(s,q,d)}_m := \bigcup_{j\neq i} \bigcup_{a_j} \mathbf{I}^{\alpha(s,q,d)}_m(j, a_j)$. Note that, for all $h \in \alpha(s, q, d)$, all $j \neq i$, and all a_j, $l^\alpha_j[a_j] \leq \sigma_j(h)[a_j] \leq L^\alpha_j[a_j]$; $L^\alpha_j[a_j] - l^\alpha_j[a_j] \leq \bar{\xi}^i \diagup 4$. From this and Lemma 4 it follows that $\mu_\sigma(\mathbf{I}^{\alpha(s,q,d)}_m) \leq (\sum_{j\neq i} \# A_j) 2 \exp(-\frac{1}{8} m(\bar{\xi}^i)^2)$. Furthermore, for all $s' \geq s_0 + 1$,

$$\begin{aligned} & \mu_\sigma(\bigcap_{s'\geq 1} \bigcup_{s\geq s'} \bigcup_{q=0}^{\infty} \bigcup_{\alpha\in\mathcal{P}^i_{s+q}} \bigcup_{d=1}^{\infty} \bigcup_{m\geq \underline{m}^i_{s+q}+d-1} \mathbf{I}^{\alpha(s,q,d)}_m) \\ \leq \; & \mu_\sigma(\bigcup_{s\geq s'} \bigcup_{q=0}^{\infty} \bigcup_{\alpha\in\mathcal{P}^i_{s+q}} \bigcup_{d=1}^{\infty} \bigcup_{m\geq \underline{m}^i_{s+q}+d-1} \mathbf{I}^{\alpha(s,q,d)}_m). \end{aligned}$$

However, then, for all $s' \geq s_0 + 1$,

$$
\begin{aligned}
&\mu_\sigma(\bigcup_{s\geq s'}\bigcup_{q=0}^{\infty}\bigcup_{\alpha\in\mathcal{P}^i_{s+q}}\bigcup_{d=1}^{\infty}\bigcup_{m\geq \underline{m}^i_{s+q}+d-1}\mathbf{I}_m^{\alpha(s,q,d)}) \\
\leq\ & \sum_{s\geq s'}\sum_{q\geq 0}\sum_{\alpha\in\mathcal{P}^i_{s+q}}\sum_{d\geq 1}\sum_{m\geq \underline{m}^i_{s+q}+d-1}\mu_\sigma(\mathbf{I}_m^{\alpha(s,q,d)}) \\
\leq\ & \sum_{s\geq s'}\sum_{q\geq 0}\sum_{\alpha\in\mathcal{P}^i_{s+q}}\sum_{d\geq 1}\sum_{m\geq \underline{m}^i_{s+q}+d-1}(\sum_{j\neq i}\#A_j)2\exp(-\frac{1}{8}m(\bar{\xi}^i)^2) \\
\leq\ & 2(\sum_{j\neq i}\#A_j)\sum_{s\geq s'}\sum_{q\geq 0}\sum_{\alpha\in\mathcal{P}^i_{s+q}}\sum_{d\geq 1}\sum_{m\geq \underline{m}^i_{s+q}+d-1}\exp(-\frac{1}{8}m(\bar{\xi}^i)^2) \\
=\ & 2(\sum_{j\neq i}\#A_j)\sum_{s\geq s'}\sum_{q\geq 0}\sum_{\alpha\in\mathcal{P}^i_{s+q}}\sum_{d\geq 1}(1-\exp(-\frac{1}{8}(\bar{\xi}^i)^2))^{-1}\exp(-\frac{1}{8}(\underline{m}^i_{s+q}+d-1)(\bar{\xi}^i)^2) \\
=\ & 2(\sum_{j\neq i}\#A_j)\sum_{s\geq s'}\sum_{q\geq 0}\sum_{\alpha\in\mathcal{P}^i_{s+q}}(1-\exp(-\frac{1}{8}(\bar{\xi}^i)^2))^{-2}\exp(-\frac{1}{8}\underline{m}^i_{s+q}(\bar{\xi}^i)^2) \\
=\ & 2(\sum_{j\neq i}\#A_j)(1-\exp(-\frac{1}{8}(\bar{\xi}^i)^2))^{-1}\sum_{s\geq s'}\sum_{q\geq 0}(\#\mathcal{P}^i_{s+q})\sum_{m\geq \underline{m}^i_{s+q}}\exp(-\frac{1}{8}\underline{m}^i_{s+q}(\bar{\xi}^i)^2) \\
\leq\ & 2(\sum_{j\neq i}\#A_j)(1-\exp(-\frac{1}{8}(\bar{\xi}^i)^2))^{-1}\sum_{s\geq s'}\sum_{q\geq 0}R^i_{s+q}(\#\mathcal{P}^i_{s+q})\sum_{m\geq \underline{m}^i_{s+q}}\exp(-\frac{1}{8}m(\xi^i_{s+q})^2) \\
\leq\ & 2(\sum_{j\neq i}\#A_j)(1-\exp(-\frac{1}{8}(\bar{\xi}^i)^2))^{-1}\sum_{s\geq s'}\sum_{q\geq 0}\exp(-s-q) \\
=\ & 2(\sum_{j\neq i}\#A_j)(1-\exp(-\frac{1}{8}(\bar{\xi}^i)^2))^{-1}(1-\exp(-1))^{-2}\exp(-s').
\end{aligned}
$$

The seventh inequality holds by Condition 7. The other inequalities are obvious. Therefore, $\mu_\sigma(\bigcap_{s'\geq 1}\bigcup_{s\geq s'}\bigcup_{q=0}^{\infty}\bigcup_{\alpha\in\mathcal{P}^i_{s+q}}\bigcup_{d=1}^{\infty}\bigcup_{m\geq \underline{m}^i_{s+q}+d-1}\mathbf{G}_m^{\alpha(s,q,d)}) = 0$. Define

$$
\mathbf{I} := \bigcup_{s'\geq 1}\bigcap_{s\geq s'}\bigcap_{q=0}^{\infty}\bigcap_{\alpha\in\mathcal{P}^i_{s+q}}\bigcap_{d=1}^{\infty}\bigcap_{m\geq \underline{m}^i_{s+q}+d-1}(\mathbf{I}_m^{\alpha(s,q,d)})^c.
$$

Then, $\mu_\sigma(\mathbf{I}) = 1$.

• We say that f^i is rejected with *type I error* if f^i is rejected by some $\alpha-$test but f^i is *statistically accurate* in $\alpha-$active periods, i.e., $\|f^i_j(h) - \sigma_j(h)\| \leq \bar{\xi}^i/4$ for all $j \neq i$ in all $\alpha-$active periods (since the $\alpha-$test started) in which (enough) samples have been

collected. In addition, if f^i is rejected with type I error, we say that the rejection is of type I error. Then, for all $s = s_0 + 1, s_0 + 2, \cdots$, all $q = 0, 1, 2, \cdots$, all $\alpha \in \mathcal{P}^i_{s+q}$ and all $d = 1, 2, \cdots$, define the corresponding class $\tilde{\alpha}(s, q, d)$ such that $h_T \in \tilde{\alpha}(s, q, d)$ if and only if (1) time $T + 1$ is in epoch s (of player i), (2) $h_T \in \alpha$, (3) the $d-$th $\alpha-$test (in epoch s) is effective at time $T + 1$, (4) for some $h_\infty > h_T$, $T_0(h_\infty) \leq T$ and the $d-$th $\alpha-$test (in epoch s) obtains the *first* sample between time $T_0(h_\infty) + 1$ and time $T + 1$ in h_T, and (5) for all $h_{T_0(h_\infty)} \leq h_t \leq h_T$ such that $h_t \in \alpha$ and the $d-$th $\alpha-$test is effective at time $t + 1$,

$$\|f^i_j(h_t) - \sigma_j(h_t)\| \leq \frac{\bar{\xi}^i}{4} \text{ for all } j \neq i,$$

where temporary belief f^i has been formed just after the most recent rejection (of player i) in h_T.

Let $\mathbf{d}^{\tilde{\alpha}(s,q,d)}_{j,m}[a_j]$ denote the number of times that a_j has been realized in the first m $\tilde{\alpha}(s, q, d)-$active periods, i.e., the number of a_j in the first m samples obtained for the $d-$th $\alpha-$test against accurate belief f^i. Note that for all $h \in \tilde{\alpha}(s, q, d)$, all $j \neq i$, and all a_j, $l^\alpha_j[a_j] \leq \sigma_j(h)[a_j] \leq L^\alpha_j[a_j]$; $L^\alpha_j[a_j] - l^\alpha_j[a_j] \leq \bar{\xi}^i/4$. Then, define

$$\mathbf{J}^{\tilde{\alpha}(s,q,d)}_m(j, a_j) := \{h_\infty \mid \mathcal{T}^{\tilde{\alpha}(s,q,d)}_m < \infty, \frac{\mathbf{d}^{\tilde{\alpha}(s,q,d)}_{j,m}[a_j]}{m} > L^\alpha_j[a_j] + \frac{\bar{\xi}^i}{4} \text{ or } \frac{\mathbf{d}^{\tilde{\alpha}(s,q,d)}_{j,m}[a_j]}{m} < l^\alpha_j[a_j] - \frac{\bar{\xi}^i}{4}\}$$

and $\mathbf{J}^{\tilde{\alpha}(s,q,d)}_m := \bigcup_{j \neq i} \bigcup_{a_j} \mathbf{J}^{\tilde{\alpha}(s,q,d)}_m(j, a_j)$. Then, from Lemma 4 it follows that $\mu_\sigma(\mathbf{J}^{\tilde{\alpha}(s,q,d)}_m) \leq (\sum_{j \neq i} \#A_j) 2 \exp(-\frac{1}{8} m (\bar{\xi}^i)^2)$.

On the other hand, define $\mathbf{W} := \{h_\infty \mid$there are infinitely many rejections of type I error in $h_\infty\}$. Let $\mathbf{X}_n$ denote the event that the $n-$th rejection (of player i) occurs such that a preliminary test chooses epoch-dependent toleraton level ξ^i_s and then $\bar{\xi}^i/4-$accurate belief f^i is rejected with ξ^i_s, and let $\mathbf{Y}_n$ denote the event that the $n-$th rejection (of

player i) occurs such that a preliminary test chooses constant toleration level $\bar{\xi}^i$ and then $\bar{\xi}^i/4-$accurate belief f^i is rejected with $\bar{\xi}^i$. By the definition, $\mathbf{W}=\bigcap_{n'\geq 1}\bigcup_{n\geq n'}(\mathbf{X}_n\bigcup\mathbf{Y}_n)$.

We first show that $\mathbf{I}\bigcap\mathbf{W}\bigcap\mathbf{Z}_0\subset\bigcap_{s'\geq 1}\bigcup_{s\geq s'}\bigcup_{q=0}^{\infty}\bigcup_{\alpha\in\mathcal{P}^i_{s+q}}\bigcup_{d=1}^{\infty}\bigcup_{m\geq \underline{m}^i_{s+q}+d-1}\mathbf{J}_m^{\tilde{\alpha}(s,q,d)}$. Indeed, suppose that $h_\infty\in\mathbf{I}\bigcap\mathbf{W}\bigcap\mathbf{Z}_0$. Then, since $h_\infty\in\mathbf{W}\bigcap\mathbf{Z}_0$, it is obvious that after time $T_0(h_\infty)+1$, there are infinitely many rejections of type I error in h_∞. Furthermore, since $h_\infty\in\mathbf{I}$, there exists $\bar{s}(\geq s_0+1)$ such that for all $s\geq\bar{s}$, all q, all $\alpha\in\mathcal{P}^i_{s+q}$, all d, and all $m\geq\underline{m}^i_{s+q}+d-1$, if $\mathcal{T}_m^{\tilde{\alpha}(s,q,d)}<\infty$, then $l^\alpha_j[a_j]-\bar{\xi}^i/4\leq\mathbf{d}^{\tilde{\alpha}(s,q,d)}_{j,m}[a_j]/m\leq L^\alpha_j[a_j]+\bar{\xi}^i/4$ for all $j\neq i$ and all a_j. Then, for any $\tilde{s}\geq\bar{s}$, there exists $s\geq\tilde{s}$ such that (1) epoch s starts after time $T_0(h_\infty)+1$ and (2) type I error rejection occurs in epoch s: let the rejection be the $n-$th one. This implies that for some q, some $\alpha\in\mathcal{P}^i_{s+q}$ and some d, the $d-$th $\alpha-$test (in epoch s) rejects a $\bar{\xi}^i/4-$accurate belief, say, f^i; f^i is generated by $\mathcal{P}^i_s$. Clearly, there are two possible cases of rejecting f^i. One case is that a preliminary test chooses epoch-dependent toleration level ξ^i_s and then f^i is rejected with toleration level ξ^i_s. In this case, there exist $\alpha',\alpha''\in\mathcal{P}^i_{s+q}$ and $\beta\in\mathcal{P}^i_{s-1}$ such that $\alpha',\alpha''\subset\beta$, and $\tilde{m}^{\alpha'},\tilde{m}^{\alpha''}\geq\underline{m}^i_{s+q}+d-1$, and $\|D^i_j(\alpha')-D^i_j(\alpha'')\|>\bar{\xi}^i$ for some $j\neq i$. Since epoch s starts after time $T_0(h_\infty)+1$, $D^i_j(\alpha')=\mathbf{d}^{\alpha'}_{j,\tilde{m}^{\alpha'}}/\tilde{m}^{\alpha'}=\mathbf{d}^{\alpha'(s,q,d)}_{j,\tilde{m}^{\alpha'}}/\tilde{m}^{\alpha'}$ and $D^i_j(\alpha'')=\mathbf{d}^{\alpha''}_{j,\tilde{m}^{\alpha''}}/\tilde{m}^{\alpha''}=\mathbf{d}^{\alpha''(s,q,d)}_{j,\tilde{m}^{\alpha''}}/\tilde{m}^{\alpha''}$. However, then, σ_{-i} takes almost same (mixed) actions in all $\beta-$active periods after time $T_0(h_\infty)+1$ because $s-1\geq s_0$. From this and $h_\infty\in\mathbf{I}$, it easily derived that $\|D^i_j(\alpha')-D^i_j(\alpha'')\|=\|\mathbf{d}^{\alpha'(s,q,d)}_{j,\tilde{m}^{\alpha'}}/\tilde{m}^{\alpha'}-\mathbf{d}^{\alpha''(s,q,d)}_{j,\tilde{m}^{\alpha''}}/\tilde{m}^{\alpha''}\|\leq\bar{\xi}^i/2$ for all $j\neq i$. Thus, this case never happens in h_∞. The other case is that a preliminary test chooses constant toleration level $\bar{\xi}^i$ and then the $d-$th $\alpha-$test rejects f^i with $\bar{\xi}^i$, which is of type I error. This implies that $m^\alpha\geq\underline{m}^i_{s+q}+d-1$ and $\|D^i_j(\alpha)-f^i_j(\alpha)\|>\bar{\xi}^i$ for some $j\neq i$.[34] However, then, because f^i is statistically accurate in the $d-$th $\alpha-$test,

[34]For each $\alpha\in\mathcal{P}^i_{s+q}$, $f^i(\alpha):=f^i(h)$ for $h\in\alpha$. Since f^i is generated by $\mathcal{P}^i_s$, $f^i(\alpha)$ is well-defined.

$\|f^i_j(h) - \sigma_j(h)\| \leq \bar{\xi}^i/4$ for all $j \neq i$ in any $\alpha-$active period (during the $d-$th $\alpha-$test is effective), so that $D^i_j(\alpha) = \mathbf{d}^{\alpha}_{j,m^\alpha}/m^\alpha = \mathbf{d}^{\tilde{\alpha}(s,q,d)}_{j,m^\alpha}/m^\alpha$ for all $j \neq i$. These imply that, for some $j \neq i$ and some a_j, either $\mathbf{d}^{\tilde{\alpha}(s,q,d)}_{j,m^\alpha}[a_j]/m^\alpha = D^i_j(\alpha)[a_j] > L^\alpha_j[a_j] + \bar{\xi}^i/2$, or $\mathbf{d}^{\tilde{\alpha}(s,q,d)}_{j,m^\alpha}[a_j]/m^\alpha = D^i_j(\alpha)[a_j] < l^\alpha_j[a_j] - \bar{\xi}^i/2$. Thus, $h_\infty \in \mathbf{J}^{\tilde{\alpha}(s,q,d)}_{m^\alpha}$. Hence, $h_\infty \in \bigcap_{s'\geq 1}\bigcup_{s\geq s'}\bigcup_{q=0}^{\infty}\bigcup_{\alpha\in\mathcal{P}^i_{s+q}}\bigcup_{d=1}^{\infty}\bigcup_{m\geq \underline{m}^i_{s+q}+d-1}\mathbf{J}^{\tilde{\alpha}(s,q,d)}_m$.

Next, we need to show that $\mu_\sigma(\bigcap_{s'\geq 1}\bigcup_{s\geq s'}\bigcup_{q=0}^{\infty}\bigcup_{\alpha\in\mathcal{P}^i_{s+q}}\bigcup_{d=1}^{\infty}\bigcup_{m\geq \underline{m}^i_{s+q}+d-1}\mathbf{J}^{\tilde{\alpha}(s,q,d)}_m) = 0$. However, the proof is quite the same as in the case of $\mathbf{I}^{\alpha(s,q,d)}_m$. Therefore, $\mu_\sigma(\mathbf{I}\bigcap\mathbf{W}\bigcap\mathbf{Z}_0) = 0$. Since $\mu_\sigma(\mathbf{I}\bigcap\mathbf{Z}_0) = 1$, it means that $\mu_\sigma(\mathbf{W}) = 0$. Thus, we have shown the following lemma.

Lemma C.2 *With $\mu_\sigma-$probability one, there are at most finite test rejections of type I error.*

Lemma C.1 shows that with probability one, if there are infinitely many rejections, then a $\frac{1}{4}\bar{\xi}^i-$accurate belief is chosen infinitely many times. This implies that if there are infinitely many rejections, there are infinitely many rejections of type I error. However, then, Lemma C.2 shows that *with probability one*, there are at most *finite* rejections of type I error. Therefore, these imply that with probability one, there are at most *finite* rejections. In other words, we have obtained the following lemma.

Lemma C.3 *With $\mu_\sigma-$probability one, there is no rejection from some period on.*

• Finally, we prove Proposition 2. Let $\mathbf{S} := \{h_\infty \mid \tilde{\rho}^i_*$ does *not* $2\bar{\xi}^i-$learns to predict σ_{-i} with σ_i in $h_\infty\}$. It suffices to show that $\mu_\sigma(\mathbf{S}) = 0$ because $\epsilon > 2\bar{\xi}^i$ by Condition 8. Let $\mathbf{U} := \{h_\infty \mid$there are at most finite rejections along $h_\infty\}$; by Lemma C.3, $\mu_\sigma(\mathbf{U}) = 1$. Furthermore, for all s, q such that $s + q \geq s_0$, all $\alpha \in \mathcal{P}^i_{s+q}$, and all $d = 1, 2, \cdots$, consider class $\alpha(s, q, d)$ as defined above. Let $\mathbf{d}^{\alpha(s,q,d)}_{j,m}[a_j]$ be as defined above, i.e., the number of

times that a_j has been realized in the first m $\alpha(s,q,d)-$active periods. Then, define

$$\mathbf{K}_m^{\alpha(s,q,d)}(j,a_j) := \{h_\infty \mid \mathcal{T}_m^{\alpha(s,q,d)} < \infty, \frac{\mathbf{d}_{j,m}^{\alpha(s,q,d)}[a_j]}{m} > L_j^\alpha[a_j] + \frac{\bar{\xi}^i}{2} \text{ or } \frac{\mathbf{d}_{j,m}^{\alpha(s,q,d)}[a_j]}{m} < l_j^\alpha[a_j] - \frac{\bar{\xi}^i}{2}\}$$

and $\mathbf{K}_m^{\alpha(s,q,d)} := \bigcup_{j\neq i}\bigcup_{a_j} \mathbf{K}_m^{\alpha(s,q,d)}(j,a_j)$. Letting $q_0(s) := \max[0, s_0 - s]$, we first show that

$$\mathbf{S}\bigcap\mathbf{U}\bigcap\mathbf{Z}_0 \subset \bigcup_{s\geq 1}\bigcup_{\bar{q}\geq q_0(s)}\bigcap_{q\geq\bar{q}}\bigcup_{\alpha\in\mathcal{P}^i_{s+q}}\bigcup_{d\geq 1}\bigcup_{m\geq \underline{m}^i_{s+q}+d-1} \mathbf{K}_m^{\alpha(s,q,d)}.$$

Suppose that $h_\infty \in \mathbf{S}\bigcap\mathbf{U}\bigcap\mathbf{Z}_0$. Then, since $h_\infty \in \mathbf{U}$, there exists a (temporary) belief f^i such that player i keeps f^i *forever* from some period, say, time $\tilde{T}$; he also keeps being in the same epoch, say, epoch s^i, forever from time $\tilde{T}$. Then, player i uses either toleration level $\xi^i_{s^i}$, or $\bar{\xi}^i$ for all tests after the last rejection (in h_∞). On the other hand, since $h_\infty \in \mathbf{S}\bigcap\mathbf{Z}_0$, if $s^i \geq s_0$, then there exist $\eta > 0$, $\beta \in \mathcal{P}^i_{s^i}$, $j \neq i$, and $\hat{T} \geq \tilde{T}$ such that $\|f^i_j(\beta) - \sigma_j(h_T)\| \geq 2\bar{\xi}^i + \eta$ for *infinitely many* $T \geq \hat{T}$ such that $h_T \in \beta$ (and $h_T < h_\infty$); otherwise, i.e., $s_0 > s^i$, replace $\beta \in \mathcal{P}^i_{s^i}$ by $\beta \in \mathcal{P}^i_{s_0}$ in the previous sentence. It implies that letting $\check{T} := \max[\hat{T}, T_0(h_\infty)]$, $\|f^i_j(\beta) - \sigma_j(h_T)\| \geq \frac{7}{4}\bar{\xi}^i + \eta$ for *all* $T \geq \check{T}$ such that $h_T \in \beta$ and $h_T < h_\infty$. However, then, since f^i is employed forever in h_∞, there exists $\bar{q} \geq q_0(s^i) = \max[0, s_0 - s^i]$ such that, for all $q \geq \bar{q}$, there exist $\alpha \in \mathcal{P}^i_{s^i+q}$ and $d \geq 1$ with $\alpha \subset \beta$ such that the $d-$th $\alpha-$test is passed against $f^i(\alpha)(= f^i(\beta))$.[35] That is, $\|f^i_j(\alpha) - D^i_j(\alpha)\| \leq (\xi^i_s \leq)\bar{\xi}^i$, where $D^i_j(\alpha)(= \mathbf{d}^{\alpha(s^i,q,d)}_{j,m^\alpha}/m^\alpha)$ is based on enough samples, i.e., $m^\alpha \geq \underline{m}^i_{s^i+q} + d - 1$. From these it follows that $\|\mathbf{d}^{\alpha(s^i,q,d)}_{j,m^\alpha}/m^\alpha - \sigma_j(h_T)\| \geq \frac{3}{4}\bar{\xi}^i + \eta$ for all $T \geq \check{T}$ such that $h_T \in \beta$ and $h_T < h_\infty$. Recall that, for all $T \geq \check{T}$ such that $h_T \in \beta$ and $h_T < h_\infty$, $l^\beta_j[a_j] \leq \sigma_j(h_T)[a_j] \leq L^\beta_j[a_j]$ for all a_j and that, for

[35]Since $\mathcal{P}^i_{s^i} \leq \mathcal{P}^i_{s^i+q}$ and f^i is generated by $\mathcal{P}^i_{s^i}$, for all $\alpha \in \mathcal{P}^i_{s^i+q}$, $f^i(\alpha)$ is well-defined: $f^i(\alpha) := f^i(h)$ for $h \in \alpha$. Note also that, for all $\alpha \in \mathcal{P}^i_{s^i+q}$ there exists a *unique* $\beta \in \mathcal{P}^i_{s^i}$ such that $\beta \supset \alpha$, so that $f^i(\alpha) = f^i(\beta)$.

all a_j, $L_j^\beta[a_j] - l_j^\beta[a_j] \leq \frac{1}{4}\bar{\xi}^i$, $L_j^\beta[a_j] = L_j^\alpha[a_j]$, and $l_j^\beta[a_j] = l_j^\alpha[a_j]$. These imply that, for some a_j, either $\mathbf{d}_{j,m^\alpha}^{\alpha(s^i,q,d)}[a_j]/m^\alpha \geq L_j^\alpha[a_j] + \frac{1}{2}\bar{\xi}^i + \eta$, or $\mathbf{d}_{j,m^\alpha}^{\alpha(s^i,q,d)}[a_j]/m^\alpha \leq l_j^\alpha[a_j] - \frac{1}{2}\bar{\xi}^i - \eta$. Therefore, $h_\infty \in \bigcup_{s\geq 1}\bigcup_{\bar{q}\geq q_0(s)}\bigcap_{q\geq\bar{q}}\bigcup_{\alpha\in\mathcal{P}_{s+q}^i}\bigcup_{d\geq 1}\bigcup_{m\geq \underline{m}_{s+q}^i+d-1}\mathbf{K}_m^{\alpha(s,q,d)}$. Thus, $\mathbf{S}\bigcap\mathbf{U}\bigcap\mathbf{Z}_0 \subset \bigcup_{s\geq 1}\bigcup_{\bar{q}\geq q_0(s)}\bigcap_{q\geq\bar{q}}\bigcup_{\alpha\in\mathcal{P}_{s+q}^i}\bigcup_{d\geq 1}\bigcup_{m\geq \underline{m}_{s+q}^i+d-1}\mathbf{K}_m^{\alpha(s,q,d)}$.

Finally, from Lemma 4 and Conditions 7 and 8 it follows that, for all s, q such that $s+q \geq s_0$,

$$
\begin{aligned}
& \mu_\sigma(\bigcup_{\alpha\in\mathcal{P}_{s+q}^i}\bigcup_{d\geq 1}\bigcup_{m\geq \underline{m}_{s+q}^i+d-1}\mathbf{K}_m^{\alpha(s,q,d)}) \\
\leq\ & \#\mathcal{P}_{s+q}^i\sum_{d=1}^{\infty}\sum_{m\geq \underline{m}_{s+q}^i+d-1} 2\exp(-\frac{1}{2}m(\bar{\xi}^i)^2) \\
=\ & 2\#\mathcal{P}_{s+q}^i\sum_{d=1}^{\infty}(1-\exp(-\frac{1}{2}(\bar{\xi}^i)^2))^{-1}\exp(-\frac{1}{2}(\underline{m}_{s+q}^i+d-1)(\bar{\xi}^i)^2) \\
=\ & 2\#\mathcal{P}_{s+q}^i(1-\exp(-\frac{1}{2}(\bar{\xi}^i)^2))^{-2}\exp(-\frac{1}{2}(\underline{m}_{s+q}^i)(\bar{\xi}^i)^2) \\
=\ & 2\#\mathcal{P}_{s+q}^i(1-\exp(-\frac{1}{2}(\bar{\xi}^i)^2))^{-1}\sum_{m\geq \underline{m}_{s+q}^i}\exp(-\frac{1}{2}m(\bar{\xi}^i)^2) \\
\leq\ & 2\#\mathcal{P}_{s+q}^i(1-\exp(-\frac{1}{2}(\bar{\xi}^i)^2))^{-1}\sum_{m\geq \underline{m}_{s+q}^i}\exp(-\frac{1}{2}m(\xi_{s+q}^i)^2) \\
\leq\ & 2(1-\exp(-\frac{1}{2}(\bar{\xi}^i)^2))^{-1}R_{s+q}^i(\#\mathcal{P}_{s+q}^i)\sum_{m\geq \underline{m}_{s+q}^i}\exp(-\frac{1}{8}m(\xi_{s+q}^i)^2) \\
\leq\ & 2(1-\exp(-\frac{1}{2}(\bar{\xi}^i)^2))^{-1}\exp(-s-q).
\end{aligned}
$$

Therefore, for all $s \geq 1$ and all $\bar{q} \geq q_0(s) = \max[0, s_0 - s]$,

$$
\begin{aligned}
\mu_\sigma(\bigcap_{q\geq\bar{q}}\bigcup_{\alpha\in\mathcal{P}_{s+q}^i}\bigcup_{d\geq 1}\bigcup_{m\geq \underline{m}_s^i+d-1}\mathbf{K}_m^{\alpha(s,q,d)}) &\leq \mu_\sigma(\bigcup_{\alpha\in\mathcal{P}_{s+q}^i}\bigcup_{d\geq 1}\bigcup_{m\geq \underline{m}_s^i+d-1}\mathbf{K}_m^{\alpha(s,q,d)}) \\
&\leq 2(1-\exp(-\frac{1}{2}(\bar{\xi}^i)^2))^{-1}\exp(-s)\exp(-q) \text{ for all } q \geq \bar{q}.
\end{aligned}
$$

Thus, $\mu_\sigma(\bigcap_{q\geq\bar{q}}\bigcup_{\alpha\in\mathcal{P}_{s+q}^i}\bigcup_{d\geq 1}\bigcup_{m\geq \underline{m}_s^i+d-1}\mathbf{K}_m^{\alpha(s,q,d)}) = 0$ for all $s \geq 1$ and all $\bar{q} \geq q_0(s)$,

so that $\mu_\sigma(\bigcup_{s\geq 1}\bigcup_{\bar{q}\geq q_0(s)}\bigcap_{q\geq\bar{q}}\bigcup_{\alpha\in\mathcal{P}^i_{s+q}}\bigcup_{d\geq 1}\bigcup_{m\geq\underline{m}^i_s+d-1}\mathbf{K}^{\alpha(s,q,d)}_m)=0$. Therefore,

$$\mu_\sigma(\mathbf{S}\bigcap\mathbf{U}\bigcap\mathbf{Z}_0)\leq\mu_\sigma(\bigcup_{s\geq 1}\bigcup_{\bar{q}\geq q_0(s)}\bigcap_{q\geq\bar{q}}\bigcup_{\alpha\in\mathcal{P}^i_{s+q}}\bigcup_{d\geq 1}\bigcup_{m\geq\underline{m}^i_s+d-1}\mathbf{K}^{\alpha(s,q,d)}_m)=0.$$

Since $\mu_\sigma(\mathbf{U}\bigcap\mathbf{Z}_0)=1$, it implies that $\mu_\sigma(\mathbf{S})=0$. This completes the proof. ■

13 Appendix D

In preparation.

References

[1] Arthur, W. B. (1994): "Inductive Reasoning and Bounded Rationality," American Economic Association Papers and Proceedings, 84, 406-411.

[2] Foster, D. P., and H. P. Young (2001): "On the Impossibility of Predicting the Behavior of Rational Agents," Proceedings of the National Academy of Sciences of the USA, 98, 222, 12848-53.

[3] Foster, D. P., and H. P. Young (2003): "Learning, Hypothesis Testing, and Nash Equilibrium," Games and Economic Behavior, 45, 73-96.

[4] Fudenberg, D., and D. K. Levine (1998): The Theory of Learning in Games. Cambridge, MA: MIT Press.

[5] Gilboa, I., A. Postlewaite, and D. Schmeidler (2004): "Rationality of Belief Or: Why Bayesianism Is neither Necessary nor Sufficient for Rationality," Mimeo., Tel Aviv Univ.

[6] Kalai, E., and E. Lehrer (1993): "Rational Learning Leads to Nash Equilibrium," Econometrica, 61, 1019-1045.

[7] Kalai, E., and E. Lehrer (1994): "Weak and Strong Merging of Opinions," Journal of Mathematical Economics, 23, 73-86.

[8] Miller, R. I., and C. W. Sanchirico (1997): "Almost Everybody Disagrees Almost All Time: The Genericity of Weakly Merging Nowhere," Mimeo., Columbia Univ.

[9] Nachbar, J. H. (1997): "Prediction, Optimization, and Learning in Repeated Games," Econometrica, 65, 275-309.

[10] Nachbar, J. H. (2005): "Beliefs in Repeated Games," Econometrica, 73, 459-480.

[11] Noguchi, Y. (2005): "Merging with a Set of Probability Measures: A Characterization," Mimeo., Kanto Gakuin Univ.

[12] Sandroni, A. (2000): "Reciprocity and Cooperation in Repeated Coordination Games: The Principled-Player Approach," Games and Economic Behavior, 32, 157-182.

[13] Shiryaev, A. N. (1984): Probability, Second Edition. New York: Springer.

[14] Young, H. P. (2004): Strategic Learning and Its Limits. Oxford Univ. Press.