DISCUSSION PAPER SERIES E

Discussion Paper No. E-29

Stochastic adaptive learning with committed players in games with strict Nash equilibria

Naoki Funai

November 2023

The Institute for Economic and Business Research
Faculty of Economics
SHIGA UNIVERSITY

1-1-1 BANBA, HIKONE,
SHIGA 522-8522, JAPAN

Stochastic adaptive learning with committed players in games with strict Nash equilibria

Naoki Funai*

Faculty of Economics, Shiga University, Shiga 522-8522, Japan

October 19, 2023

Abstract

We investigate the convergence properties of an adaptive learning model that overlaps those of stochastic fictitious play learning and experience-weighted attraction learning in normal form games with strict Nash equilibria. In particular, we consider the case in which adaptive players play a game against not only other adaptive players but also committed players, who do not revise their behaviour and follow a fixed (strict Nash equilibrium or corresponding logit quantal response equilibrium) action. We then provide conditions under which the adaptive learning process, the choice probability profile of adaptive players, almost surely converges to the logit quantal response equilibrium that committed players follow. We also provide conditions under which the adaptive learning process of a more general adaptive learning model which overlaps those of payoff assessment learning and delta learning converges to a logit quantal response equilibrium different from the equilibrium that committed players follow with positive probability. Lastly, we also consider the case without committed players and provide conditions under which the adaptive learning process of the general learning model converges to each of the logit quantal response equilibria corresponding to strict Nash equilibria with positive probability.

Keywords: Adaptive learning; stochastic fictitious play learning; experience-weighted attraction learning; quantal response equilibrium; stochastic approximation; equilibrium selection

JEL classification: C72; D83

*Email addresses: naoki-funai@biwako.shiga-u.ac.jp, nfunaijp@yahoo.co.jp

1 Introduction

In the standard learning-in-games literature, we mostly consider the case in which adaptive players (e.g. incomers in a touristic city) following the same behaviour and learning rules interact only with each other and play the same game (e.g. incomers choose a side when passing by each other on a narrow street) repeatedly over periods. However, we can also consider the case in which adaptive players sometimes face other types of players (e.g. one-time visitors, local residents) who do not adjust their behaviour and commit to the same action (e.g. keeping on the right) in each period. In the experiments of finitely repeated prisoner's dilemma games and public goods games for instance, we also observe that some players seem to be not adjusting their behaviour but committing to a cooperative behaviour over periods.[1] One intriguing question is whether the existence of such committed players affects adaptive players' long-run behaviour. In particular, it is interesting to know (i) whether the existence of committed players guarantees the convergence of adaptive players' behaviour; (ii) if so, which equilibrium concept describes adaptive players' long-run behaviour; and (iii) if there exist multiple equilibria, whether adaptive players end up playing one specific equilibrium or one of any equilibria in the long run.

In this paper, to address these questions, we theoretically investigate the convergence properties of an adaptive learning model which overlaps those of stochastic fictitious play learning (Fudenberg and Kreps, 1993; SFPL hereafter) and experience-weighted attraction learning (Camerer and Ho, 1999; EWAL hereafter) with a possibility that adaptive players play a game against not only other adaptive players but also committed players, who do not revise their behaviour over periods.

In particular, we consider the following situation. In each period, from the adaptive players and committed players, a pair is randomly chosen to play a fixed finite normal form game. When facing the game, each adaptive player assigns a subjective payoff assessment, which corresponds to the expected payoff with the empirical distribution in the SFPL model and an attraction in the EWAL model, to each of her actions and chooses the action which has the highest assessment with the highest probability: she may choose the other actions with some positive probability. If an adaptive player is picked and plays the game, she observes payoff information for each of her actions, including the foregone/counterfactual payoff information, the payoff that she could have obtained if she had chosen the other actions. Using the payoff information, the adaptive player revises her assessment in such a way that the new assessment becomes a weighted average of the past payoffs. Therefore, each adaptive player chooses an action which has performed relatively better than the other actions with higher probability and adjusts the assessments towards the payoffs. Regarding the behaviour rule of committed players, we assume that each of them chooses an action according to a fixed probability distribution in each period and does not revise

[1] See Andreoni (1995) and Andreoni and Miller (1993) for instance.

her behaviour. We also assume that each adaptive player cannot identify whether her opponent is an adaptive player or a committed player, so she follows the same behaviour rule even when she interacts with a committed player.

We first consider the simplest case in which there exist only two adaptive players 1 and 2 and one committed player ν, where in each period, (i) adaptive players interact with each other with probability $p_{1,2}$ and interact with a committed player ν with probabilities $p_{1,\nu}$ and $p_{2,\nu}$, respectively, and (ii) paired players play a fixed symmetric 2×2 game. In particular, we assume that $p_{1,2} \in [0, 1]$, and thus this model corresponds to (i) the typical learning model in games when $p_{1,2} = 1$ and (ii) the model in a single-person decision problem,[2] in which each player plays the game against nature, choosing a state according to some fixed probability distribution, when $p_{1,2} = 0$. We then consider a more general case in which there may exist more than two adaptive players and committed players, a pair of whom are randomly chosen to play a fixed finite two-player normal form game in each period.

To investigate the convergence properties, we additionally assume that (i) each adaptive player follows the logit choice rule and (ii) committed players follow a logit quantal response equilibrium (McKelvey and Palfrey, 1995; LQRE hereafter) that is close to a strict Nash equilibrium. In particular, we can show that any strict Nash equilibrium can be LQRE approachable (Goeree et al., 2016), meaning that there exists a sequence of LQREs which approaches the strict Nash equilibrium as the precision parameter of the logit choice rule increases. Therefore, we can pick any strict Nash equilibrium for committed players to follow.

From the existing literature, such as the SFPL model, we know that (a) in a single-person decision problem, adaptive players learn to choose the optimal action, and (b) in the case where adaptive players do not interact with committed players, each of the Nash equilibria can arise with positive probability.[3] In this paper, we also consider the intermediate case, that is, the case in which the probabilities of adaptive players interacting with other adaptive players are between 0 and 1, and we provide a condition for the probabilities under which the equilibrium that committed players follow is uniquely chosen by adaptive players in the end. In particular, we show that if the probabilities of adaptive players facing other adaptive players are smaller than some normalised value of what players lose by unilaterally deviating from the equilibrium that committed players follow, then adaptive players learn to follow the equilibrium almost surely. We also relax assumption (ii), that is, committed players may not follow an LQRE, and provide the condition for the convergence.

As an auxiliary argument, we also focus on the result of Benaïm and Hirsch (1999) showing that the SFPL process without committed players converges to any of the strict Nash equilibria with positive probability. In this paper, we extend the result and show

[2] The case in which an adaptive player faces a single-person decision problem is also investigated in Funai (2014).

[3] We show that the same result holds in a more general adaptive learning model.

Figure 1: Symmetric 2×2 game

	s	t
s	$\pi_{s,s}, \pi_{s,s}$	$\pi_{s,t}, \pi_{t,s}$
t	$\pi_{t,s}, \pi_{s,t}$	$\pi_{t,t}, \pi_{t,t}$

that without committed players, an adaptive learning process of a more general learning model which overlaps those of SFPL, EWAL, payoff assessment learning (Cominetti et al., 2010; Funai, 2019; Leslie and Collins, 2005 and Sarin and Vahid, 1999) and delta learning (Yechiam and Busemeyer, 2005, 2006) also converges to any of the strict Nash equilibria with positive probability.

Then, by utilising the result, we show that in the general learning model, if the probabilities of adaptive players interacting with other adaptive players are greater than some normalised value of what players lose by unilaterally deviating from the equilibrium that committed players follow and shifting to another equilibrium, then adaptive players end up following the different equilibrium with positive probability.

The rest of paper is organised as follows. In Section 2, we focus on the simplest case in which there exist only two adaptive players and one committed player, a pair of whom is randomly chosen to play a fixed symmetric 2×2 game. In Section 3, we consider the case in which there may exist more than two adaptive players and committed players, a pair of whom play a finite two-player normal form game. In Section 4, we consider the case in which adaptive players may learn to follow an equilibrium which may be different from that which committed players follow. In Section 5, we provide a brief literature review. In Section 6, we conclude the argument.

2 The simplest case: two adaptive players and one committed player playing a symmetric 2×2 game

In this section, we focus on the simplest case in which there exist only two adaptive players, who are labelled 1 and 2, and a committed player, who is labelled ν, a pair of whom are randomly picked and play a fixed symmetric 2×2 game, whose payoff matrix is shown in Figure 1, repeatedly.[4] In detail, in each period, adaptive players 1 and 2 are chosen with probability $p_{1,2}$, adaptive player 1 and committed player ν are chosen with probability $p_{1,\nu}$, and adaptive player 2 and committed player ν are chosen with probability $p_{2,\nu}$, where the probabilities are fixed over periods and $p_{1,2} + p_{1,\nu} + p_{2,\nu} = 1$.

When adaptive player $i \in \{1, 2\}$ faces the committed player, the committed player chooses an action according to a probability distribution which is fixed over periods. Let $\overline{x}_u,\ u \in \{s, t\}$, be the probability that the committed player chooses action u and $\overline{x} =$

[4] General cases are discussed in Section 3.

Figure 2: Pedestrians' coordination game

	L	R
L	$1,1$	$0,0$
R	$0,0$	$1,1$

$(\overline{x}_s, \overline{x}_t)$ be the probability profile.

Note that (i) when $p_{1,2} = 1$, this model corresponds to a typical adaptive learning model in games; (ii) when $p_{1,2} = 0$, this model corresponds to the model in a single-person decision problem, in which the committed player corresponds to nature, her actions correspond to states and $\overline{x}$ corresponds to the probability distribution over states; and (iii) when $p_{1,2} \in (0,1)$, each adaptive player learns through the interaction with the other adaptive player and the committed player.

For instance, consider the case in which, on a narrow street in a city, pedestrians have to choose a side when they pass by each other. In this example, there exist two incomers, who correspond to adaptive players, who (i) interact with each other with probability $p_{1,2}$ and (ii) interact with the local resident, who corresponds to the committed player choosing the same side each time, with probabilities $p_{1,\nu}$ and $p_{2,\nu}$. In this example, during the interaction, they play the coordination game of Figure 2, in which each player chooses L (left) or R (right) and receives a payoff of 1 when they smoothly pass by each other, and a payoff of 0 otherwise.

Now, we describe the behaviour rule of adaptive players. In each period, each adaptive player assigns a subjective payoff assessment to each of her actions: let $Q_{n,i,u}$ denote player i's payoff assessment on action u in period $n \in \mathbb{N} \cup \{0\}$, where we assume that the initial assessment is bounded.[5] After playing a fixed game in each period n, for each action, each player i observes a payoff, which is denoted by $\pi_{n,i,u}$, and updates the payoff assessment using the payoff information in the following manner: for each n, i and u,

$$\begin{aligned} Q_{n+1,i,u} =& \begin{cases} Q_{n,i,u} + \lambda_{n,i}\big(\pi_{n,i,u} - Q_{n,i,s_i}\big) & \text{if } i \text{ is picked,} \\ Q_{n,i,u} & \text{otherwise,} \end{cases} \\ =& Q_{n,i,u} + \lambda_{n,i}\mathbb{1}_{n,i}\big(\pi_{n,i,u} - Q_{n,i,u}\big), \end{aligned}$$

where $\lambda_{n,i} \in [0,1]$ represents player i's weighting parameter in period n describing how much the payoff information of each action affects the next period's assessment, and $\mathbb{1}_{n,i}$ represents the indicator function such that $\mathbb{1}_{n,i} = 1$ if player i is chosen to play in period n, and 0 otherwise. Therefore, the subjective payoff assessment for each action is a weighted average of the past payoffs, and if the observed payoff of an action is greater (less) than

[5] In the SFPL model, the payoff assessment of each action corresponds to the sample average of the past (realised or foregone) payoffs of the action, and in the EWAL model, it corresponds to the action's attraction.

the current assessment of the action, the player revises and raises (lowers) the assessment. Regarding $\pi_{n,i,u}$, the payoff that adaptive player i observes for action u in period n, we assume that she observes $\pi_{u,v}$ if the opponent player chooses v: for each n, i and u, if the opponent player chooses action v,

$$\pi_{n,i,u} = \pi_{u,v}.$$

Note that since the opponent player can be the other adaptive player or the committed player, $\pi_{n,i,u}$ can be expressed as follows:

$$\pi_{n,i,u} = \mathbb{1}_{n,i,-i} \sum_{v \in \{s,t\}} \pi_{u,v} \mathbb{1}_{n,-i,v} + \mathbb{1}_{n,i,\nu} \sum_{v \in \{s,t\}} \pi_{u,v} \mathbb{1}_{n,\nu,v},$$

where $\mathbb{1}_{n,i,-i}$ and $\mathbb{1}_{n,i,\nu}$ are the indicator functions for the events that the other adaptive player $-i$ and the committed player ν, respectively, are chosen to play in period n, and $\mathbb{1}_{n,-i,v}$ and $\mathbb{1}_{n,\nu,v}$ are the indicator functions for the events that the other adaptive player and the committed player, respectively, choose action v in period n.

Note that when an adaptive player is chosen to play, she also updates the assessment of the action which is not chosen by using the payoff information which would have been obtained if she had chosen the action. Therefore, in this section, we consider the adaptive learning model with (partial) foregone/counterfactual payoff information; in other words, we focus on the belief-based part of the EWAL model.

In terms of the weighting parameter profile for each player, we assume that the profile satisfies the following condition: for each i,

$$\sum_n \lambda_{n,i} = \infty \text{ and } \sum_n (\lambda_{n,i})^2 < \infty \tag{1}$$

almost surely. This condition means that the influence of the payoff information on her behaviour diminishes over periods but does not vanish completely in later periods. Note that as opposed to the standard assumption in the SFPL model, we do not need to assume that the weighting parameter profile should be the same among players: how the payoff information affects her behaviour can be different among players. However, we assume that for each player, the weighting parameter for each of her actions should be the same. Note that when $\lambda_{n,i} = \frac{1}{n+1}$, the model corresponds to the SFPL model.

Next, we describe the choice behaviour of each adaptive player. In this paper, we assume that she follows the logit choice rule. In detail, for each n, i, u and her assessment profile $Q_{n,i} = (Q_{n,i,s}, Q_{n,i,t})$, player i's choice probability for action u in period n, which is denoted by $x_{n,\sigma,i,u}$, is defined as follows:

$$x_{n,\sigma,i,u} = \frac{e^{\sigma Q_{n,i,u}}}{\sum_{v \in \{s,t\}} e^{\sigma Q_{n,i,v}}},$$

Figure 3: Symmetric 2×2 game

	s	t
s	A, A	$0, 0$
t	$0, 0$	B, B

where σ represents player i's precision on her decision. Note that (i) the precision parameter for each player is fixed over periods; (ii) if $\sigma \to \infty$, the choice rule approaches the one which chooses the actions with the highest assessment with equal probability; and (iii) if $\sigma \to 0$, the choice rule approaches the one which chooses each of her actions with equal probability. We also let $x_{n,\sigma,i} = (x_{n,\sigma,i,s}, x_{n,\sigma,i,t})$ denote adaptive player i's choice probability profile and $x_{n,\sigma} = (x_{n,\sigma,1}, x_{n,\sigma,2})$ denote the adaptive players' choice probability profile in period n. We also assume that players' decisions are conditionally independent: the probability that adaptive players choose action profile (u, v) in period n is given by $x_{n,\sigma,1,u} \times x_{n,\sigma,2,v}$. In the following argument, we call the sequence $\{x_{n,\sigma} : n \in \mathbb{N} \cup \{0\}\}$ an adaptive learning process.

Also, we assume that each adaptive player follows the same choice and updating rules during the interaction with her opponent, whether the opponent is the other adaptive player or the committed player, in each period. One interpretation of this assumption is that the adaptive player cannot distinguish or does not pay attention to whether her opponent is the other adaptive player or a committed player. In the context of the pedestrians' coordination problem, players never ask each of the other pedestrians whether they follow some local rule or whether they are local residents, so they never learn the types of her opponents.

In this paper, we first focus on the case in which players face a game with multiple strict Nash equilibria: letting $A := \pi_{s,s} - \pi_{t,s}$ and $B := \pi_{t,t} - \pi_{s,t}$, we focus on the case in which $A, B > 0$ or $A, B < 0$. One example of this case is the symmetric game shown in Figure 3.[6]

Before providing a sufficient condition for the convergence in this case, we additionally assume that the committed player follows an LQRE with σ. That is, we assume that $(\overline{x}_s, \overline{x}_t) = (x^*_{\sigma,s}, x^*_{\sigma,t})$, where $x^*_{\sigma,u}$ denotes the equilibrium choice probability for action u and satisfies the following condition: for each $u \in \{s, t\}$,

$$x^*_{\sigma,u} = \frac{e^{\sigma(\pi_{u,s}x^*_{\sigma,s} + \pi_{u,t}x^*_{\sigma,t})}}{\sum_{v\in\{s,t\}} e^{\sigma(\pi_{v,s}x^*_{\sigma,s} + \pi_{v,t}x^*_{\sigma,t})}}.$$

In particular, for the LQRE that committed players follow, we pick the one which corresponds to strict Nash equilibrium (s, s), that is, the one which is closest to the strict Nash equilibrium; in the following section, we show that any strict Nash equilibrium is LQRE

[6] As each player observes the payoff information of unchosen actions, it is sufficient to focus on the 2×2 in Figure 3 where $A := \pi_{s,s} - \pi_{t,s}$ and $B := \pi_{t,t} - \pi_{s,t}$ to analyse the original game.

approachable (Goeree et al., 2016). Thus, in the following argument, we pick large enough σ and the LQRE such that $x^*_{\sigma,s} > \frac{1}{2}$.[7]

In the following statement, we provide a condition for the conditional probability of each adaptive player facing the other adaptive player, given that she is chosen, under which adaptive players learn to follow the LQRE that the committed player follows almost surely. Let

$$p_{\max} := \max\{\frac{p_{1,2}}{p_{1,2}+p_{1,\nu}}, \frac{p_{1,2}}{p_{1,2}+p_{2,\nu}}\}$$

be the maximum probability of each adaptive player facing another adaptive player given that she is chosen to play the game.

Proposition 1. If $p_{\max} < \frac{|A|}{|A+B|}$, then there exists $\overline{\sigma}$ such that for any $\sigma > \overline{\sigma}$, the adaptive learning process $x_{n,\sigma}$ almost surely converges to the LQRE x^*_σ that the committed player ν follows.

Proof. See Appendix A. □

It is worth noting that in Appendix A we provide a more general condition for the convergence. Note also that the condition $p_{\max} < \frac{|A|}{|A+B|}$ coincides with the one under which players do not choose t at the Bayesian Nash equilibrium when the committed player chooses s with probability one. However, players in this model require less knowledge than in the equilibrium theory: each adaptive player does not need to know the probability of facing the committed player, the (mixed) action that the committed player chooses or even the existence of the committed player.

Also, note that since we can pick any σ which is greater than $\overline{\sigma}$ and the LQRE approaches the Nash equilibrium as σ approaches infinity, the argument holds when the committed player (almost) follows a strict Nash equilibrium.

To understand the condition further, it is helpful to focus on some specific games. We first consider the pedestrians' coordination game with the payoff matrix in Figure 2. Note that $|A+B| = 2$ and $|A| = 1$. Therefore, incomers learn to follow the local rule if the conditional probability of each incomer facing the other incomer is less than that of facing the local resident: $p_{\max} < \frac{1}{2}$ if σ is large enough. We can also show that if the conditional probability of each incomer facing the the other incomer is greater than that of facing the local resident, then there is a possibility that incomers fail to follow the local rule and form a different rule.[8]

Next, consider the game of Figure 4. Note that $|A+B| = 3$, $|A| = 2$. In this case, for large enough σ, (i) if the committed player follows the LQRE corresponding to (L, L) and if $p_{\max} < \frac{2}{3}$, then the adaptive learning process almost surely converges to the equilibrium,

[7] For instance, there exist three LQREs if $\sigma > 2$ in the coordination game of Figure 2. Note that the number of equilibria depends on the precision parameter. If the precision parameter becomes small, then there exists only one LQRE. For the analysis of the case, see Funai (2019).

[8] This is shown in Section 4.1.

Figure 4: symmetric 2×2 game with risk- and payoff-dominant equilibria

	L	R
L	$2, 2$	$3, 0$
R	$0, 3$	$4, 4$

Figure 5: Prisoner's dilemma game

	s	t
s	$1, 1$	$6, 0$
t	$0, 6$	$5, 5$

and (ii) if the committed player follows the LQRE corresponding to (R, R) and if $p_{\max} < \frac{1}{3}$, the process almost surely converges to the equilibrium. Note that (L, L) risk-dominates (R, R) but (R, R) payoff-dominates (L, L). Note also that the condition for the convergence to the risk-dominant equilibrium is weaker than the one for the convergence to the payoff-dominant equilibrium: for the preservation of the risk-dominant local rule, incomers need to interact with the local resident less often than for the preservation of the payoff-dominant local rule.

We next consider the case in which there exists only one strict Nash equilibrium, the symmetric games with a strictly dominant action, in which, without loss of generality, we assume that s is the dominant action for each i: for each $v \in \{s, t\}$, $\pi_{s,v} > \pi_{t,v}$. The class of such games includes prisoner's dilemma games, whose payoff matrix is shown in Figure 5. Then, in the following statement, we show that in these games, adaptive players in the end follow the LQRE corresponding to the strictly dominant strategy equilibrium for any interaction probability profile.

Proposition 2. In symmetric 2×2 games with a strictly dominant action, for any $p_{\max} \in [0, 1]$, there exists $\overline{\sigma}$ such that for any $\sigma > \overline{\sigma}$, adaptive learning process $x_{n,\sigma}$ almost surely converges to the LQRE x^*_σ that corresponds to the strictly dominant strategy equilibrium and that the committed player ν follows.

Proof. See Appendix B. □

3 Generalisation

3.1 Finite two-player normal form games

In this section, we generalise the argument above so that (i) there may exist more than two adaptive players and (ii) the game that they play may not be symmetric and may have more than two actions. In detail, in each period $n \in \mathbb{N} \cup \{0\}$, a pair of players is randomly

chosen from the set of players denoted by $\mathcal{N} = \{1, 2, ..., N\}$ to play a fixed finite two-player normal form game. In particular, there exist two types of players, type 1 and type 2, and in each period, a type-1 player and a type-2 player are chosen to play a game $(\mathcal{T}, S, \pi)$, where (i) $\mathcal{T} = \{1, 2\}$ corresponds to the set of types, (ii) $S = S_1 \times S_2$ is the set of action profiles with S_τ being the set of actions for type-τ's players and $|S_\tau| < \infty$ for each $\tau \in \mathcal{T}$, and (iii) $\pi = (\pi_1, \pi_2) : S \to \mathbb{R}^2$ is the payoff function, where for each $\tau \in \mathcal{T}$, $\pi_\tau : S \to \mathbb{R}$ represents the payoff function of type-τ players. Therefore, for each $i \in \mathcal{N}$, if player i's type is τ, then the set of actions of player i, S_i, and the payoff function of player i, π_i, are S_τ and π_τ respectively.

Let π_{n,i,s_i} denote the payoff that player i observes for action $s_i \in S_i$ in period n and $\pi_{n,i,s_i} := \pi_i(s_i, s_{n,-i}) = \sum_{s_{-i} \in S_{-i}} \pi_i(s_i, s_{-i}) \mathbb{1}_{n,-i,s_{-i}}$, where (i) $\mathbb{1}_{n,-i,s_{-i}}$ is the indicator function such that $\mathbb{1}_{n,-i,s_{-i}} = 1$ if player i's opponent chooses s_{-i} in period n and 0 otherwise and (ii) $s_{n,-i}$ is the action that player i's opponent chooses in period n.

As in the standard argument, we extend the domain of the payoff function to the set of mixed actions: let $\pi_i(x) = \sum_{(s_i, s_{-i}) \in S} \pi_i(s_i, s_{-i}) x_{i,s_i} x_{-i,s_{-i}}$ denote the expected payoff of player i given a mixed action profile $x \in \Delta(S) := \{(x_{i,s_i}) \in [0,1]^{|S|} : \sum_{s_i \in S_i} x_{i,s_i} = 1 \text{ for each } i\}$ and let $\pi_i(s_i, x_{-i}) = \sum_{s_{-i} \in S_{-i}} \pi_i(s_i, s_{-i}) x_{-i,s_{-i}}$ denote the expected payoff of player i when player i chooses s_i with probability one and the opponent player follows mixed action x_{-i}.

For the purpose of formal analysis, we introduce the following notation. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space on which all the random variables that appear in this paper are based. Let $\mathcal{F}_n$ denote the σ-algebra which is generated by the initial payoff assessments, the information about which pair is picked and about all the choices of players up to, but not included, period n: $\mathcal{F}_n := \sigma(Q_0, \mathbb{1}_{m,i,j}, \mathbb{1}_{m,i,s_i} : m < n, i, j \in \mathcal{N}, s_i \in S)$, where Q_0 denotes the initial assessment profile and $\mathbb{1}_{m,i,j}$ denotes the indicator function such that $\mathbb{1}_{m,i,j} = 1$ if players i and j are picked in period m and 0 otherwise.[9] This generates a filtration $\{\mathcal{F}_n\}$, where $\mathcal{F}_m \subset \mathcal{F}_n$ for $m < n$.

For any i and $j \in \mathcal{N}$, we assume that the event that they are picked is independent of any other events and its probability is fixed over periods: for each n, let $p_{i,j} := \mathbb{E}[\mathbb{1}_{n,i,j} \mid \mathcal{F}_n] = \mathbb{P}(\{i \text{ and } j \text{ are picked in period } n\} \mid \mathcal{F}_n)$ denote the probability that players i and j are picked in period n. We assume that different types of players are picked in each period: for each τ, $p_{i,j} = 0$ for $i, j \in \mathcal{N}_\tau$, where $\mathcal{N}_\tau$ denotes the set of players of type τ and $\mathcal{N} = \mathcal{N}_1 \cup \mathcal{N}_2$. Letting $p_i := \sum_{j \in \mathcal{N}} p_{i,j}$ be the probability that player i is chosen to play in period n, we assume that $p_i > 0$ for each $i \in \mathcal{N}$: we only focus on the decision problems of players who actually play the game.

As in the previous section, there exist committed players, who follow a fixed (mixed) action in each period, for each type.[10] In particular, let (i) $\mathcal{N}_{\tau,l} = \{1, ..., N_{\tau,l}\}$ and $\mathcal{N}_{\tau,\nu} = \{1, ..., N_{\tau,\nu}\}$ denote the sets of adaptive players and committed players, respectively, for

[9] The initial payoff assessment profile is formally defined in the following argument.

[10] In Section 2, there exists only one committed player. Since in that section, we consider the case in which two adaptive players face a symmetric game, we omit the argument for their types and regard two

type $\tau \in \{1,2\}$ and (ii) $\mathcal{N}_l = \cup_\tau \mathcal{N}_{\tau,l}$ and $\mathcal{N}_\nu = \cup_\tau \mathcal{N}_{\tau,\nu}$ denote the set of adaptive players and committed players, respectively. We assume that each player is either an adaptive player or a committed player: $\mathcal{N}_\tau = \mathcal{N}_{\tau,l} \cup \mathcal{N}_{\tau,\nu}$ for each τ and $\mathcal{N} = \mathcal{N}_l \cup \mathcal{N}_\nu$.

Next, we introduce notation regarding the interaction probabilities between players. Let (i) $p_{(i,j)|i} := \frac{p_{i,j}}{p_i}$ denote the conditional probability of i and j being matched given that i is chosen, (ii) for a set of players $\mathcal{J}$, $p_{(i,\mathcal{J})|i} := \sum_{j\in\mathcal{J}} p_{(i,j)|i}$ denote the conditional probability that i and a player in set $\mathcal{J}$ are matched given that i is chosen, and (iii) for $j \in \mathcal{J}$, $p_{(i,j)|(i,\mathcal{J})} := \frac{p_{i,j}}{\sum_{k\in\mathcal{J}} p_{i,k}}$ denote the probability that players i and j are matched given that i is matched with a player in set $\mathcal{J}$.

Next, we describe adaptive players' behaviour. In each period, each adaptive player (i) assigns a subjective payoff assessment to each of her actions, which represents what she expects from choosing the action, and (ii) chooses an action which has the highest assessment with some randomness. For each $n \in \mathbb{N}\cup\{0\}$, $i \in \mathcal{N}_l$ and $s_i \in S_i$, let $Q_{n,i,s_i} \in \mathbb{R}$ denote the payoff assessment of player i on action s_i, let $Q_{n,i} = (Q_{n,i,s_i})_{s_i\in S_i}$ denote player i's assessment profile, and let $Q_n = (Q_{n,i})_{i\in\mathcal{N}_l}$ denote the assessment profile in period n. We assume that the initial assessment profile is bounded: there exists some $Q_{0,\max} \in \mathbb{R}$ such that $||Q_0||_\infty < Q_{0,\max}$ almost surely, where $||\cdot||_\infty$ denotes the maximum norm.

Given her payoff assessments, each adaptive player chooses an action which has the highest assessment. In particular, we consider the case in which each of the payoff assessments is subject to some random perturbation. In detail, given that noise is added to each of the assessments, each adaptive player chooses an action which has the highest perturbed assessment: for each n, i and s_i,

$$\mathbb{P}(\text{player } i \text{ chooses } s_i \text{ in period } n \mid \mathcal{F}_n) = \mathbb{P}(s_i = \arg\max_{t_i\in S_i}(Q_{n,i,t_i} + \eta_{i,t_i})),$$

where η_{i,t_i} represents the perturbation of the assessment of action t_i of player i.[11] In this paper, we assume that $(\eta_{i,t_i})_{i,t_i}$ is independent and identically distributed with the extreme value distribution, $F(\eta_{i,t_i}) = \exp(-\exp(-\sigma\eta_{i,t_i}))$, so that adaptive players follow the logit choice rule: for each n, i and s_i,

$$\begin{aligned} x_{n,\sigma,i,s_i} :=& \mathbb{P}(\text{player } i \text{ chooses } s_i \text{ in period } n \mid \mathcal{F}_n) \\ =& \frac{e^{\sigma Q_{n,i,s_i}}}{\sum_{t_i\in S_i} e^{\sigma Q_{n,i,t_i}}}, \end{aligned}$$

where x_{n,σ,i,s_i} denotes the probability that player i chooses s_i in period n given her assessments $Q_{n,i}$ and precision parameter σ. We assume that (i) the precision parameter is

committed players for both types as just one committed player.

[11] In the payoff assessment learning model, the perturbations are sometimes interpreted as emotional noise on the assessment (Sarin and Vahid, 1999), while in the SFPL model, the perturbations are interpreted as random payoffs (Fudenberg and Kreps, 1993).

the same among players and fixed over periods and (ii) players' decisions are conditionally independent: for each n, $\mathcal{J} \subset \mathcal{N}_l$ and $(s_j)_{j\in\mathcal{J}} \in \times_{j\in\mathcal{J}} S_j$,

$$\mathbb{P}((s_j)_{j\in\mathcal{J}} \text{ is chosen in period } n \mid \mathcal{F}_n) = \prod_{j\in\mathcal{J}} x_{n,\sigma,j,s_j}.$$

In the following argument, we call the sequence $\{x_{n,\sigma} = (x_{n,\sigma,i,s_i})_{i,s_i} : n \in \mathbb{N} \cup \{0\}\}$ an adaptive learning process.

Next, we describe the updating rule, the way in which adaptive players revise their payoff assessments. In each period, each adaptive player updates her assessments using the payoff information in the following manner: for each n, i and s_i,

$$\begin{aligned} Q_{n+1,i,s_i} =& \begin{cases} Q_{n,i,s_i} + \lambda_{n,i}\big(\pi_i(s_i, s_j) - Q_{n,i,s_i}\big) & \text{if } i \text{ and } j \text{ are picked and } s_j \\ & \text{is chosen in period } n, \\ Q_{n,i,s_i} & \text{if } i \text{ is not picked in period } n \end{cases} \\ =& Q_{n,i,s_i} + \lambda_{n,i} \sum_{j\in\mathcal{N}} \mathbb{1}_{n,i,j}\Big(\sum_{s_{-i}\in S_{-i}} \pi_i(s_i, s_{-i}) \mathbb{1}_{n,j,s_{-i}} - Q_{n,i,s_i}\Big). \end{aligned}$$

Note that when player i is chosen, the assessment of each of her actions is updated. It means that each player also observes what she could have received if she had chosen other actions: players observe foregone/counterfactual payoffs.[12] However, when players are not chosen, they do not obtain payoff information and do not update their payoff assessments. Therefore, each adaptive player observes payoff information for all of her actions only when she has been chosen to play the game.

Regarding the behaviour of committed players in each type, we assume that in each period, the choice probability profile of each committed player is fixed: for each n and τ, let $\overline{x}_{\tau,k,s_k}$ denote the probability of committed player $k \in \mathcal{N}_{\tau,\nu}$ of type τ choosing action $s_k \in S_\tau$, that is,

$$\mathbb{P}(\text{Committed player } k \text{ of type } \tau \text{ chooses action } s_k \text{ in period } n \mid \mathcal{F}_n) = \overline{x}_{\tau,k,s_k}.$$

Also, we assume that each committed player chooses an action independently: for $\mathcal{J} \subset \mathcal{N}_l$, $\mathcal{K} \subset \mathcal{N}_\nu$, $(s_j)_{j\in\mathcal{J}} \in \times_{j\in\mathcal{J}} S_j$, $(s_k)_{k\in\mathcal{K}} \in \times_{k\in\mathcal{K}} S_k$ and $s = (s_j, s_k) \in \times_{j\in\mathcal{J}} S_j \times \times_{k\in\mathcal{K}} S_k$,

$$\mathbb{P}(s \text{ is chosen in period } n \mid \mathcal{F}_n) = \prod_{j\in\mathcal{J}} x_{n,\sigma,j,s_j} \times \prod_{k\in\mathcal{K}} \overline{x}_{\tau,k,s_k}.$$

In particular, we assume that committed players follow an LQRE action of their own type, where the LQRE is close to $s^* = (s_1^*, s_2^*)$. First, we provide the definition of LQRE for a more general game $(\mathcal{I}, S, \pi)$, where (i) the set of players, $\mathcal{I} := \{1, 2, ..., I\}$, may consist

[12] If players know the payoff function, then the assumption means that each player imagines what she could have obtained if she had chosen the other actions.

of more than two players; (ii) $S = \times_{i \in \mathcal{I}} S_i$ denotes the finite set of action profiles; and (iii) the payoff function is $\pi : S \to \mathbb{R}^I$, where $\pi_i : S \to \mathbb{R}$ denotes the i-th component and corresponds to player i's payoff function.

Definition 1. A logit quantal response equilibrium of the normal form game $(\mathcal{I}, S, \pi)$ is a mixed action profile $x^*_\sigma = (x^*_{\sigma,i,s_i})_{i,s_i}$ such that for each $i \in \mathcal{I}$ and $s_i \in S_i$,

$$x^*_{\sigma,i,s_i} = \frac{e^{\sigma \pi_i(s_i, x^*_{\sigma,-i})}}{\sum_{t_i \in S_i} e^{\sigma \pi_i(t_i, x^*_{\sigma,-i})}},$$

where for each $t_i \in S_i$, $\pi_i(t_i, x^*_{\sigma,-i}) := \sum_{s_{-i} \in \times_{j \neq i} S_j} \pi_i(t_i, s_{-i}) \prod_{j \neq i} x^*_{\sigma,j,s_j}$.

Then we assume that for each $i \in \mathcal{N}_\tau$, $\sum_{k \in \mathcal{N}_{-\tau,\nu}} p_{(i,k)|(i,\mathcal{N}_{-\tau,\nu})} \overline{x}_{-\tau,k,s_{-\tau}} = x^*_{\sigma,-\tau,s_{-\tau}}$ for each $s_{-\tau} \in S_{-\tau}$, where $x^*_\sigma = (x^*_{\sigma,1}, x^*_{\sigma,2})$ is an LQRE which approaches s^* as $\sigma \to \infty$, $x^*_{\sigma,\tau} = (x^*_{\sigma,\tau,s_\tau})$ is the equilibrium choice probability profile for type τ, and x^*_{σ,τ,s_τ} is the choice probability that the equilibrium assigns to action $s_\tau \in S_\tau$. This means that in the population level, committed players follow an LQRE. If each committed player follows the equilibrium, meaning that $\overline{x}_{\tau,k,s_\tau} = x^*_{\sigma,\tau,s_\tau}$ for each τ, k and s_τ, then the condition holds.

The strict Nash equilibrium that committed players follow, s^*, is considered as a convention (e.g. keeping on the right on a narrow street) that the existing members of a society (e.g. local residents) have already formed. In this paper, we allow some perturbation of players' behaviour so that their behaviour can be expressed as an LQRE. However, in this paper, we do not consider a big perturbation of their behaviour. We take a perturbation small enough so that there exist multiple LQREs if there exist multiple strict Nash equilibria.

We say that a Nash equilibrium of the normal form game $(\mathcal{I}, S, \pi)$ is LQRE approachable (Goeree et al. 2016) if there exists a sequence of LQREs with respect to σ such that the sequence approaches the Nash equilibrium as σ approaches infinity. To be more specific, the strict Nash equilibrium $s^* = (s^*_i)_i$ is LQRE approachable if there exists a sequence $\{\sigma_n : n \in \mathbb{N}\}$ converging to infinity and a corresponding sequence of LQREs $\{x^*_{\sigma_n} : n \in \mathbb{N}\}$ such that for each i, $x^*_{\sigma_n,i,s^*_i} \to 1$ as $\sigma_n \to \infty$.[13]

One natural question is whether any strict Nash equilibrium is LQRE approachable, so that we can pick any strict Nash equilibrium for the one that committed players follow. The following statement answers the question. Note that in the following statement, we consider a more general game $(\mathcal{I}, S, \pi)$ in which there may exist more than two players.

Proposition 3. In any finite normal form game $(\mathcal{I}, S, \pi)$, any strict Nash equilibrium is LQRE approachable.

[13] The definition of Goeree et al. (2016), which also includes the case for a mixed Nash equilibrium, is as follows: a Nash equilibrium $\varsigma^* = (\varsigma^*_{i,s_i})$ is LQRE approachable if there exists a sequence $\{\sigma_n : n \in \mathbb{N}\}$ converging to infinity and a corresponding sequence of LQRE $\{x^*_{\sigma_n} : n \in \mathbb{N}\}$ such that for each i and s_i, $x^*_{\sigma_n,i,s_i} \to \varsigma_{i,s_i}$ as $\sigma_n \to \infty$.

Proof. See Appendix C. □

Now we consider a condition under which the adaptive learning process almost surely converges to an LQRE corresponding to a strict Nash equilibrium s^*. Here, we take σ large enough so that the expected payoff of the equilibrium action is greater than those of the other actions: we take σ such that $\pi^{d*}_{i,s_i} := \pi_i(s_i, x^*_{\sigma,-i}) - \pi_i(s^*_i, x^*_{\sigma,-i}) < 0$ for each $s_i \neq s^*_i$. Note that $\pi^{d*}_{i,s_i} \to \pi_i(s_i, s^*_{-i}) - \pi_i(s^*_i, s^*_{-i}) < 0$ as $\sigma \to \infty$, so that the condition holds if we take large enough σ. Let $p_{\max} := \max_i p_{(i,\mathcal{N}_l)|i}$ and $\mathcal{M} := \max_i(|S_i| - 1)$.

Proposition 4. If

$$p_{\max}\frac{\mathcal{M}\max_{i,s_i,s_{-i}\neq s^*_{-i}}|\pi_i(s_i,s_{-i}) - \pi_i(s^*_i,s_{-i}) - \pi_i(s_i,s^*_{-i}) + \pi_i(s^*_i,s^*_{-i})|}{\min_{i,s_i}|\pi_i(s_i,s^*_{-i}) - \pi_i(s^*_i,s^*_{-i})|} < 1,$$

there exists $\overline{\sigma}$ such that for any $\sigma > \overline{\sigma}$, adaptive learning process $x_{n,\sigma}$ almost surely converges to the LQRE x^*_σ that committed players follow.

Proof. See Appendix D. □

Roughly speaking, Proposition 4 says that if the conditional probability of each adaptive player facing another adaptive player is lower than some normalised value of what each player loses by unilaterally deviating from the equilibrium s^*, then adaptive players learn to follow what committed players follow. If (i) there exist only one adaptive player and one committed player for each type, (ii) paired players play the symmetric 2×2 game in Figure 3 and (iii) committed players follow the LQRE corresponding to s^*, then $|S_i| = 2$, $\mathcal{M} = 1$, $|\pi_i(s_i, s_{-i}) - \pi_i(s^*_i, s_{-i}) - \pi_i(s_i, s^*_{-i}) + \pi_i(s^*_i, s^*_{-i})| = |A+B|$, $\min_{i,s_i}|\pi_i(s_i, s^*_{-i}) - \pi_i(s^*_i, s^*_{-i})| = A$ and $p_{\max} = \max\{\frac{p_{1,2}}{p_{1,2}+p_{1,\nu}}, \frac{p_{1,2}}{p_{1,2}+p_{2,\nu}}\}$, and thus the condition corresponds to that of Proposition 1.

Next, we consider the case in which there exists a strictly dominant action s^*_i for each player. That is, for each i, there exists s^*_i such that for $s_i \neq s^*_i$, $\pi_i(s^*_i, s_{-i}) > \pi_i(s_i, s_{-i})$ for $s_{-i} \in S_{-i}$. Then we can first show that there exists a period after which the assessment of the dominant action is always greater than those of the other actions.

Lemma 1. With probability one, there exists N such that for $n > N$, $Q^d_{n,i,s_i} := Q_{n,i,s_i} - Q_{n,i,s^*_i} < 0$ for $s_i \neq s^*_i$.

Proof. See Appendix E. □

Given Lemma 1, we can show that for any interaction probability profile, adaptive players learn to play the LQRE corresponding to the strictly dominant strategy equilibrium.

Proposition 5. For a game with a strictly dominant strategy equilibrium, for any $p_{i,j} \in [0, 1]$, there exists $\overline{\sigma}$ such that for any $\sigma > \overline{\sigma}$, the adaptive learning process $x_{n,\sigma}$ almost surely converges to the LQRE x^*_σ that committed players follow and corresponds to the strictly dominant strategy equilibrium.

Proof. See Appendix F. □

3.2 When committed players do not follow a logit quantal response equilibrium

In the previous section, to show that the behaviour of adaptive players corresponds to that of committed players in the long run, we assumed that (i) the precision parameters for adaptive players and committed players are the same and (ii) committed players follow an LQRE. In this section, we first show that given the condition (i), if the assessment profiles and choice probability profiles converge and the limit choice probability profile coincides with the choice probability profile of committed players, then the choice probability profile of committed players should correspond to an LQRE. In the following statement, we assume that committed players of each type follow the same choice probability profile: $\overline{x}_{\tau,k,s_\tau} = \overline{x}_{\tau,l,s_\tau} =: \overline{x}_{\tau,s_\tau}$ for each $\tau \in \{1,2\}$, $k,l \in \mathcal{N}_{\tau,\nu}$, and $s_\tau \in S_\tau$.

Proposition 6. If the assessment profile almost surely converges and the limit choice probability profile of each adaptive player is the same as the choice probability profile of committed players of the same type, then the choice probability profile of committed players corresponds to an LQRE.

Proof. See Appendix G. □

Next, we relax the condition (i) and consider the case in which the precision parameters of adaptive players and committed players may not be the same. For instance, we can consider the case in which each committed player chooses the same action with probability one and adaptive players follow the logit choice rule in each period.

To analyse the case, we consider the following auxiliary payoff function: for each $i \in \mathcal{N}_\tau$ and (s_i, s_{-i}),

$$\pi'_i(s_i, s_{-i}) = p_{(i,\mathcal{N}_l)|i}\pi_i(s_i, s_{-i}) + (1 - p_{(i,\mathcal{N}_l)|i})\pi_i(s_i, \overline{x}_{-i}),$$

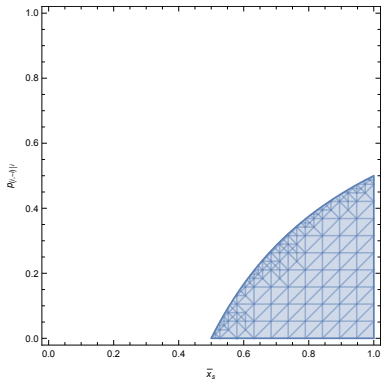
where $\overline{x}_{-i} = (\overline{x}_{-i,s_{-i}})_{s_{-i}}$ and $\overline{x}_{-i,s_{-i}} := \sum_{k \in N_{-\tau,\nu}} p_{(i,k)|(i,\mathcal{N}_{-\tau,\nu})}\overline{x}_{-\tau,k,s_{-i}}$ for each s_{-i}. That is, the payoff of adaptive player i given action profile (s_i, s_{-i}) under the auxiliary payoff function is the weighted average of (i) the payoff obtained from playing against an adaptive player and (ii) the expected payoff obtained from facing committed players. Note that in this case, we do not need to assume that $\overline{x} = x^*_\sigma$; we can obtain the convergence even when $\overline{x}_{i,s_i} = 1$ for some action s_i.

Proposition 7. If $\pi'_i(s^*_i, s^*_{-i}) > \pi'_i(s_i, s^*_{-i})$ for each i and $s_i \neq s^*_i$ and

$$\frac{\mathcal{M} \max_{i,s_i,s_{-i} \neq s^*_{-i}} |\pi'_i(s_i, s_{-i}) - \pi'_i(s^*_i, s_{-i}) - \pi'_i(s_i, s^*_{-i}) + \pi'_i(s^*_i, s^*_{-i})|}{\min_{i,s_i} |\pi'_i(s_i, s^*_{-i}) - \pi'_i(s^*_i, s^*_{-i})|} < 1,$$

there exists $\overline{\sigma}$ such that for any $\sigma > \overline{\sigma}$, adaptive learning process $x_{n,\sigma}$ almost surely converges to the LQRE x^*_σ that corresponds to strict Nash equilibrium s^* under π'.

Figure 6: A graphical expression of the condition in Proposition 7



Proof. See Appendix H. □

We now focus on the case in which (i) there exist only one adaptive player and one committed player for each type, (ii) committed players follow $\overline{x}$ and (iii) paired players play a fixed symmetric 2×2 game. Note that the payoff function π' in this case is expressed as follows: for $i \in \{1,2\}$ and $v, w \in \{s,t\}$, $\pi'_{i,v,w} := \frac{p_{i,-i}}{p_{i,-i}+p_{i,\nu}}\pi_{v,w} + \frac{p_{i,\nu}}{p_{i,-i}+p_{i,\nu}} \sum_{u\in\{s,t\}} \pi_{v,u}\overline{x}_u$. Also, let $A'_i := \pi'_{i,s,s} - \pi'_{i,t,s}$ and $B'_i := \pi'_{i,t,t} - \pi'_{i,s,t}$. In particular, we focus on the case in which $A'_i > 0$ for each i, that is, (s,s) is a strict Nash equilibrium under π'. Also, since any strict Nash equilibrium is LQRE approachable, we focus on the LQRE such that $x^*_{\sigma,i,s} > \frac{1}{2}$ for each i. Then, by Proposition 7, we know that if $A'_i > 0$ and $\frac{\max_i |A'_i+B'_i|}{\min_i |A'_i|} < 1$ for each i, then there exists $\overline{\sigma}$ such that for any $\sigma > \overline{\sigma}$, the adaptive learning process converges to the LQRE x^*_σ under π'.

To understand the conditions further, we focus on the case in which (i) players play the game in Figure 3 and (ii) $p_{(i,-i)|i} = p_{(i,-i)|-i}$ so that $A'_i = A'_{-i} =: A'$ and $B'_i = B'_{-i} =: B'$ and $\frac{\max_i |A'_i+B'_i|}{\min_i |A'_i|} = \frac{|A'+B'|}{|A'|}$. Then we express the conditions $A' > 0$ and $\frac{|A'+B'|}{|A'|} < 1$ graphically using a two-dimensional graph with x-axis representing $\overline{x}_s$ and y-axis representing $p_{(i,-i)|i}$.[14] In particular, when $A = 1$ and $B = 1$, the conditions are graphically expressed as in Figure 6.

[14] Note that

$$\begin{aligned}
A' =& Ap_{(i,-i)|i} + A(1-p_{(i,-i)|i})\overline{x}_s - B(1-p_{(i,-i)|i})(1-\overline{x}_s)\\
=& (A+B)(\overline{x}_s + p_{(i,-i)|i} - \overline{x}_s p_{(i,-i)|i}) - B,\\
B' =& Bp_{(i,-i)|i} + B(1-p_{(i,-i)|i})(1-\overline{x}_s) - A(1-p_{(i,-i)|i})\overline{x}_s\\
=& (A+B)(-\overline{x}_s + \overline{x}_s p_{(i,-i)|i}) + B \text{ and}\\
A'+B' =& (A+B)p_{(i,-i)|i}.
\end{aligned}$$

Note that (i) the dark shaded area on the lower right part of the diagram represents the profiles $(\overline{x}_s, p_{(i,-i)|i})$ which satisfy the conditions; (ii) when $\overline{x}_s < 1$, the upper limit for the probability $p_{(i,-i)|i}$ to satisfy the conditions should be less than 1/2.

Also, it is worth noting that if the precision parameters are the same among adaptive players and committed players, even for the case in which $\overline{x}_s = x^*_{\sigma,s} < 1$, the upper limit for the probability $p_{(i,-i)|i}$ to satisfy the conditions is 1/2. Therefore, if we impose the restriction that the precision parameters for adaptive players and committed players are the same, we have a better upper limit for $p_{(i,-i)|i}$ to satisfy the conditions.

4 Locking into a strict Nash equilibrium, and the possibility of convergence to any strict Nash equilibria

In this section, first, we do not consider the existence of committed players. In particular, there exist only two adaptive players facing each other to play a finite two-player game repeatedly. By focusing on this case, we may provide one possible explanation for the existence of committed players: if adaptive players have enough experience of the game and their behaviour is close to one strict Nash equilibrium, then with high probability, their behaviour converges to the equilibrium. Also, we show that in any finite period, adaptive players' behaviour can be close to any strict Nash equilibrium with positive probability, which means that any strict Nash equilibrium can be selected by the adaptive learning process.

In particular, we consider the following updating rule:

$$Q_{n+1,i,s_i} = Q_{n,i,s_i} + \lambda_n \gamma_{n,i,s_i} \big(\sum_{s_{-i}} \pi_i(s_i, s_{-i}) \mathbb{1}_{n,-i,s_{-i}} - Q_{n,i,s_i} \big),$$

where (i) $\{\lambda_n\}$ satisfies condition (1) in Section 2; (ii) γ_{n,i,s_i} is such that, almost surely, $\mathbb{E}[\gamma_{n,i,s_i} \mathbb{1}_{n,-i,s_{-i}} \mid \mathcal{F}_n] = \mathbb{E}[\gamma_{n,i,s_i} \mid \mathcal{F}_n] \mathbb{E}[\mathbb{1}_{n,-i,s_{-i}} \mid \mathcal{F}_n]$ for each n, i and s_i; and (iii) $\mathbb{E}[\mathbb{1}_{n,i,t_i} \mid \mathcal{F}_n] = \frac{\exp(\sigma Q_{n,i,t_i})}{\sum_{u_i} \exp(\sigma Q_{n,i,u_i})}$. In particular, we focus on the case in which $\gamma_{n,i,s_i} = \sum_{t_i} \gamma'_{i,s_i,t_i} \mathbb{1}_{n,i,t_i}$, where $\gamma'_{i,s_i,t_i} \in [0,1]$ and $\gamma'_{i,s_i,s_i} = 1$ for any i, s_i and t_i, and thus

$$Q_{n+1,i,s_i} = Q_{n,i,s_i} + \lambda_n (\sum_{t_i} \gamma'_{i,s_i,t_i} \mathbb{1}_{n,i,t_i}) \big(\sum_{s_{-i}} \pi_i(s_i, s_{-i}) \mathbb{1}_{n,-i,s_{-i}} - Q_{n,i,s_i} \big). \tag{2}$$

Note that (i) when $\gamma'_{i,s_i,t_i} = 1$ if $t_i = s_i$ and 0 otherwise, the learning model coincides with those of Cominetti et al.(2010), Funai (2019), Leslie and Collins (2005) and Sarin and Vahid (1999); (ii) when $\gamma'_{i,s_i,t_i} = 1$ for each s_i and t_i, the model coincides with the one in the previous sections; and (iii) when $\gamma_{i,s_i,t_i} \in (0,1)$ for $s_i \neq t_i$ and $\gamma_{i,s_i,s_i} = 1$, the model coincides with that of Yechiam and Busemeyer (2005, 2006).

Here, γ'_{i,s_i,t_i} represents how much/whether the assessment of action s_i is updated when player i chooses action t_i. In particular, the parameter can represent a situation in which the

player may obtain foregone payoff information for only some of their actions. For instance, among multiple routes to a destination, when player i chooses route A, she may obtain the traffic congestion information for routes B and C, but might not obtain the information for routes D and E as they are far from route A. In this case, we can represent the situation by assuming that $\gamma'_{i,D,A} = \gamma'_{i,E,A} = 0$. Also, γ'_{i,s_i,t_i} can represent what player i can guess for the payoff of s_i when choosing action t_i. For instance, when player i walks on the left-hand side of street A and passes by another pedestrian smoothy, she may be able to guess what could have happened if she had walked on the different side of the street; however, if player i had chosen street B instead, she might not be able to guess what could have happened if she had walked on the left-side of street A and faced a pedestrian. Therefore, γ'_{i,s_i,t_i} can represent the physical and psychological distance between the actions.[15]

For analytical purposes, we rewrite equation (2) in the following manner:

$$\begin{aligned} Q_{n+1,i,s_i} =& Q_{n,i,s_i} + \lambda_n \gamma_{n,i,s_i} \big(\sum_{s_{-i}} \pi_i(s_i, s_{-i}) \mathbb{1}_{n,-i,s_{-i}} - Q_{n,i,s_i} \big) \\ =& Q_{n,i,s_i} + \lambda_n \big(\overline{\gamma}_{n,i,s_i} (\sum_{s_{-i}} \pi_i(s_i, s_{-i}) x_{n,\sigma,-i,s_{-i}} - Q_{n,i,s_i}) + M_{n,i,s_i} \big), \end{aligned} \tag{3}$$

where

$$\begin{aligned} M_{n,i,s_i} :=& \gamma_{n,i,s_i} \big(\sum_{s_{-i}} \pi_i(s_i, s_{-i}) \mathbb{1}_{n,-i,s_{-i}} - Q_{n,i,s_i} \big) \\ & - \overline{\gamma}_{n,i,s_i} (\sum_{s_{-i}} \pi_i(s_i, s_{-i}) x_{n,\sigma,-i,s_{-i}} - Q_{n,i,s_i}). \end{aligned}$$

Then, by the stochastic approximation method, the discrete-time learning process can be approximated by the solution of the ODE $\dot{Q}_t = \frac{dQ_t}{dt} = h(Q_t)$,[16] where function $h = (h_{i,s_i}) : \mathbb{R}^{|S|} \to \mathbb{R}^{|S|}$ is defined as follows: for each i, s_i and $Q_t = (Q_{t,i,s_i})$,

$$h_{i,s_i}(Q_t) := \overline{\gamma}_{t,i,s_i} (\pi_i(s_i, x_{t,\sigma,-i}) - Q_{t,i,s_i}),$$

where $x_{t,\sigma,-i} = (x_{t,\sigma,-i,s_{-i}})$ and $x_{t,\sigma,-i,s_{-i}} := \frac{\exp(\sigma Q_{t,-i,s_{-i}})}{\sum_{t_{-i}} \exp(\sigma Q_{t,-i,t_{-i}})}$.

[15] Sarin and Vahid (2004) also consider the case in which players observe partial foregone payoffs. In particular, γ'_{i,s_i,t_i} corresponds to $f_i(s_i, t_i)$ in their model. However, the difference between their model and the model in this paper is that in their model, players use the payoff obtained by choosing an action to update the assessments of similar but different actions. While in this paper, players use the foregone payoff of each action to update its assessment.

[16] See Benaïm (1999) and Borkar (2008) for details. Note that we can show that the assumptions (A1) to (A4) in Section 2.1 of Borkar (2008) hold: (A1) holds as the logit choice rule is Lipschitz continuous and the payoff function, assessments and $\{\gamma_n\}$ are all bounded; (A2) holds as we impose the same assumption on $\{\lambda_n\}$; (A3) holds as $\{M_n\}$ is a martingale difference sequence and the payoff function, assessments and $\{\gamma_n\}$ are all bounded; (A4) holds since the initial assessment profile Q_0 is bounded and each period's assessment of each action is a convex combination of the assessment in the previous period and a payoff.

Let (s_i^*, s_{-i}^*) be a strict Nash equilibrium and x_σ^* be an LQRE which approaches (s_i^*, s_{-i}^*) as $\sigma \to \infty$. Let $\pi^* = (\pi_{i,s_i}^*)$ denote the LQRE payoff profile such that for large enough σ, $\pi_{i,s_i^*}^* > \pi_{i,s_i}^*$ for each i and $s_i \neq s_i^*$.

We now find the domain such that

$$V(Q) = ||Q - \pi^*||_\infty$$

becomes a Lyapunov function for the ODE. Let $\mathcal{Q} \subset \mathbb{R}^{|S|}$ be such that for each $Q \in \mathcal{Q}$ and i, $-K < Q_{i,s_i} - Q_{i,s_i^*} < \varepsilon$ for $s_i \neq s_i^*$ and $0 < \varepsilon < K$, where we pick ε and K such that $(\pi_i(s_i, s_i^*))_{i,s_i} \in \mathcal{Q}$ and $|Q_{n,i,s_i} - Q_{n,i,t_i}| < K$ for any n, i, $Q_{n,i}$, s_i and t_i.[17] Note also that $\mathcal{Q}$ is an open set.

Lemma 2. There exists $\overline{\sigma}$ such that for $\sigma > \overline{\sigma}$, V becomes a Lyapunov function for $\dot{Q}_t = h(Q_t)$ on $\mathcal{Q}$.

Proof. See Appendix I. □

Now we consider the case in which the assessment process converges to an LQRE payoff profile with high probability. Here, to utilise Corollary 12 of Borkar (2008), we additionally assume that (i) there exists c such that $\lambda_n \leq c\lambda_m$ for any $n \geq m$,[18] (ii) there exists a constant C such that $||h(Q)||_\infty \leq C$,[19] and (iii) there exists a Lipschitz constant L for h: for any Q and Q',

$$||h(Q) - h(Q')||_\infty \leq L||Q - Q'||_\infty.$$

Now we pick an open set B such that $\pi^* \in B \subset \overline{B} \subset \mathcal{Q}$, where $\overline{B}$ is the compact closure of B.

Proposition 8. For any B such that $\pi^* \in B \subset \overline{B} \subset \mathcal{Q}$, there exists $\overline{n}$ such that for any $n_0 \geq \overline{n}$,

$$\mathbb{P}(Q_n \to \pi^* \mid Q_{n_0} \in B) \geq 1 - \eta(b(n_0)),$$

where $b(n_0) := \sum_{n \geq n_0} (\lambda_n)^2$ and $\eta : \mathbb{R} \to (0, \infty)$ is such that $\eta(x) \to 0$ as $x \to 0$, and thus

$$\lim_{n_0 \to \infty} \mathbb{P}(Q_n \to \pi^* \mid Q_{n_0} \in B) = 1.$$

Proof. See Appendix J. □

[17] Note that since (s_i^*, s_{-i}^*) is a strict Nash equilibrium, $\pi_i(s_i, s_{-i}^*) - \pi_i(s_i^*, s_{-i}^*) < 0$ for each i and $s_i \neq s_i^*$. In addition, since we focus on a finite game, we can pick $\varepsilon > 0$ such that $\pi_i(s_i, s_{-i}^*) - \pi_i(s_i^*, s_{-i}^*) < -\varepsilon$ for each i and $s_i \neq s_i^*$. In terms of K, since each assessment is updated in a convex combination manner, for each n, i and s_i, we have $Q_{n,i,s_i} \in [\min\{Q_{0,i,s_i}, \min_{s_{-i}} \pi_i(s_i, s_{-i})\}, \max\{Q_{0,i,s_i}, \max_{s_{-i}} \pi_i(s_i, s_{-i})\}]$, and thus we pick K such that $K > |\max_{i,s_i} \max\{Q_{0,i,s_i}, \max_{s_{-i}} \pi_i(s_i, s_{-i})\} - \min_{i,s_i} \min\{Q_{0,i,s_i}, \min_{s_{-i}} \pi_i(s_i, s_{-i})\}|$.

[18] Note that for the SFPL model, $c = 1$.

[19] Such C exists due to the fact that the assessment, payoff function and γ are all bounded.

Roughly speaking, Proposition 8 says that in later periods, if the assessments are aligned so that the assessment of the strict Nash equilibrium action becomes the highest for each player, then with high probability, the process converges to the Nash equilibrium. Ianni (2014) shows a similar result on the reinforcement learning model of Roth and Erev (1995) and Erev and Roth (1998). Her result can be extended to the adaptive learning models of Cominetti et al (2006), Funai (2019), Leslie and Collins (2005), Sarin and Vahid (1999) and Yechiam and Busemeyer (2005, 2006).

Next, we show that assessments can be close to the payoff profile of any strict Nash equilibrium with positive probability. Now, pick ε' and K' such that $\pi^* = (\pi_i(s_i, x^*))_{i,s_i} \in B_{\varepsilon'} := \{Q = (Q_{i,s_i}) : -K' < Q_{i,s_i} - Q_{i,s_i^*} < -\varepsilon' \text{ for each } s_i \neq s_i^*\}$ and $\overline{B}_{\varepsilon'} = \{Q = (Q_{i,s_i}) : -K' \leq Q_{i,s_i} - Q_{i,s_i^*} \leq -\varepsilon \text{ for each } s_i \neq s_i^*\} \subset \mathcal{Q}$. Note that $B_{\varepsilon'}$ is an open set.

Lemma 3. For any $\overline{n}$, there exists $n_0 > \overline{n}$ such that $P(Q_{n_0} \in B_{\varepsilon'}) > 0$.

Proof. See Appendix K. □

Given Proposition 8 and Lemma 3, we can show that any strict Nash equilibria can be realised with positive probability by the adaptive learning process.

Proposition 9. For any strict Nash equilibrium, the probability of the adaptive learning process converging to the LQRE corresponding to the strict Nash equilibrium is positive.

Proof. It follows from Proposition 8 and Lemma 3. □

Remark: Benäim and Hirsch (1999) show the same convergence result for the SFPL process; we can show the result even when players are not required to know the payoff function, as in the adaptive learning model of Sarin and Vahid (1999), for instance.

4.1 With committed players

In this section, we extend the result above to the case considered in Section 3, in which there is a possibility that each adaptive player interacts with a committed player. In particular, we consider the case in which there exist only one adaptive player and one committed player for each type. As in the previous section, the updating rule of each assessment is given as follows:

$$Q_{n+1,i,s_i} = Q_{n,i,s_i} + \lambda_n \sum_{j \in \mathcal{N}} \mathbb{1}_{n,i,j} \big(\gamma_{n,i,s_i} (\sum_{s_{-i}} \pi_i(s_i, s_{-i}) \mathbb{1}_{n,j,s_{-i}} - Q_{n,i,s_i}) \big),$$

where $\gamma_{n,i,s_i} = \sum_{t_i} \gamma'_{i,s_i,t_i} \mathbb{1}_{n,i,t_i}$. Note that each player updates her assessments only when she is picked, and γ_{n,i,s_i} is the discount factor for the foregone/counterfactual payoff information for action s_i, where how much the information is discounted depends on which action is actually chosen and $\gamma'_{i,s_i} = (\gamma'_{i,s_i,t_i})_{t_i}$, where $\gamma'_{i,s_i,t_i} \in [0,1]$ and $\gamma'_{i,s_i,s_i} = 1$ for any

i, s_i and t_i. Note also that since we focus on the same matching structure as in Section 3, we have

$$\mathbb{E}[\mathbb{1}_{n,j,s_{-i}} \mid \mathcal{F}_n] = \begin{cases} x_{n,\sigma,j,s_{-i}} & \text{if } j \in \mathcal{N}_l, \\ \overline{x}_{-i,s_{-i}} & \text{if } j \in \mathcal{N}_\nu, \end{cases}$$

where without any confusion, $\overline{x}_{-i,s_{-i}}$ denotes the probability that the committed player whose type is opposite to player i chooses action s_{-i}.

Now recall the auxiliary payoff function π', which is defined in Section 3.2 as follows: for each (s_i, s_{-i}),

$$\pi'_i(s_i, s_{-i}) := p_{(i,-i)|i}\pi_i(s_i, s_{-i}) + (1 - p_{(i,-i)|i})\pi_i(s_i, \overline{x}_{-i}),$$

where for each i, $p_{(i,-i)|i}$ represents the conditional probability of adaptive players being matched given that adaptive player i is chosen and $\overline{x}_{-i} = (\overline{x}_{-i,s_{-i}})$.

Proposition 10. If $\pi'_i(s^*_i, s^*_{-i}) > \pi'_i(s_i, s^*_{-i})$ for each i and $s_i \neq s^*_i$, then there exists $\overline{\sigma}$ such that for any $\sigma > \overline{\sigma}$, the adaptive learning process with committed players converges to the LQRE in $(\{1,2\}, S, \pi')$ which corresponds to the strict Nash equilibrium (s^*_i, s^*_{-i}) under π' with positive probability.

Proof. See Appendix L. □

Note that in the case in which (i) players face the 2×2 game in Figure 3 with $A, B > 0$ and (ii) $\overline{x}_{i,s} = 1$ for each i, the auxiliary payoff π' is given as follows:

$$\begin{aligned} \pi'_i(s,s) =& A, \\ \pi'_i(s,t) =& A(1 - p_{(i,-i)|i}), \\ \pi'_i(t,s) =& 0 \text{ and} \\ \pi'_i(t,t) =& Bp_{(i,-i)|i}. \end{aligned}$$

Therefore, (t,t) is a strict Nash equilibrium under π' if $Bp_{(i,-i)|i} > A(1 - p_{(i,-i)|i})$ for each i, that is, if $p_{\min} := \min\{\frac{p_{1,2}}{p_{1,2}+p_{1,\nu}}, \frac{p_{1,2}}{p_{1,2}+p_{2,\nu'}}\} > \frac{A}{A+B}$ for $\nu \in \mathcal{N}_2$ and $\nu' \in \mathcal{N}_1$. In other words, given each adaptive player being chosen, if the conditional probability of her playing against the other adaptive player exceeds $\frac{A}{A+B}$, then the probability of adaptive players following the equilibrium (t,t), which is different from that which committed players follow, is positive. In the example of the pedestrians' coordination game with $p_{1,\nu} = p_{2,\nu'}$, if each adaptive player interacts with the other adaptive player more often than the committed player, then there is a chance that adaptive players in the end choose a different side from that which the local residents choose.

Next, we consider a more general case in which committed players follow strict Nash equilibrium $t^* = (t^*_1, t^*_2)$ while adaptive players may end up following strict Nash equilibrium

$s^* = (s_1^*, s_2^*)$ in a finite two-player game. By Proposition 10, this may happen if

$$\begin{aligned}\pi_i'(s_i^*, s_{-i}^*) &= p_{(i,-i)|i}\pi_i(s_i^*, s_{-i}^*) + (1 - p_{(i,-i)|i})\pi_i(s_i^*, t_{-i}^*) \\ &> p_{(i,-i)|i}\pi_i(s_i, s_{-i}^*) + (1 - p_{(i,-i)|i})\pi_i(s_i, t_{-i}^*) = \pi_i'(s_i, s_{-i}^*)\end{aligned}$$

for each i and $s_i \neq s_i^*$. In particular, this condition holds when

$$p_{(i,-i)|i} > \frac{\pi_i(t_i^*, t_{-i}^*) - \pi_i(s_i^*, t_{-i}^*)}{\pi_i(s_i^*, s_{-i}^*) - \pi_i(t_i^*, s_{-i}^*) + \pi_i(t_i^*, t_{-i}^*) - \pi_i(s_i^*, t_{-i}^*)}$$

for each i.[20] That is, if the conditional probability of each adaptive player interacting with the other adaptive player is greater than some normalised value of what each player loses by unilaterally deviating from the equilibrium that the committed players follow and shifting to a different equilibrium, then adaptive players may end up following the different equilibrium with positive probability.

Lastly, note that this result does not require players to observe foregone/counterfactual payoffs: this result holds in the models of not only SFPL and EWAL, but also payoff assessment learning (Sarin and Vahid, 1999; Cominetti et al., 2010; Funai, 2019; Leslie and Collins, 2005) and delta learning (Yechiam and Busemeyer, 2005, 2006).

[20] If we rearrange the inequality in Proposition 10, we obtain

$$\Big(\pi_i(s_i^*, s_{-i}^*) - \pi_i(s_i, s_{-i}^*) + \pi_i(s_i, t_{-i}^*) - \pi_i(s_i^*, t_{-i}^*)\Big)p_{(i,-i)|i} > \pi_i(s_i, t_{-i}^*) - \pi_i(s_i^*, t_{-i}^*)$$

for each i and $s_i \neq s_i^*$. Note that (i) $\pi_i(s_i^*, s_{-i}^*) - \pi_i(s_i, s_{-i}^*) > 0$ and (ii) if the term inside the parentheses on the left-hand side of the inequality above is equal to zero, $\pi_i(s_i^*, s_{-i}^*) - \pi_i(s_i, s_{-i}^*) + \pi_i(s_i, t_{-i}^*) - \pi_i(s_i^*, t_{-i}^*) = 0$, then $\pi_i(s_i, t_{-i}^*) - \pi_i(s_i^*, t_{-i}^*) < 0$ and the inequality above holds for any $p_{(i,-i)|i}$ without contradiction. If $\pi_i(s_i, t_{-i}^*) - \pi_i(s_i^*, t_{-i}^*) < 0$ and $\pi_i(s_i^*, s_{-i}^*) - \pi_i(s_i, s_{-i}^*) + \pi_i(s_i, t_{-i}^*) - \pi_i(s_i^*, t_{-i}^*) > 0$, then

$$p_{(i,-i)|i} \geq 0 > \frac{\pi_i(s_i, t_{-i}^*) - \pi_i(s_i^*, t_{-i}^*)}{\Big(\pi_i(s_i^*, s_{-i}^*) - \pi_i(s_i, s_{-i}^*) + \pi_i(s_i, t_{-i}^*) - \pi_i(s_i^*, t_{-i}^*)\Big)},$$

which holds for any $p_{(i,-i)|i}$ without contradiction. If $\pi_i(s_i, t_{-i}^*) - \pi_i(s_i^*, t_{-i}^*) < 0$ and $\pi_i(s_i^*, s_{-i}^*) - \pi_i(s_i, s_{-i}^*) + \pi_i(s_i, t_{-i}^*) - \pi_i(s_i^*, t_{-i}^*) < 0$, then

$$p_{(i,-i)|i} \leq 1 < \frac{\pi_i(s_i, t_{-i}^*) - \pi_i(s_i^*, t_{-i}^*)}{\Big(\pi_i(s_i^*, s_{-i}^*) - \pi_i(s_i, s_{-i}^*) + \pi_i(s_i, t_{-i}^*) - \pi_i(s_i^*, t_{-i}^*)\Big)},$$

which also holds for any $p_{(i,-i)|i}$ without contradiction. Lastly, if $\pi_i(s_i, t_{-i}^*) - \pi_i(s_i^*, t_{-i}^*) \geq 0$,

$$p_{(i,-i)|i} > \frac{\pi_i(s_i, t_{-i}^*) - \pi_i(s_i^*, t_{-i}^*)}{\Big(\pi_i(s_i^*, s_{-i}^*) - \pi_i(s_i, s_{-i}^*) + \pi_i(s_i, t_{-i}^*) - \pi_i(s_i^*, t_{-i}^*)\Big)},$$

where the right-hand side of the inequality above takes the highest value when $s_i = t_i^*$. Therefore, if the inequality above holds at $s_i = t_i^*$, then the condition in Proposition 10 also holds.

5 Literature Review

In this paper, we focus on the case in which adaptive players and committed players are randomly matched to play a finite two-player game. Utilising the SFPL model, which the model of this paper overlaps, Fudenberg and Takahashi (2011) investigate a similar situation in which players in some populations are randomly matched and play a symmetric two-player game in each period, in which they only observe their own outcomes. They also consider the case in which only some of the players are chosen to be active in each period, which causes asynchronous belief updating among players. The model in this paper shares the same motivation as theirs: we also consider the case in which players are randomly matched and update their assessments only when they are chosen. However, they do not consider the possibility that players are matched with committed players, who never revise their behaviour. Also, they investigate the impact of initial conditions on the long-run outcome via simulations; in this paper, we provide a theoretical prediction of the impact of the initial conditions for the assessments on the long-run behaviour of adaptive players.

In the game theory literature, there have already been arguments concerning the existence of players who commit to a specific action or behaviour. For instance, in the finitely repeated prisoner's dilemma, Kreps et al. (1982) theoretically investigate how the belief in the existence of players who commit to a "tit-for-tat" strategy or cooperative strategy affects the behaviour of rational players in the equilibrium model. In the experimental literature, for instance, Andreoni and Miller (1993) test the model of Kreps et al. (1982) and argue that there actually exist such committed players, and Andreoni (1995) suggests that we should incorporate such players into learning models.

Also, in the evolutionary game theory and learning-in-games literature, there have already been analyses which consider the existence of players who do not revise their behaviour. For instance, Skyrms and Pemantle (2000) simultaneously consider the evolutions of networks among players and of strategies that players choose in some games via a reinforcement learning scheme. However, they assume that as the evolution of strategies is slower than the evolution of the structure, each player always chooses a fixed strategy, but which can be different among players, during the evolution of the networks among players.[21] Although motivation for the existence of the committed players is the same, in their paper, they mainly focus on the evolution of the networks and each player does not learn which strategy to choose in the game.

We can also consider some other reasons for players committing to some behaviour. If we consider that the commitment is due to inertia, Chmura, Goerg and Selten (2012) show that on the experimental investigation of several learning models in 2×2 games, the rate of the inertia of their lab agents tends to increase over time; they suggest incorporating the inertia element into learning models. If we consider the commitment is due to some religious

[21] In some non-game-theory literature, Singh et al. (2012) for instance, the effect of the existence of committed agents on the social networks and the spread of some behaviour or opinions on some social networks are considered.

or cultural restrictions, Carvalho (2017) investigates perturbed best-response dynamics with two populations, each of which may face a restriction on its own action set, which may cause a commitment on a specific action, in coordination games.

For the stochastic stability analysis of an imitative model, Sandholm (2012) incorporates committed players instead of mutation.[22] He justifies the existence of the players in the following manner: they may not change their behaviour if they think the game that they are facing is less important than other problems or if they engage in other activities which require more cognitive capacities.[23] He shows that in two-action games where committed players focus on one action and are not negligible, the imitative players end up choosing the action that the committed players choose.[24] In this paper, however, we also consider the case in which players play the game with more than two actions available and provide conditions for the interaction probabilities under which we obtain the convergence/non-convergence to the equilibrium that committed players follow.

Block et al. (2019) also incorporate the existence of committed players in a learning model in which they assume that each action is taken by committed players to guarantee experimentation on all the actions. However, they do not provide any implication for how the degree of the interaction with the committed players affects the long-run behaviour of their learning process. Also, they assume that for each action, there is at least one player who follows the action, which we do not need to assume in this paper. Also, our motivation for introducing committed players is different from theirs: in this paper, we focus on the way in which the equilibrium that committed players follow is transmitted to adaptive players.

When we do not consider the existence of committed players, Benaïm and Hirsch (1999) show that in coordination games, the SFPL process converges to any strict Nash equilibrium with positive probability. Ianni (2014) shows the same convergence result for a different learning process, the reinforcement learning process of Roth and Erev (1995) and Erev and Roth (1998). While we also show that the convergence result is obtained in other adaptive learning models, we mainly focus on the condition under which the equilibrium that committed players follow is transmitted to adaptive players.

6 Conclusion

In this paper, we consider the case in which adaptive players interact with not only other adaptive players but also committed players, who never revise their behaviour. We first consider the case in which adaptive players in a model overlapping those of SFPL (Fuden-

[22] Heller and Mohlin (2018) also consider the existence of committed players in prisoner's dilemma games but in the equilibrium model: players do not revise their behaviour via their experience.

[23] In the evolutionary game context, Sawa and Zusai (2019) consider the situation where each player faces multiple games and investigate the evolutions of actions in the games simultaneously.

[24] Nakajima and Masuda (2015) also argue that in a similar evolutionary model, players in a fixed population always end up following what committed players play if committed players follow one specific action.

berg and Kreps, 1993) and EWAL (Camerer and Ho, 1999) learn to play the equilibrium that committed players follow. In particular, we show that if the conditional probabilities of each adaptive player interacting with other adaptive players are lower than some normalised value of what each player loses by unilaterally deviating from the equilibrium, then adaptive players learn to follow the equilibrium.

We also extend the argument to a general model overlapping those of SFPL, EWAL, payoff assessment learning (Sarin and Vahid, 1999; Cominetti et al., 2010; Funai, 2019; Leslie and Collins, 2005) and delta learning (Yechiam and Busemeyer, 2005, 2006) and consider the case in which adaptive players may not end up following the equilibrium that committed players follow. In particular, if the conditional probabilities of adaptive players interacting with other adaptive players are greater than some normalised value of what each player loses by unilaterally deviating from the equilibrium that committed players follow and shifting to a different equilibrium, then adaptive players may end up following the different equilibrium with positive probability.

Lastly, as an auxiliary result, we also show that when adaptive players of the general model interact with each other only, if their behaviour is close to an equilibrium in later periods, then their behaviour converges to the equilibrium with probability close to one.

Appendix A Proof of Proposition 1

Before providing a proof of Proposition 1, we first introduce the following notation relating to LQRE. For each $u \in \{s,t\}$, let $\pi^*_u := \pi_{u,s}x^*_{\sigma,s} + \pi_{u,t}x^*_{\sigma,t}$ and $x^*_{\sigma,u} := \frac{e^{\sigma\pi^*_u}}{e^{\sigma\pi^*_s}+e^{\sigma\pi^*_t}}$ denote the equilibrium payoff and the equilibrium choice probability, respectively, for action u. Also, let $\pi^{d*} := \pi^*_t - \pi^*_s$ denote the equilibrium payoff difference. Then, in this Appendix, we prove the following statement:

Proposition. If $p_{\max}|A+B|\frac{|x^*_{\sigma,s}|}{|\pi^{d*}|} < 1$, then adaptive learning process $x_{n,\sigma}$ almost surely converges to the LQRE $x^*_\sigma = (x^*_{\sigma,s}, x^*_{\sigma,t})$ that the committed player follows and corresponds to (s,s). In particular, if $p_{\max}\frac{|A+B|}{|A|} < 1$, then there exists $\overline{\sigma}$ such that for any $\sigma > \overline{\sigma}$, the convergence is obtained.

To utilise a stochastic approximation method, we first rewrite the updating rule of each adaptive player in the following manner: for each i and $u \in \{s,t\}$,

$$\begin{aligned} Q_{n+1,i,u} =& Q_{n,i,u} + \lambda_{n,i}\mathbb{1}_{n,i}\big(\pi_{n,i,u} - Q_{n,i,u}\big) \\ =& Q_{n,i,u} + \lambda_{n,i}(p_{i,-i}+p_{i,\nu})\big(\overline{\pi}_{n,i,u} - Q_{n,i,u} + M_{n,i,u}\big), \end{aligned}$$

where

$$\overline{\pi}_{n,i,u} := \frac{p_{i,-i}}{p_{i,-i}+p_{i,\nu}} \sum_{v\in\{s,t\}} \pi_{u,v}x_{n,\sigma,-i,v} + \frac{p_{i,\nu}}{p_{i,-i}+p_{i,\nu}} \sum_{v\in\{s,t\}} \pi_{u,v}\overline{x}_v$$

and

$$M_{n,i,u} := \frac{\mathbb{1}_{n,i}}{p_{i,-i} + p_{i,\nu}}\big(\pi_{n,i,u} - Q_{n,i,u}\big) - \big(\overline{\pi}_{n,i,u} - Q_{n,i,u}\big).$$

Note that for each i and u, $\{M_{n,i,u}\}$ is a martingale difference sequence.

Since only two actions are available for each player, it is enough for the following convergence analysis to focus on the assessment difference. In particular, let $Q^d_{n,i}$ denote the difference of player i's assessments in period n: for $s, t \in S$ and $s \neq t$,

$$Q^d_{n,i} := Q_{n,i,t} - Q_{n,i,s}.$$

Let $Q^d_n = (Q^d_{n,1}, Q^d_{n,2})$ be the assessment difference profile of adaptive players. Then each player's choice rule can be expressed as a function of the difference: let $f_\sigma : \mathbb{R} \to \mathbb{R}$ be the choice rule of each player choosing s and defined as follows: for each i and $D_i \in \mathbb{R}$,

$$f_\sigma(D_i) := \frac{1}{1 + \exp(\sigma D_i)}$$

and let $x_{n,\sigma,i,s} = f_\sigma(Q^d_{n,i})$ denote the choice probability of player i for action s in period n.

Note also that LQRE can also be described by the equilibrium payoff difference and f_σ. In particular, we have

$$\begin{aligned}
\pi^{d*} =& (\pi_{t,s} - \pi_{s,s})x^*_{\sigma,s} + (\pi_{t,t} - \pi_{s,t})x^*_{\sigma,t} = -(A+B)x^*_{\sigma,s} + B \text{ and} \\
x^*_{\sigma,s} =& \frac{1}{1 + \exp(\sigma\pi^{d*})} = f_\sigma(\pi^{d*}) = f_\sigma(-(A+B)x^*_{\sigma,s} + B),
\end{aligned}$$

recalling that $A = \pi_{s,s} - \pi_{t,s}$ and $B = \pi_{t,t} - \pi_{s,t}$.

To show the convergence, we express the updating rule of the payoff assessment differences in the following manner: for each n and i,

$$Q^d_{n+1,i} = Q^d_{n,i} + \lambda_{n,i}(p_{i,-i} + p_{i,\nu})\big(G_i(Q^d_n) - Q^d_{n,i} + M^d_{n,i}\big)$$

where

1. $M^d_{n,i} := M_{n,i,t} - M_{n,i,s}$, which is still a martingale difference noise;
2. $G = (G_i, G_{-i}) : \mathbb{R}^2 \to \mathbb{R}^2$ is defined in the following manner: for each i and $D = (D_1, D_2) \in \mathbb{R}^2$,

$$\begin{aligned}
G_i(D) :=& \big(p_{(i,-i)|i} \sum_{v\in\{s,t\}} \pi_{t,v} x_{\sigma,-i,v} + (1 - p_{(i,-i)|i}) \sum_{v\in\{s,t\}} \pi_{t,v}\overline{x}_v\big) \\
&- \big(p_{(i,-i)|i} \sum_{v\in\{s,t\}} \pi_{s,v} x_{\sigma,-i,v} + (1 - p_{(i,-i)|i}) \sum_{v\in\{s,t\}} \pi_{s,v}\overline{x}_v\big) \\
=& -(A+B)(p_{(i,-i)|i} x_{\sigma,-i,s} + (1 - p_{(i,-i)|i})\overline{x}_s) + B,
\end{aligned}$$

where

(a) $p_{(i,-i)|i} := \frac{p_{i,-i}}{p_{i,-i}+p_{i,\nu}}$ and

(b) $x_{\sigma,i,s} := f_\sigma(D_i)$, $x_{\sigma,i,t} := 1 - x_{\sigma,i,s}$ and $x_{\sigma,i} := (x_{\sigma,i,s}, x_{i,t})$ for each i.

Now to utilise the asynchronous stochastic approximation method of Tsitsiklis (1994), we consider the following difference: for each i,

$$\begin{aligned}|G_i(D) - \pi^{d*}| =& p_{(i,-i)|i}|A+B||x_{\sigma,-i,s} - x^*_{\sigma,s}| \\ =& k_{n,i}|D_{-i} - \pi^{d*}|,\end{aligned}$$

where

$$k_{n,i} := \begin{cases} p_{(i,-i)|i}|A+B|\frac{|f_\sigma(D_{-i})-f_\sigma(\pi^{d*})|}{|D_{-i}-\pi^{d*}|} & \text{if } D_{-i} \neq \pi^{d*}, \\ p_{(i,-i)|i}|A+B| \cdot |f'_{\sigma,}(\pi^{d*})| & \text{otherwise.}\end{cases}$$

Note also that in the first equality, we utilise the fact that $\overline{x}_s = x^*_{\sigma,s}$, that is, the committed player follows the LQRE x^*_σ. Then, if there is a scholar $\beta \in [0,1)$ such that for $D \in \mathbb{R}^2$ and $\boldsymbol{\pi}^{d*} := (\pi^{d*}, \pi^{d*}) \in \mathbb{R}^2$,

$$||G(D) - \boldsymbol{\pi}^{d*}||_\infty \leq \beta||D - \boldsymbol{\pi}^{d*}||_\infty,$$

then by Theorem 3 of Tsitsiklis (1994), Q^d_n almost surely converges to $\boldsymbol{\pi}^{d*}$, that is, $x_{n,\sigma}$ almost surely converges to x^*_σ. Note that (i) $||\cdot||_\infty$ denotes the maximum norm; (ii) $||(Q_{0,1}, Q_{0,2})||_\infty < Q_{0,\max}$ for some $Q_{0,\max} \in \mathbb{R}$, as we assume that the initial assessments are bounded; and (iii) payoffs, assessments and the martingale difference noises are all bounded in each period. Therefore, Assumptions 1 to 3 of Tsitsiklis (1994) are all satisfied.

Now, since f is positive and decreasing on the whole domain and concave on the non-positive domain, we know that for any $D_{-i} \in \mathbb{R}$,

$$\frac{|f_\sigma(D_{-i}) - f_\sigma(\pi^{d*})|}{|D_{-i} - \pi^{d*}|} \leq \frac{|f_\sigma(\pi^{d*})|}{|\pi^{d*}|}$$

and

$$|f'_\sigma(\pi^{d*})| \leq \frac{|f_\sigma(\pi^{d*})|}{|\pi^{d*}|}$$

for $\pi^{d*} < 0$ and $f_\sigma(\pi^{d*}) > \frac{1}{2}$.[25] Then

$$\begin{aligned}|G_i(D) - \pi^{d*}| \leq& p_{(i,-i)|i}|A+B|\frac{|f_\sigma(\pi^{d*})|}{|\pi^{d*}|}|D_{-i} - \pi^{d*}| \\ \leq& p_{\max}|A+B|\frac{|f_\sigma(\pi^{d*})|}{|\pi^{d*}|}||D - \boldsymbol{\pi}^{d*}||_\infty.\end{aligned}$$

[25]Note that (i) for $Q \leq 0$, the first inequality holds since f is a concave function on the non-positive domain; (ii) for $Q > 0$, if the first inequality does not hold ($\frac{f_\sigma(\pi^{d*})-f_\sigma(Q)}{Q-\pi^{d*}} > \frac{f_\sigma(\pi^{d*})}{-\pi^{d*}} \Rightarrow \pi^{d*}f_\sigma(Q) > f_\sigma(\pi^{d*})Q$), then $f_\sigma(\pi^{d*})$ or Q should be negative, which contradicts the assumptions and (iii) for the

Therefore, by the asynchronous stochastic approximation method of Tsitsiklis (1994), if $p_{\max}|A+B|\frac{f_\sigma(\pi^{d*})}{|\pi^{d*}|} = p_{\max}|A+B|\frac{x^*_{\sigma,s}}{|\pi^{d*}|} < 1$, we know that $x_{n,\sigma}$ almost surely converges to the LQRE x^*_σ.

Lastly, note that if $\sigma \to \infty$, $|A+B|\frac{x^*_{\sigma,s}}{|\pi^{d*}|}$ approaches $\frac{|A+B|}{|A|}$. Therefore, there exists $\overline{\sigma}$ such that for any $\sigma > \overline{\sigma}$, the condition $p_{\max}|A+B|\frac{|x^*_{\sigma,s}|}{|\pi^{d*}|} < 1$ holds if $p_{\max}\frac{|A+B|}{|A|} < 1$.[26]

Appendix B Proof of Proposition 2

We consider the case in which action s strictly dominates action t, that is, $A > 0$ and $B < 0$. Since (i) the payoff from action t is strictly lower than the one from action s for any action of her opponent and (ii) for the payoff difference, $-A$ or B is realised, we can show that for any ε, there exists N_ε such that for all $n > N_\varepsilon$, $Q^d_{n,i} \in [\min\{-A,B\} - \varepsilon, \max\{-A,B\} + \varepsilon]$ for each i.[27] In particular, we focus on ε such that $\max\{-A,B\} + \varepsilon < 0$ and $n > N_\varepsilon$. Now, since f_σ is decreasing and concave on the negative domain, for $D_{-i} \neq \pi^{d*}$,

$$\begin{aligned}\frac{|f_\sigma(D_{-i}) - f_\sigma(\pi^{d*})|}{|D_{-i} - \pi^{d*}|} =& |f'_\sigma(D_M)| \\ \leq& |f'_\sigma(\max\{-A,B\} + \varepsilon)|\end{aligned}$$

for some $D_M \in (\min\{D_{-i}, \pi^{d*}\} - \varepsilon, \max\{D_{-i}, \pi^{d*}\} + \varepsilon)$. As $\max\{-A,B\} + \varepsilon < 0$, $|f'_\sigma(\max\{-A,B\} + \varepsilon)| \to 0$ as $\sigma \to \infty$. Therefore, for any $p_{\max} \in [0,1]$, we take $\overline{\sigma}$ such that for any $\sigma > \overline{\sigma}$

$$p_{\max}|A+B||f'_\sigma(\max\{-A,B\} + \varepsilon)| < 1,$$

and thus

$$p_{(i,-i)|i}|A+B|\frac{|f_\sigma(D_{-i}) - f_\sigma(\pi^{d*})|}{|D_{-i} - \pi^{d*}|} \leq p_{\max}|A+B||f'_\sigma(\max\{-A,B\} + \varepsilon)| < 1.$$

---

second inequality, note that for $Q < 0$,

$$\begin{aligned}|f'_\sigma(Q)| =& |\frac{\sigma \exp(\sigma Q)}{(1+\exp(\sigma Q))^2}| \\ =& |f_\sigma(Q)| \cdot |\frac{\sigma}{1+\exp(-\sigma Q)}| \\ \leq& |f_\sigma(Q)| \cdot |\frac{1}{\frac{1}{\sigma} - Q}| \\ \leq& \frac{|f_\sigma(Q)|}{|Q|}.\end{aligned}$$

[26] Since $\frac{|f_\sigma(\pi^{d*})|}{|\pi^{d*}|} \to \frac{1}{|A|}$ as $\sigma \to \infty$, for any ε, there exists $\overline{\sigma}_\varepsilon$ such that for any $\sigma > \sigma_\varepsilon$, $\frac{|f_\sigma(\pi^{d*})|}{|\pi^{d*}|} < \frac{1}{|A|}(1+\varepsilon)$. Now pick ε such that $p_{\max}|A+B|\frac{1}{|A|}(1+\varepsilon) < 1$ (since $p_{\max} < \frac{|A|}{|A+B|}$, there exists $\varepsilon > 0$ such that $p_{\max}(1+\varepsilon) < \frac{|A|}{|A+B|}$).

[27] See Lemma 1 for details.

For $D_{-i} = \pi^{d*}$, we have

$$p_{(i,-i)|i}|A+B||f'_\sigma(\pi^{d*})| \leq p_{\max}|A+B||f'_\sigma(\max\{-A,B\}+\varepsilon)| < 1.$$

Therefore, as in the argument of Proposition 1, we have the convergence for any $p_{\max} \in [0,1]$.

Appendix C Proof of Proposition 3

We first characterise an LQRE by players' payoff difference profiles and the logit choice rule defined on the profiles. In detail, for each i, $D_i \in \mathbb{R}^{(|S_i|-1)}$ and s_i^*, which corresponds to player i's action at a strict Nash equilibrium (s_i^*, s_{-i}^*), let $f_\sigma = (f_{\sigma,i,s_i})_{s_i \neq s_i^*} : \mathbb{R}^{(|S_i|-1)} \to \mathbb{R}^{(|S_i|-1)}$ be such that for each i and $s_i \neq s_i^*$,

$$f_{\sigma,i,s_i}(D_i) := \frac{e^{\sigma D_{i,s_i}}}{1+\sum_{t_i \neq s_i^*} e^{\sigma D_{i,t_i}}}.$$

Then an LQRE is characterised as follows: for each i and s_i,

$$x^*_{\sigma,i,s_i} = f_{\sigma,i,s_i}(\pi_i^{d*}),$$

where (i) $\pi^{d*}_{i,s_i} = \pi_i(s_i, x^*_{\sigma,-i}) - \pi_i(s_i^*, x^*_{\sigma,-i})$ represents the equilibrium payoff difference between actions s_i and $s_i^*$and (ii) $\pi_i^{d*} = (\pi^{d*}_{i,s_i})_{s_i \neq s_i^*}$ represents player i's equilibrium payoff difference profile.

Next, let function $F_\sigma : \times_i \mathbb{R}^{(|S_i|-1)} \to \times_i \mathbb{R}^{(|S_i|-1)}$ be such that for each i, $D \in \times_i \mathbb{R}^{(|S_i|-1)}$ and $s_i \neq s_i^*$,

$$F_{\sigma,i,s_i}(D) := \sum_{s_{-i}=(s_j)_{j\neq i}} (\pi_i(s_i,s_{-i}) - \pi_i(s_i^*, s_{-i})) \prod_{j\neq i} f_{\sigma,j,s_j}(D).$$

Note that for $s_i \neq s_i^*$ and $D \in \times_i \mathbb{R}^{(|S_i|-1)}$,

$$\begin{aligned} F_{\sigma,i,s_i}(D) = &\sum_{s_{-i} \neq s^*_{-i}} (\pi_i(s_i,s_{-i}) - \pi_i(s_i^*,s_{-i}) - \pi_i(s_i,s^*_{-i}) + \pi_i(s_i^*,s^*_{-i})) \prod_{j\neq i} f_{\sigma,j,s_j}(D) \\ &+ \pi_i(s_i,s^*_{-i}) - \pi_i(s_i^*,s^*_{-i}). \end{aligned}$$

Now, we restrict the domain of F to

$$\mathcal{D} := \times_{i,s_i \neq s_i^*} [\pi_i(s_i,s^*_{-i}) - \pi_i(s_i^*,s^*_{-i}) - \varepsilon, \pi_i(s_i,s^*_{-i}) - \pi_i(s_i^*,s^*_{-i}) + \varepsilon]$$

for some $\varepsilon > 0$ such that $\pi_i(s_i,s^*_{-i}) - \pi_i(s_i^*,s^*_{-i}) + \varepsilon < 0$ for each i and s_i. Note that since (s_i^*, s^*_{-i}) is a strict Nash equilibrium, there exists such ε.

Next, consider a sequence $\{\sigma_n\}$ of precision parameters such that $\sigma_n \to \infty$ as $n \to \infty$. Since for any $D \in \mathcal{D}$, i and $s_i \neq s_i^*$, $\{f_{\sigma_n,i,s_i}(D)\}_n$ is monotonically decreasing and converging to zero and $\{f_{\sigma_n,i,s_i}\}_n$ is a sequence of continuous functions on $\mathcal{D}$, we know by Dini's theorem that f_{σ_n,i,s_i} uniformly converges to zero on $\mathcal{D}$: for any $\varepsilon' > 0$, there exists N_{ε',s_i} such that for any $n > N_{\varepsilon',s_i}$, we have $f_{\sigma_n,i,s_i}(D) < \varepsilon'$ for any $D \in \mathcal{D}$. In particular, we pick N such that for any $D \in \mathcal{D}$ and $n > N$,[28]

$$\sum_{s_{-i} \neq s_{-i}^*} |\pi_i(s_i, s_{-i}) - \pi_i(s_i, s_{-i}^*) - \pi_i(s_i^*, s_{-i}) + \pi_i(s_i^*, s_{-i}^*)| \prod_{j \neq i} f_{\sigma_n,j,s_j}(D) < \varepsilon.$$

Then for each $n > N$,

$$\begin{aligned} F_{\sigma_n,i,s_i}(D) \leq & \pi_i(s_i, s_{-i}^*) - \pi_i(s_i^*, s_{-i}^*) \\ & + \sum_{s_{-i} \neq s_{-i}^*} |\pi_i(s_i, s_{-i}) - \pi_i(s_i, s_{-i}^*) - \pi_i(s_i^*, s_{-i}) + \pi_i(s_i^*, s_{-i}^*)| \prod_{j \neq i} f_{\sigma_n,j,s_j}(D) \\ \leq & \pi_i(s_i, s_{-i}^*) - \pi_i(s_i^*, s_{-i}^*) + \varepsilon \end{aligned}$$

and

$$\begin{aligned} F_{\sigma_n,i,s_i}(D) \geq & \pi_i(s_i, s_{-i}^*) - \pi_i(s_i^*, s_{-i}^*) \\ & - \sum_{s_{-i} \neq s_{-i}^*} |\pi_i(s_i, s_{-i}) - \pi_i(s_i, s_{-i}^*) - \pi_i(s_i^*, s_{-i}) + \pi_i(s_i^*, s_{-i}^*)| \prod_{j \neq i} f_{\sigma_n,j,s_j}(D) \\ \geq & \pi_i(s_i, s_{-i}^*) - \pi_i(s_i^*, s_{-i}^*) - \varepsilon \end{aligned}$$

Therefore, F_{σ_n} is a continuous mapping on the domain $\mathcal{D}$ to itself, which is nonempty, compact and convex. Therefore, by Brouwer's fixed point theorem, there exists $D_n^* \in \mathcal{D}$ such that $F_{\sigma_n}(D_n^*) = D_n^*$. Note that for each n, D_n^* corresponds to an LQRE payoff

[28] For each j and s_j, pick N_{s_j} such that for $n > N_{s_j}$,

$$f_{\sigma_n,s_j}(D) \leq \left(\frac{\varepsilon}{\max_{i,s_i}(\sum_{s_{-i} \neq s_{-i}^*} |\pi_i(s_i, s_{-i}) - \pi_i(s_i, s_{-i}^*) - \pi_i(s_i^*, s_{-i}) + \pi_i(s_i^*, s_{-i}^*)|)}\right)^{\frac{1}{N-1}}$$

if $\max_{i,s_i}(\sum_{s_{-i} \neq s_{-i}^*} |\pi_i(s_i, s_{-i}) - \pi_i(s_i, s_{-i}^*) - \pi_i(s_i^*, s_{-i}) + \pi_i(s_i^*, s_{-i}^*)|) \neq 0$. Then $N = \max_{i,s_i} N_{s_i,\varepsilon'}$, where $\varepsilon' = \left(\frac{\varepsilon}{\max_{i,s_i}(\sum_{s_{-i} \neq s_{-i}^*} |\pi_i(s_i, s_{-i}) - \pi_i(s_i, s_{-i}^*) - \pi_i(s_i^*, s_{-i}) + \pi_i(s_i^*, s_{-i}^*)|)}\right)^{\frac{1}{N-1}}$. If $\max_{i,s_i}(\sum_{t_{-i} \neq s_{-i}^*} |\pi_i(s_i, s_{-i}) - \pi_i(s_i, s_{-i}^*) - \pi_i(s_i^*, s_{-i}) + \pi_i(s_i^*, s_{-i}^*)|) = 0$, we can pick any N.

difference profile: for each i and s_i,

$$\begin{aligned} x^*_{\sigma_n,i,s_i} :=& f_{\sigma_n,i,s_i}(D^*_{n,i}) = \frac{e^{\sigma_n D^*_{n,s_i}}}{1+\sum_{t_i\neq s_i^*} e^{\sigma_n D^*_{n,t_i}}} \\ =& \frac{e^{\sigma_n \sum_{s_{-i}} (\pi_i(s_i,s_{-i})-\pi_i(s_i^*,s_{-i}))\prod_{j\neq i} f_{\sigma_n,j,s_j}(D^*_{n,-i})}}{1+\sum_{t_i\neq s_i^*} e^{\sigma_n (\sum_{s_{-i}} (\pi_i(t_i,s_{-i})-\pi_i(s_i^*,s_{-i}))\prod_{j\neq i} f_{\sigma_n,j,s_j}(D^*_{n,-i}))}} \\ =& \frac{e^{\sigma_n \pi_i(s_i,x^*_{\sigma_n,-i})}}{\sum_{t_i} e^{\sigma_n \pi_i(t_i,x^*_{\sigma_n,-i})}}. \end{aligned}$$

Also, note that for each $n > N$, $D^*_n \in \mathcal{D}$ and thus, for $s_i \neq s_i^*$, $x^*_{\sigma_n,i,s_i} = f_{\sigma_n,i,s_i}(D^*_{n,i}) \to 0$ as $\sigma_n \to \infty$. That is, the sequence of LQREs $\{x^*_{\sigma_n}\}$ converges to the strict Nash equilibrium $s^* = (s_i^*)_i$ as $\sigma_n \to \infty$, which shows that s^* is LQRE approachable. Since we pick a strict Nash equilibrium arbitrarily, the argument above holds for any strict Nash equilibrium.

Appendix D Proof of Proposition 4

Without any confusion, let i denote both a player and her type. For analytical purposes, we first express the updating rule in the following manner: for each n, i and s_i,

$$\begin{aligned} Q_{n+1,i,s_i} =& Q_{n,i,s_i} + \lambda_{n,i} \sum_j \mathbb{1}_{n,i,j} \Big(\sum_{s_{-i}\in S_{-i}} \pi_i(s_i,s_{-i}) \mathbb{1}_{n,j,s_{-i}} - Q_{n,i,s_i} \Big) \\ =& Q_{n,i,s_i} + \alpha_{n,i} \Big(\overline{\pi}_{n,i,s_i} - Q_{n,i,s_i} + M_{n,i,s_i} \Big), \end{aligned}$$

where (i)$\alpha_{n,i} := \lambda_{n,i} p_i$, (ii) $\overline{\pi}_{n,i,s_i} := \sum_{s_{-i}\in S_{-i}} \pi_i(s_i,s_{-i}) (\sum_j p_{(i,j)|i} \mathbb{E}[\mathbb{1}_{n,j,s_{-i}} \mid \mathcal{F}_n])$ and (iii)

$$\begin{aligned} M_{n,i,s_i} :=& \frac{1}{p_i} \Big(\sum_j \mathbb{1}_{n,i,j} \Big(\sum_{s_{-i}\in S_{-i}} \pi_i(s_i,s_{-i}) \mathbb{1}_{n,j,s_{-i}} - Q_{n,i,s_i} \Big) \\ & - \mathbb{E}[\sum_j \mathbb{1}_{n,i,j} \Big(\sum_{s_{-i}\in S_{-i}} \pi_i(s_i,s_{-i}) \mathbb{1}_{n,j,s_{-i}} - Q_{n,i,s_i} \Big) \mid \mathcal{F}_n] \Big), \end{aligned}$$

which is a martingale difference noise. Note that

$$\mathbb{E}[\mathbb{1}_{n,j,s_{-i}} \mid \mathcal{F}_n] = \begin{cases} x_{n,\sigma,j,s_{-i}} & \text{for } j \in \mathcal{N}_l, \\ \overline{x}_{\sigma,j,s_{-i}} & \text{for } j \in \mathcal{N}_\nu, \end{cases}$$

and

$$\begin{aligned}\sum_j p_{(i,j)|i}\mathbb{E}[\mathbb{1}_{n,j,s_{-i}} \mid \mathcal{F}_n] &= \sum_{j\in\mathcal{N}_l} p_{(i,j)|i}x_{n,\sigma,j,s_{-i}} + \sum_{j\in\mathcal{N}_\nu} p_{(i,j)|i}\overline{x}_{\sigma,j,s_{-i}} \\ &= \sum_{j\in\mathcal{N}_l} p_{(i,j)|i}x_{n,\sigma,j,s_{-i}} + p_{(i,\mathcal{N}_\nu|i)} \sum_{j\in\mathcal{N}_\nu} p_{(i,j)|(i,\mathcal{N}_\nu)}\overline{x}_{\sigma,j,s_{-i}} \\ &= \sum_{j\in\mathcal{N}_l} p_{(i,j)|i}x_{n,\sigma,j,s_{-i}} + (1-p_{(i,\mathcal{N}_l)|i})x^*_{\sigma,-i,s_{-i}} \\ &=: x'_{n,\sigma,-i,s_{-i}}.\end{aligned}$$

In addition, we assume that for each i,

$$\sum_n \alpha_{n,i} = \infty \text{ and } \sum_n(\alpha_{n,i}) < \infty$$

with probability one, which holds when $\{\lambda_{n,i}\}$ satisfies condition (1) in Section 2.

Now, we follow the argument in Appendix A: we define the payoff assessment difference profile and utilise the stochastic approximation method of Tsitsiklis (1994). Let $Q^d_{n,i,s_i} := Q_{n,i,s_i} - Q_{n,i,s^*_i}$ denote the payoff assessment difference of actions s_i and s^*_i in period n. Note that for $s_i \neq s^*_i$,

$$\begin{aligned}Q^d_{n+1,i,s_i} =& Q^d_{n,i,s_i} + \alpha_{n,i}(\overline{\pi}_{n,i,s_i} - \overline{\pi}_{n,i,s^*_i} - Q^d_{n,i,s_i} - (M_{n,i,s_i} - M_{n,i,s^*_i})) \\ =& Q^d_{n,i,s_i} + \alpha_{n,i}(\overline{\pi}^d_{n,i,s_i} - Q^d_{n,i,s_i} + M^d_{n,i,s_i})\end{aligned}$$

where $M^d_{n,i,s_i} := M_{n,i,s_i} - M_{n,i,s^*_i}$, which is still a martingale difference noise, and

$$\begin{aligned}\overline{\pi}^d_{n,i,s_i} :=& \overline{\pi}_{n,i,s_i} - \overline{\pi}_{n,i,s^*_i} \\ =& \sum_{s_{-i}\in S_{-i}} (\pi_i(s_i,s_{-i}) - \pi_i(s^*_i,s_{-i}))x'_{n,\sigma,-i,s_{-i}} \\ =& \sum_{s_{-i}\neq s^*_{-i}} (\pi_i(s_i,s_{-i}) - \pi_i(s^*_i,s_{-i}))x'_{n,\sigma,-i,s_{-i}} + (\pi_i(s_i,s^*_{-i}) - \pi_i(s^*_i,s^*_{-i}))(1 - \sum_{s_{-i}\neq s^*_{-i}} x'_{n,\sigma,-i,s_{-i}}) \\ =& \sum_{s_{-i}\neq s^*_{-i}} (\pi_i(s_i,s_{-i}) - \pi_i(s^*_i,s_{-i}) - \pi_i(s_i,s^*_{-i}) + \pi_i(s^*_i,s^*_{-i}))x'_{n,\sigma,-i,s_{-i}} + (\pi_i(s_i,s^*_{-i}) - \pi_i(s^*_i,s^*_{-i})).\end{aligned}$$

To utilise the asynchronous stochastic approximation method, letting $G = (G_{i,s_i}) : \times_i\mathbb{R}^{|S_i|-1} \to \times_i\mathbb{R}^{|S_i|-1}$ be such that for each s_i and $D \in \times_i\mathbb{R}^{|S_i|-1}$,

$$\begin{aligned}G_{i,s_i}(D) =& \sum_{s_{-i}\neq s^*_{-i}} (\pi_i(s_i,s_{-i}) - \pi_i(s^*_i,s_{-i}) - \pi_i(s_i,s^*_{-i}) + \pi_i(s^*_i,s^*_{-i})) \\ & \times (\sum_{j\in\mathcal{N}_l} p_{(i,j)|i}f_{\sigma,j,s_{-i}}(D) - (1-p_{(i,\mathcal{N}_l)|i})x^*_{\sigma,-i,s_{-i}}) \\ & + (\pi_i(s_i,s^*_{-i}) - \pi_i(s^*_i,s^*_{-i})),\end{aligned}$$

we consider the following difference:

$$
\begin{aligned}
&|G_{i,s_i}(D) - \pi^{d*}_{i,s_i}| \\
=&|\sum_{j\in N_l} p_{(i,j)|i} \sum_{s_{-i}\neq s^*_{-i}} (\pi_i(s_i,s_{-i}) - \pi_i(s^*_i,s_{-i}) - \pi_i(s_i,s^*_{-i}) + \pi_i(s^*_i,s^*_{-i}))(x_{\sigma,j,s_{-i}} - x^*_{\sigma,-i,s_{-i}})| \\
\leq &\sum_{j\in N_l} p_{(i,j)|i} b \sum_{s_{-i}\neq s^*_{-i}} |x_{\sigma,j,s_{-i}} - x^*_{\sigma,-i,s_{-i}}|,
\end{aligned}
$$

where (i) $x_{\sigma,j,s_{-i}} = f_{\sigma,j,s_{-i}}(D_j)$, (ii) $b = \max_{i,s_i,s_{-i}\neq s^*_{-i}} |\pi_i(s_i,s_{-i}) - \pi_i(s^*_i,s_{-i}) - \pi_i(s_i,s^*_{-i}) + \pi_i(s^*_i,s^*_{-i})|$ and (iii) for the equilibrium payoff difference profile $\pi^{d*} = (\pi^{d*}_{i,s_i})$, we have

$$
\begin{aligned}
\pi^{d*}_{i,s_i} =& \sum_{s_{-i}} (\pi_i(s_i,s_{-i}) - \pi_i(s^*_i,s_{-i}))x^*_{\sigma,-i,s_{-i}} \\
=& \sum_{s_{-i}\neq s^*_{-i}} \big(\pi_i(s_i,s_{-i}) - \pi_i(s^*_i,s_{-i}) - \pi_i(s_i,s^*_{-i}) + \pi_i(s^*_i,s^*_{-i}))x^*_{\sigma,-i,s_{-i}} + (\pi_i(s_i,s^*_{-i}) - \pi_i(s^*_i,s^*_{-i})\big).
\end{aligned}
$$

Now, if there exists some $\beta \in [0,1)$ such that for any $D \in \times_i \mathbb{R}^{|S_i|-1}$,

$$||G(D) - \pi^{d*}||_\infty \leq \beta ||D - \pi^{d*}||_\infty,$$

then we can show the convergence by the method.

To obtain the inequality above, we consider the inequality between the difference $x_{\sigma,j,s_{-i}} - x^*_{\sigma,-i,s_{-i}}$ and $||D - \pi^{d*}||_\infty$. In particular, we express the difference $x_{\sigma,j,s_{-i}} - x^*_{\sigma,-i,s_{-i}}$ as a telescoping sum and consider the inequality between each term of the sum and $||Q^d - \pi^{d*}||_\infty$. To do so, given $D_j = (D_{j,s_{-i}}, D_{j,t_{-i}}, ..., D_{j,z_{-i}})$ and $\pi^{d*}_{-i} = (\pi^{d*}_{-i,s_{-i}}, \pi^{d*}_{-i,t_{-i}}, ..., \pi^{d*}_{-i,z_{-i}})$, we define the sequence $(D_{0,j}, D_{1,j}, ..., D_{|S_{-i}|-1,j})$ such that

$$
\begin{aligned}
&D_{0,j} := (D_{j,s_{-i}}, D_{j,t_{-i}}, ..., D_{j,z_{-i}}) = D_j, \\
&D_{1,j} := (\pi^{d*}_{-i,s_{-i}}, D_{j,t_{-i}}, ..., D_{j,z_{-i}}), \\
&..., \\
&D_{|S_{-i}|-2,j} := (\pi^{d*}_{-i,s_{-i}}, ..., \pi^{d*}_{-i,y_{-i}}, D_{j,z_{-i}}), \\
&D_{|S_{-i}|-1,j} := (\pi^{d*}_{-i,s_{-i}}, ..., \pi^{d*}_{-i,y_{-i}}, \pi^{d*}_{-i,z_{-i}}) = \pi^{d*}_{-i}.
\end{aligned}
$$

Then the difference $x_{\sigma,j,s_{-i}} - x^*_{\sigma,-i,s_{-i}}$ can be expressed as the following telescoping sum:

$$x_{\sigma,j,s_{-i}} - x^*_{\sigma,-i,s_{-i}} = \sum_{m=0}^{|S_{-i}|-2} (f_{\sigma,-i,s_{-i}}(D_{m,j}) - f_{\sigma,-i,s_{-i}}(D_{m+1,j})). \tag{4}$$

Next, we see an upper bound for each term of the summation in equation (4). For $m = 0$,

$$\begin{aligned}|f_{\sigma,-i,s_{-i}}(D_{0,j}) - f_{\sigma,-i,s_{-i}}(D_{1,j})| =& \frac{|f_{\sigma,-i,s_{-i}}(D_{0,j}) - f_{\sigma,-i,s_{-i}}(D_{1,j})|}{||D_{0,j} - D_{1,j}||_\infty}||D_{0,j} - D_{1,j}||_\infty \\ =& \frac{|f_{\sigma,-i,s_{-i}}(D_{0,-i}) - f_{\sigma,-i,s_{-i}}(D_{1,j})|}{|D_{j,s_{-i}} - \pi^{d*}_{-i,s_{-i}}|}|D_{j,s_{-i}} - \pi^{d*}_{-i,s_{-i}}|.\end{aligned}$$

If $D_{j,s_{-i}} \geq 0$, we have $f_{\sigma,-i,s_{-i}}(D_{0,-i}) > f_{\sigma,-i,s_{-i}}(D_{1,-i})$, $|D_{j,s_{-i}} - \pi^{d*}_{-i,s_{-i}}| \geq |\pi^{d*}_{-i,s_{-i}}|$ and $|f_{\sigma,-i,s_{-i}}(D_{0,j}) - f_{\sigma,-i,s_{-i}}(D_{1,j})| \leq |1 - f_{\sigma,-i,s_{-i}}(D_{1,j})|$. Then

$$|f_{\sigma,-i,s_{-i}}(D_{0,j}) - f_{\sigma,-i,s_{-i}}(D_{1,j})| \leq \frac{|1 - f_{\sigma,-i,s_{-i}}(D_{1,j})|}{|\pi^{d*}_{-i,s_{-i}}|}|D_{j,s_{-i}} - \pi^{d*}_{-i,s_{-i}}|.$$

If $D_{j,s_{-i}} < 0$, since $f_{\sigma,-i,s_{-i}}$ is convex and increasing in $D_{j,s_{-i}}$,

$$\begin{aligned}\frac{f_{\sigma,-i,s_{-i}}(D_{0,j}) - f_{\sigma,-i,s_{-i}}(D_{1,j})}{D_{j,s_{-i}} - \pi^{d*}_{-i,s_{-i}}} \leq& \frac{f_{\sigma,-i,s_{-i}}((0, D_{j,t_{-i}}, ..., D_{j,z_{-i}})) - f_{\sigma,-i,s_{-i}}(D_{1,j})}{-\pi^{d*}_{-i,s_{-i}}} \\ \leq& \frac{|1 - f_{\sigma,-i,s_{-i}}(D_{1,j})|}{|\pi^{d*}_{-i,s_{-i}}|}\end{aligned}$$

where $f_{\sigma,-i,s_{-i}}((0, D_{j,t_{-i}}, ..., D_{j,z_{-i}})) - f_{\sigma,-i,s_{-i}}(D_{1,j}) > 0$. Therefore, for any $D_{j,s_{-i}}$,

$$|f_{\sigma,-i,s_{-i}}(D_{0,j}) - f_{\sigma,-i,s_{-i}}(D_{1,j})| \leq \frac{|1 - f_{\sigma,-i,s_{-i}}(D_{1,j})|}{|\pi^{d*}_{-i,s_{-i}}|}|D_{j,s_{-i}} - \pi^{d*}_{-i,s_{-i}}|.$$

Next, for $m \neq 0$, let u_{-i} be such that $||D_{m,j} - D_{m+1,j}||_\infty = |D_{j,u_{-i}} - \pi^{d*}_{-i,u_{-i}}|$. Then

$$\begin{aligned}|f_{\sigma,-i,s_{-i}}(D_{m,j}) - f_{\sigma,-i,s_{-i}}(D_{m+1,j})| =& \frac{|f_{\sigma,-i,s_{-i}}(D_{m,-i}) - f_{\sigma,-i,s_{-i}}(D_{m+1,-i})|}{||D_{m,j} - D_{m+1,j}||_\infty}||D_{m,j} - D_{m+1,j}||_\infty \\ =& \frac{|f_{\sigma,-i,s_{-i}}(D_{m,-i}) - f_{\sigma,-i,s_{-i}}(D_{m+1,-i})|}{|D_{j,u_{-i}} - \pi^{d*}_{-i,u_{-i}}|}|D_{j,u_{-i}} - \pi^{d*}_{-i,u_{-i}}|.\end{aligned}$$

If $D_{j,u_{-i}} \leq 0$, since $f_{\sigma,-i,s_{-i}}$ is concave and decreasing in $D_{j,u_{-i}}$, we have

$$\begin{aligned}&\frac{|f_{\sigma,-i,s_{-i}}(D_{m,-i}) - f_{\sigma,-i,s_{-i}}(D_{m+1,-i})|}{|D_{j,u_{-i}} - \pi^{d*}_{-i,u_{-i}}|} \\ \leq& \frac{|f_{\sigma,-i,s_{-i}}(\pi^{d*}_{-i,s_{-i}}, ..., \pi^{d*}_{-i,t_{-i}}, 0, D_{j,v_{-i}}, ..., D_{j,z_{-i}}) - f_{\sigma,-i,s_{-i}}(D_{m+1,-i})|}{|\pi^{d*}_{-i,u_{-i}}|} \\ \leq& \frac{f_{\sigma,-i,s_{-i}}(D_{m+1,j})}{|\pi^{d*}_{-i,u_{-i}}|},\end{aligned}$$

as $f_{\sigma,-i,s_{-i}}(\pi^{d*}_{-i,s_{-i}},...,\pi^{d*}_{-i,t_{-i}},0,D_{j,v_{-i}},...,D_{j,z_{-i}})-f_{\sigma,-i,s_{-i}}(D_{m+1,j})<0$. While if $D_{j,u_{-i}}>0$, since $|D_{j,u_{-i}}-\pi^{d*}_{-i,u_{-i}}|>|\pi^{d*}_{-i,u_{-i}}|$ and $f_{\sigma,-i,s_{-i}}(D_{m,j})-f_{\sigma,-i,s_{-i}}(D_{m+1,j})<0$, we have

$$\frac{|f_{\sigma,-i,s_{-i}}(D_{m,j})-f_{\sigma,-i,s_{-i}}(D_{m+1,j})|}{|D_{j,u_{-i}}-\pi^{d*}_{-i,u_{-i}}|}<\frac{|f_{\sigma,-i,s_{-i}}(D_{m+1,j})|}{|\pi^{d*}_{-i,u_{-i}}|}.$$

Therefore, for any $D_{j,u_{-i}}$, we have

$$|f_{\sigma,-i,s_{-i}}(D_{m,j})-f_{\sigma,-i,s_{-i}}(D_{m+1,j})|\leq\frac{|f_{\sigma,-i,s_{-i}}(D_{m+1,j})|}{|\pi^{d*}_{-i,u_{-i}}|}|D_{j,u_{-i}}-\pi^{d*}_{-i,u_{-i}}|.$$

Now, given the argument above, we have the following inequality: for $D=(D_{i,s_i})_{i,s_i\neq s_i^*}$ and $\pi^{d*}=(\pi^{d*}_{i,s_i})_{i,s_i\neq s_i^*}$,

$$\begin{aligned}|x_{\sigma,j,s_{-i}}-x^*_{\sigma,-i,s_{-i}}|\leq&\sum_{m=0}^{|S_{-i}|-2}|f_{\sigma,-i,s_{-i}}(D_{m,-i})-f_{\sigma,-i,s_{-i}}(D_{m+1,-i})|\\ \leq&\Big(\frac{|1-f_{\sigma,-i,s_{-i}}(D_{1,j})|}{|\pi^{d*}_{-i,s_{-i}}|}+\sum_{m=1}^{|S_{-i}|-2}\frac{|f_{\sigma,-i,s_{-i}}(D_{m+1,j})|}{|\pi^{d*}_{-i,u_{-i}}|}\Big)||D-\pi^{d*}||_\infty.\end{aligned}$$

Note that for each $s_i\neq s_i^*$, $\pi^{d*}_{i,s_i}\to\pi_i(s_i,s^*_{-i})-\pi_i(s_i^*,s^*_{-i})<0$ and for each $m\geq 1$, $f_{\sigma,-i,s_{-i}}(D_{m,j})\to 0$ as $\sigma\to\infty$.[29] Therefore, for any ε, we can pick large σ such that[30]

$$\begin{aligned}|x_{\sigma,j,s_{-i}}-x^*_{\sigma,-i,s_{-i}}|\leq&\Big(\frac{|1-f_{\sigma,-i,s_{-i}}(D_{1,j})|}{|\pi^{d*}_{-i,s_{-i}}|}+\sum_{m=1}^{|S_{-i}|-2}\frac{|f_{\sigma,-i,s_{-i}}(D_{m+1,j})|}{|\pi^{d*}_{-i,u_{-i}}|}\Big)||D-\pi^{d*}||_\infty\\ \leq&\frac{(1+\varepsilon)}{\min_{i,s_i}|\pi_i(s_i,s^*_{-i})-\pi_i(s_i^*,s^*_{-i})|}||D-\pi^{d*}||_\infty.\end{aligned}$$

[29] For $m\geq 1$, $D_{m,j}=(\pi^{d*}_{-i,s_{-i}},...,D_{j,z_{-i}})$, and thus the numerator of f converges to zero as σ diverges. Note that f is expressed as follows:

$$f_{\sigma,-i,s_{-i}}(D_{m,j})=\frac{e^{\sigma\pi^{d*}_{-i,s_{-i}}}}{1+\sum_{t_{-i}\neq s^*_{-i}}e^{\sigma D_{-i,t_{-i}}}},$$

where, for the denominator, (i) $D_{-i,t_{-i}}$ is $\pi^{d*}_{-i,t_{-i}}$ or $D_{j,t_{-i}}$, depending on m, and (ii) each $e^{\sigma D_{-i,t_{-i}}}$ diverges to infinity, converges to zero, or converges to one. Even in the case where $e^{\sigma D_{-i,t_{-i}}}$ for each t_{-i} converges to zero, since there exists the term 1, $f_{\sigma,-i,s_{-i}}(D_{m,j})$ converges to zero as σ diverges.

[30] Note that

$$f_{\sigma,-i,s_{-i}}(D_{m,j})=\frac{e^{\sigma\pi^{d*}_{-i,s_{-i}}}}{1+\sum_{t_{-i}\neq s^*_{-i}}e^{\sigma D_{-i,t_{-i}}}}<e^{\sigma\pi^{d*}_{-i,s_{-i}}}$$

and σ does not depend on $D_{m,j}$.

Therefore,

$$\begin{aligned}|G_{i,s_i}(D)-\pi^{d*}_{i,s_i}| \leq & \sum_{j\in N_l} p_{(i,j)|i} b \sum_{s_{-i}\neq s^*_{-i}} |x_{\sigma,j,s_{-i}} - x^*_{\sigma,-i,s_{-i}}| \\ \leq & p_{\max} b \frac{(1+\varepsilon)(|S_{-i}|-1)}{\min_{i,s_i}|\pi_i(s_i,s^*_{-i})-\pi_i(s^*_i,s^*_{-i})|}||D-\pi^{d*}||_\infty \\ \leq & p_{\max} b \frac{(1+\varepsilon)\mathcal{M}}{\min_{i,s_i}|\pi_i(s_i,s^*_{-i})-\pi_i(s^*_i,s^*_{-i})|}||D-\pi^{d*}||_\infty\end{aligned}$$

where $\mathcal{M} := \max_i(|S_i|-1)$. Now, if $p_{\max}|b|\frac{\mathcal{M}}{\min_{i,s_i}|\pi_i(s_i,s^*_{-i})-\pi_i(s^*_i,s^*_{-i})|} < 1$, there exists ε such that $p_{\max} b \frac{(1+\varepsilon)\mathcal{M}}{\min_{i,s_i}|\pi_i(s_i,s^*_{-i})-\pi_i(s^*_i,s^*_{-i})|} < 1$. Thus, if we pick σ under which the above inequality holds, then the adaptive learning process converges to the LQRE x^*_σ.

Appendix E Proof of Lemma 1

Note that

$$\begin{aligned}&Q_{n+1,i,s_i} - Q_{n+1,i,s^*_i} = (1-\alpha_{n,i})(Q_{n,i,s_i} - Q_{n,i,s^*_i}) + \alpha_{n,i}(\pi_i(s_i,x_{n,-i}) - \pi_i(s^*_i,x_{n,-i}) + M^d_{n,i,s_i}) \\ =&(1-\alpha_{n,i})(1-\alpha_{n-1,i})(Q_{n-1,i,s_i} - Q_{n-1,i,s^*_i}) \\ &+ \alpha_{n-1,i}(1-\alpha_{n,i})(\pi_i(s_i,x_{n-1,-i}) - \pi_i(s^*_i,x_{n-1,-i}) + M^d_{n-1,i,s_i}) \\ &+ \alpha_{n,i}(\pi_i(s_i,x_{n,-i}) - \pi_i(s^*_i,x_{n,-i}) + M^d_{n,i,s_i}) \\ =&\cdots \\ =&\Big(\prod_{m=0}^{n}(1-\alpha_{m,i})\Big)(Q_{0,i,s_i} - Q_{0,i,s^*_i}) + \sum_{m=1}^{n}\alpha_{m,i}\Big(\prod_{l=m+1}^{n}(1-\alpha_{i,l})\Big)(\pi_i(s_i,x_{m,-i}) - \pi_i(s^*_i,x_{m,-i}) + M^d_{m,i,s_i}),\end{aligned}$$

where for $m=n$, $\prod_{l=n+1}^{n}(1-\alpha_{i,l}) := 1$. Note that $\prod_{m=0}^{n}(1-\alpha_{m,i})$ converges to zero as $n\to 0$.[31] Therefore, for any ε, we can take N_1 such that for $n > N_1$,

$$\Big(\prod_{m=0}^{n}(1-\alpha_{m,i})\Big)|Q_{0,i,t_i} - Q_{0,i,s_i}| < \frac{\varepsilon}{2}.$$

Also, note that as $n\to\infty$,

$$\sum_{m=1}^{n}\alpha_{m,i}\Big(\prod_{l=m+1}^{n}(1-\alpha_{i,l})\Big)M^d_{m,i,s_i} \to 0$$

[31] Since $\alpha_{n,i}\to 0$ as $n\to 0$, there exists N such that for $n > N$, $\alpha_{n,i} < 1$. Note that $\prod_{m=0}^{n}(1-\alpha_{m,i}) = \prod_{m=0}^{N-1}(1-\alpha_{m,i})\prod_{m=N}^{n}(1-\alpha_{m,i})$, and thus we focus on the latter term. In particular, we now show that for $n\geq N$ and $0<\alpha_{n,i}<1$, $\prod_{m=N}^{n}(1-\alpha_{m,i})$ converges to zero as $n\to\infty$ when $\sum_m \alpha_m = \infty$. First, note that $\sum_{m=N}^{n}\alpha_{m,i} < \prod_{m=N}^{n}(1+\alpha_{m,i})$, and thus $\prod_{m=N}^{n}(1+\alpha_{m,i})$ diverges to infinity. Next, note that $\prod_{m=N}^{n}(1-\alpha_{m,i})\prod_{m=N}^{n}(1+\alpha_{m,i}) \leq 1$, as $(1-\alpha_{m,i})(1+\alpha_{m,i}) = 1-\alpha^2_{m,i} \leq 1$ for each m. Therefore, $\prod_{m=N}^{n}(1-\alpha_{m,i}) \leq \frac{1}{\prod_{m=N}^{n}(1+\alpha_{m,i})}$ converges to zero.

with probability one.[32] Therefore, we take N_2 such that for $n > N_2$,

$$\sum_{m=1}^{n} \alpha_{m,i}\Big(\prod_{l=m+1}^{n}(1-\alpha_{i,l})\Big)M^d_{m,i,s_i} < \frac{\varepsilon}{2}.$$

Since for any x_{-i},

$$\pi_i(s_i, x_{-i}) - \pi_i(s_i^*, x_{-i}) \in [\min_{s_{-i}}(\pi_i(s_i, s_{-i}) - \pi_i(s_i^*, s_{-i})), \max_{s_{-i}}(\pi_i(s_i, s_{-i}) - \pi_i(s_i^*, s_{-i}))],$$

for $n > N := \max\{N_1, N_2\}$,

$$Q_{n,i,s_i} - Q_{n,i,s_i^*} \in (\min_{s_{-i}}(\pi_i(s_i, s_{-i}) - \pi_i(s_i^*, s_{-i})) - \varepsilon, \max_{s_{-i}}(\pi_i(s_i, s_{-i}) - \pi_i(s_i^*, s_{-i})) + \varepsilon).$$

Note that s_i^* strictly dominates s_i, $\max_{s_{-i}}(\pi_i(s_i, s_{-i}) - \pi_i(s_i^*, s_{-i})) < 0$. Therefore, we can pick small enough ε such that $\max_{s_{-i}}(\pi_i(s_i, s_{-i}) - \pi_i(s_i^*, s_{-i})) + \varepsilon) < 0$.

Appendix F Proof of Proposition 5

Let s_i^* be the strictly dominant action for player i. Then, from Lemma 1, for some $\eta > 0$ such that $\max_{s_{-i}}(\pi_i(s_i, s_{-i}) - \pi_i(s_i^*, s_{-i})) + \eta) < 0$, there exists N such that for $n > N$, $Q^d_{n,s_i} \in (\min_{s_{-i}}(\pi_i(s_i, s_{-i}) - \pi_i(s_i^*, s_{-i})) - \eta, \max_{s_{-i}}(\pi_i(s_i, s_{-i}) - \pi_i(s_i^*, s_{-i})) + \eta)$ for $s_i \neq s_i^*$. We now follow the argument of Appendix D, where the domain of G and f is restricted to the set $\mathcal{H}_\eta := \prod_{i,s_i}[\min_{s_{-i}}(\pi_i(s_i, s_{-i}) - \pi_i(s_i^*, s_{-i})) - \eta, \max_{s_{-i}}(\pi_i(s_i, s_{-i}) - \pi_i(s_i^*, s_{-i})) + \eta]$. Consider the sequence $(D_{0,j}, D_{1,j}, ..., D_{|S_{-i}|-1,j})$ and equation (4) in Appendix D. For $m = 0$ of equation (4), there exist $\overline{D}_{j,s_{-i}} \in (D_{j,s_{-i}} \wedge \pi^{d*}_{-i,s_{-i}}, D_{j,s_{-i}} \vee \pi^{d*}_{-i,s_{-i}})$ and $\overline{\sigma}_{\varepsilon,j,s_{-i}}$ such that for $\overline{D}_j = (\overline{D}_{j,s_{-i}}, ..., D_{j,z_{-i}})$ and $\sigma > \overline{\sigma}_{\varepsilon,j,s_{-i}}$,

$$\begin{aligned}\frac{|f_{\sigma,-i,s_{-i}}(D_0) - f_{\sigma,-i,s_{-i}}(D_1)|}{|D_{j,s_{-i}} - \pi^{d*}_{-i,s_{-i}}|} =& |\frac{\partial}{\partial D_{-i,s_{-i}}} f_{\sigma,-i,s_{-i}}(\overline{D}_j)| \\ \leq& |\frac{\partial}{\partial D_{-i,s_{-i}}} f_{\sigma,-i,s_{-i}}(D'_j)| \\ \leq& \varepsilon,\end{aligned}$$

where $D'_j = (\max_{s_{-i}}(\pi_i(s_i, s_{-i}) - \pi_i(s_i^*, s_{-i})) + \eta, D_{j,u_{-i}}, ..., D_{j,z_{-i}})$. Note that on $\mathcal{H}_\eta$, $\frac{\partial}{\partial D_{-i,s_{-i}}} f_{\sigma,-i,s_{-i}} = \sigma f_{\sigma,-i,s_{-i}}(1 - f_{\sigma,-i,s_{-i}})$ is increasing and positive and converges to 0 as $\sigma \to \infty$. For $m > 0$, there exist $\overline{D}_{j,u_{-i}} \in (D_{j,u_{-i}} \wedge \pi^{d*}_{-i,u_{-i}}, D_{j,u_{-i}} \vee \pi^{d*}_{-i,u_{-i}})$ and $\overline{\sigma}_{\varepsilon,j,u_{-i}}$ such that for $\overline{D}_j = (\pi^{d*}_{-i,s_{-i}}, ..., \pi^{d*}_{-i,s_{-i}}, \overline{D}_{j,u_{-i}}, D_{j,v_{-i}}...., D_{j,z_{-i}})$ and $\sigma > \overline{\sigma}_{\varepsilon,j,u_{-i}}$,

$$\begin{aligned}\frac{|f_{\sigma,-i,s_{-i}}(D_m) - f_{\sigma,-i,s_{-i}}(D_{m+1})|}{|D_{j,u_{-i}} - \pi^{d*}_{-i,u_{-i}}|} =& |\frac{\partial}{\partial D_{-i,u_{-i}}} f_{\sigma,-i,s_{-i}}(\overline{D}_j)| \\ \leq& \varepsilon.\end{aligned}$$

[32] For instance, see Lemma 1 of Tsitsiklis (1994).

Note that $\frac{\partial}{\partial D_{-i,u_{-i}}} f_{\sigma,-i,s_{-i}} = -\sigma f_{\sigma,-i,u_{-i}} f_{\sigma,-i,s_{-i}} \to 0$ as $\sigma \to \infty$ on $\mathcal{H}_\eta$. Then for $\sigma > \max_{i,s_i} \overline{\sigma}_{\varepsilon,i,s_i}$,

$$\begin{aligned}
|G_{i,s_i}(D) - \pi^{d*}_{i,s_i}| =& |\sum_j p_{(i,j)|i} \sum_{s_{-i} \neq s^*_{-i}} (\pi_i(s_i, s_{-i}) - \pi_i(s^*_i, s_{-i}) - \pi_i(s_i, s^*_{-i}) + \pi_i(s^*_i, s^*_{-i}))(x_{\sigma,j,s_{-i}} - x^*_{\sigma,-i,s_{-i}})| \\
\leq& \sum_j p_{(i,j)|i} b \sum_{s_{-i} \neq s^*_{-i}} |x_{\sigma,j,s_{-i}} - x^*_{\sigma,-i,s_{-i}}| \\
\leq& \sum_j p_{(i,j)|i} b \sum_{s_{-i} \neq s^*_{-i}} \sum_{m=0}^{|S_{-i}|-2} \frac{|f_{\sigma,-i,s_{-i}}(D_m) - f_{\sigma,-i,s_{-i}}(D_{m+1})|}{|D_{j,u_{-i}} - \pi^{d*}_{-i,u_{-i}}|} |D_{j,u_{-i}} - \pi^{d*}_{-i,u_{-i}}| \\
\leq& p_{\max} b (|S_{-i}| - 1)^2 \varepsilon ||D - \pi^{d*}||_\infty
\end{aligned}$$

where $b := \max_{i,s_i} |\pi_i(s_i, s_{-i}) - \pi_i(s^*_i, s_{-i}) - \pi_i(s_i, s^*_{-i}) + \pi_i(s^*_i, s^*_{-i})|$, $p_{\max} := \max_i p_{(i,\mathcal{N}_l)|i}$. Therefore, if we pick ε such that $p_{\max} b (|S_{-i}|-1)^2 \varepsilon < 1$, the process converges to the strictly dominant strategy equilibrium.

Appendix G Proof of Proposition 6

Let Q^{**} be the limit assessment profile. Since the choice rule is continuous, the choice probability also converges; let $x^{**}_{\sigma,i,s_i} = \frac{e^{\sigma Q^{**}_{i,s_i}}}{\sum_{t_i \in S_i} e^{\sigma Q^{**}_{i,t_i}}}$ be the limit choice probability for action s_i. Since we assume that adaptive players' limit choice probability of each action coincides with that of committed players, we have $x^{**}_{\sigma,i,s_i} = \overline{x}_{i,s_i}$ for each $i \in \mathcal{N}_\tau$ and $s_i \in S_i$, where without any confusion, i and $-i$ also denote player i's type and her opponent's type, respectively. Now, let $\overline{Q}_{i,s_i} = \sum_{s_{-i} \in S_{-i}} \pi(s_i, s_{-i}) \overline{x}_{-i,s_{-i}}$ be the expected payoff given committed players' choice probability profile $\overline{x} = (\overline{x}_{i,s_i})$. Then, we prove by contradiction that $\overline{Q} = Q^{**}$, which means that $\overline{x}$ corresponds to an LQRE given precision parameter σ. Now we assume that $\overline{Q} \neq Q^{**}$. Since we assume that the choice probability profile of adaptive players almost surely converges to what committed players follow, the updating rule can be expressed as follows: for each n, i and s_i,

$$Q_{n+1,i,s_i} = Q_{n,i,s_i} + \alpha_{n,i}\big(\overline{Q}_{i,s_i} - Q_{n,i,s_i} + M_{n,i,s_i} + \eta_{n,i,s_i}\big)$$

where $\eta_{n,i,s_i} = \overline{\pi}_{n,i,s_i} - \overline{Q}_{i,s_i}$ is a noise which disappears with probability one.[33] Then by the asynchronous stochastic approximation method, the assessment profile almost surely converges to $\overline{Q}$, which contradicts the assumption.[34]

[33] $\overline{\pi}_{n,i,s_i} = \sum_{s_{-i} \in S_{-i}} \pi_i(s_i, s_{-i}) (\sum_{j \in \mathcal{N}_l} p_{(i,j)|i} x_{n,\sigma,j,s_{-i}} + (1 - p_{(i,\mathcal{N}_l)|i}) \overline{x}_{-i,s_{-i}})$ almost surely converges to $\sum_{s_{-i} \in S_{-i}} \pi_i(s_i, s_{-i}) \overline{x}_{-i,s_{-i}}$ as the choice probability of adaptive players almost surely converges to $\overline{x}$.

[34] For the stochastic approximation method with a noise which disappears with probability one, see Appendix B of Funai (2022), for instance.

Appendix H Proof for Proposition 7

Note that the updating rule is expressed as follows:

$$\begin{aligned}Q_{n+1,i,s_i} =& Q_{n,i,s_i} + \lambda_{n,i} \sum_j \mathbb{1}_{n,i,j} \Big(\sum_{s_{-i} \in S} \pi_i(s_i, s_{-i}) \mathbb{1}_{n,j,s_{-i}} - Q_{n,i,s_i} \Big) \\ =& Q_{n,i,s_i} + \alpha_{n,i} \Big(\overline{\pi}_{n,i,s_i} - Q_{n,i,s_i} + M_{n,i,s_i} \Big),\end{aligned}$$

where instead of expressing $\overline{\pi}_{n,i,s_i}$ as

$$\overline{\pi}_{n,i,s_i} = \sum_{s_{-i} \in S_{-i}} \pi_i(s_i, s_{-i}) \Big(\sum_{j \in \mathcal{N}_l} p_{(i,j)|i} x_{n,\sigma,j,s_{-i}} + (1 - p_{(i,\mathcal{N}_l)|i}) \overline{x}_{-i,s_{-i}} \Big),$$

we express it as follows:[35]

$$\overline{\pi}_{n,i,s_i} = \sum_{s_{-i} \in S_{-i}} \pi_i'(s_i, s_{-i}) \sum_{j \in \mathcal{N}_l} p_{(i,j)|(i,\mathcal{N}_l)} x_{n,\sigma,j,s_{-i}}.$$

Therefore, by utilising the argument in Proposition 4 by changing (i) π to π', (ii) $p_{(i,j)|i}$ to $p_{(i,j)|(i,\mathcal{N}_l)}$, (iii) $p_{(i,\mathcal{N}_\nu)|i}$ to 0 and (iv) $p_{\max} = \max_i \sum_{j \in \mathcal{N}_l} p_{(i,j)|i}$ to $\max_i \sum_{j \in \mathcal{N}_l} p_{(i,j)|(i,\mathcal{N}_l)} = 1$, we obtain the result.

---

[35] Note that

$$\begin{aligned}\overline{\pi}_{n,i,s_i} &= \sum_{s_{-i} \in S_{-i}} \pi_i(s_i, s_{-i}) \Big(\sum_{j \in \mathcal{N}_l} p_{(i,j)|i} x_{n,\sigma,j,s_{-i}} + \sum_{k \in \mathcal{N}_\nu} p_{(i,k)|i} \overline{x}_{k,s_{-i}} \Big) \\ &= \sum_{s_{-i} \in S_{-i}} \pi_i(s_i, s_{-i}) \Big(\sum_{j \in \mathcal{N}_l} p_{(i,j)|i} x_{n,\sigma,j,s_{-i}} + p_{(i,\mathcal{N}_\nu)|i} \sum_{k \in \mathcal{N}_\nu} p_{(i,k)|(i,\mathcal{N}_\nu)} \overline{x}_{k,s_{-i}} \Big) \\ &= \sum_{s_{-i} \in S_{-i}} \Big(\sum_{j \in \mathcal{N}_l} p_{(i,j)|i} \pi_i(s_i, s_{-i}) x_{n,\sigma,j,s_{-i}} + p_{(i,\mathcal{N}_\nu)|i} \sum_{k \in \mathcal{N}_\nu} p_{(i,k)|(\mathcal{N}_\nu,i)} \pi_i(s_i, s_{-i}) \overline{x}_{k,s_{-i}} \Big) \\ &= \sum_{s_{-i} \in S_{-i}} \sum_{j \in \mathcal{N}_l} p_{(i,j)|i} \pi_i(s_i, s_{-i}) x_{n,\sigma,j,s_{-i}} + p_{(i,\mathcal{N}_\nu)|i} \pi_i(s_i, \overline{x}_{-i}) \\ &= \sum_{s_{-i} \in S_{-i}} \sum_{j \in \mathcal{N}_l} p_{(i,j)|(i,N_l)} p_{(i,N_l)|i} \pi_i(s_i, s_{-i}) x_{n,\sigma,j,s_{-i}} \\ &\quad + (1 - p_{(i,\mathcal{N}_l)|i}) \sum_{s_{-i} \in S_{-i}} \sum_{j \in \mathcal{N}_l} p_{(i,j)|(i,\mathcal{N}_l)} \pi_i(s_i, \overline{x}_{-i}) x_{n,\sigma,j,s_{-i}} \\ &= \sum_{s_{-i} \in S_{-i}} \pi_i'(s_i, s_{-i}) \sum_{j \in \mathcal{N}_l} p_{(i,j)|(i,\mathcal{N}_l)} x_{n,\sigma,j,s_{-i}}.\end{aligned}$$

Appendix I Proof for Lemma 2

Note that for $Q \in \mathcal{Q} = \{Q \in \mathbb{R}^{|S|} : -K < Q_{i,s_i} - Q_{i,s_i^*} < -\varepsilon \text{ for any } i \text{ and } s_i \neq s_i^*\}$ with $0 < \varepsilon < K$,

$$f_{\sigma,i,s_i}(Q_i) = \frac{e^{\sigma Q_{i,s_i}}}{\sum_{t_i} e^{\sigma Q_{i,t_i}}}$$

converges to 1 if $s_i = s_i^*$ and 0 otherwise when $\sigma \to \infty$. Let $V : \mathcal{Q} \to [0, \infty)$ be such that for $Q \in \mathcal{Q}$, $V(Q) := ||Q - \pi^*||_\infty$. Let s_i be such that $||Q - \pi^*||_\infty = |Q_{i,s_i} - \pi^*_{i,s_i}|$. Also, without loss of generality, we assume that $|Q_{i,s_i} - \pi^*_{i,s_i}| = Q_{i,s_i} - \pi^*_{i,s_i}$. Note that for $t_i \in S_i$ and $c \in \mathbb{R}$, $c(Q_{i,t_i} - \pi^*_{i,t_i}) \leq |c(Q_{i,t_i} - \pi^*_{i,t_i})| \leq |c||(Q_{i,s_i} - \pi^*_{i,s_i})| = |c|(Q_{i,s_i} - \pi^*_{i,s_i})$. Now

$$\begin{aligned}
\dot{V}(Q_t) =& \overline{\gamma}_{n,i,s_i}(\pi_i(s_i, x_{t,\sigma,-i}) - Q_{t,i,s_i}) \\
=& \overline{\gamma}_{n,i,s_i}(\pi_i(s_i, x_{t,\sigma,-i}) - \pi^*_{i,s_i} - (Q_{t,i,s_i} - \pi^*_{i,s_i})) \\
=& \overline{\gamma}_{n,i,s_i}(\pi_i(s_i, x_{t,\sigma,-i}) - \pi_i(s_i, x^*_{\sigma,-i}) - (Q_{t,i,s_i} - \pi_i(s_i, x^*_{\sigma,-i})),
\end{aligned}$$

where $x^*_{\sigma,-i} = (x^*_{\sigma,-i,s_{-i}})_{s_{-i}}$ and $x^*_{\sigma,-i,s_{-i}} = \frac{e^{\sigma \pi^*_{-i,s_{-i}}}}{\sum_{t_i} e^{\sigma \pi^*_{-i,t_{-i}}}}$. Note that

$$\begin{aligned}
\pi_i(s_i, x_{t,\sigma,-i}) - \pi_i(s_i, x^*_{\sigma,-i}) =& \sum_{s_{-i}} \pi(s_i, s_{-i})(x_{t,\sigma,-i,s_{-i}} - x^*_{\sigma,-i,s_{-i}}) \\
=& \sum_{s_{-i}} \pi(s_i, s_{-i})(\sum_{u_{-i}} \frac{\partial}{\partial Q_{-i,u_{-i}}} f_{\sigma,-i,s_{-i}}(Q'_{-i})(Q_{t,-i,u_{-i}} - \pi^*_{-i,u_{-i}}))
\end{aligned}$$

for some $c \in (0,1)$ and $Q'_{-i} = cQ_{-i} + (1-c)\pi^*_{-i}$. Since $Q'_{-i} \in \mathcal{Q}_{-i} := \{Q_{-i} \in \mathbb{R}^{|S_{-i}|} : -K < Q_{-i,s_{-i}} - Q_{-i,s^*_{-i}} < -\varepsilon \text{ for } s_i \neq s_i^*\}$,

$$\begin{aligned}
\dot{V}(Q_t) =& \overline{\gamma}_{n,i,s_i}(\pi_i(s_i, x_{t,\sigma,-i}) - \pi_i(s_i, x^*_{\sigma,-i}) - (Q_{t,i,s_i} - \pi^*_{i,s_i})) \\
\leq& \overline{\gamma}_{n,i,s_i}((\sum_{s_{-i}} \sum_{u_{-i}} |\pi_i(s_i, s_{-i}) \frac{\partial}{\partial Q_{-i,u_{-i}}} f_{\sigma,-i,s_{-i}}(Q'_{-i})| - 1)(Q_{t,i,s_i} - \pi^*_{i,s_i})).
\end{aligned}$$

Note that (i) for $u_{-i} = s_{-i}$,

$$\begin{aligned}
\frac{\partial}{\partial Q_{-i,s_{-i}}} f_{\sigma,-i,s_{-i}}(Q'_{-i}) =& \frac{\sigma e^{\sigma Q'_{-i,s_{-i}}}(\sum_{v_{-i}} e^{\sigma Q'_{-i,v_{-i}}}) - \sigma e^{\sigma Q'_{-i,s_{-i}}} e^{\sigma Q'_{-i,s_{-i}}}}{(\sum_{v_{-i}} e^{\sigma Q'_{-i,v_{-i}}})^2} \\
=& \sigma f_{\sigma,-i,s_{-i}}(Q'_{-i})(1 - f_{\sigma,-i,s_{-i}}(Q'_{-i}))
\end{aligned}$$

and (ii) for $u_{-i} \neq s_{-i}$,

$$\begin{aligned}\frac{\partial}{\partial Q_{u_{-i}}} f_{\sigma,-i,s_{-i}}(Q'_{-i}) =& -\frac{\sigma e^{\sigma Q'_{s_{-i}}} e^{\sigma Q'_{u_{-i}}}}{(\sum_{v_{-i}} e^{\sigma Q'_{v_{-i}}})^2} \\ =& -\sigma f_{\sigma,-i,s_{-i}}(Q'_{-i}) f_{\sigma,-i,u_{-i}}(Q'_{-i}).\end{aligned}$$

Now, we show that for $|\frac{\partial}{\partial Q_{-i,s_{-i}}} f_{\sigma,-i,s_{-i}}(Q'_{-i})|$ in both cases (i) and (ii), there exist upper bounds which do not depend on $Q \in \mathcal{Q}$ and converge to zero as σ goes to zero, so that there exists $\overline{\sigma}$ such that for any $\sigma > \overline{\sigma}$ and $Q' \in \mathcal{Q}$,

$$(\sum_{s_{-i}} \sum_{u_{-i}} |\pi_i(s_i, s_{-i}) \frac{\partial}{\partial Q_{-i,u_{-i}}} f_{\sigma,-i,s_{-i}}(Q'_{-i})| - 1) < 0,$$

and thus V becomes a Lyapunov function.

First, when $s_{-i} \neq s^*_{-i}$,

$$\begin{aligned}f_{\sigma,-i,s_{-i}}(Q_i) =& \frac{e^{\sigma Q_{-i,s_{-i}}}}{\sum_{t_{-i}} e^{\sigma Q_{-i,t_{-i}}}} \\ =& \frac{e^{\sigma(Q_{-i,s_{-i}} - Q_{-i,s^*_{-i}})}}{1 + \sum_{t_{-i} \neq s^*_{-i}} e^{\sigma(Q_{-i,t_{-i}} - Q_{-i,s^*_{-i}})}} \\ <& \frac{e^{-\sigma\varepsilon}}{1 + (|S_i| - 1)e^{-\sigma K}}\end{aligned}$$

and

$$\sigma f_{\sigma,-i,s_{-i}}(Q_{-i}) < \frac{\sigma}{e^{\sigma\varepsilon} + (|S_i| - 1)e^{\sigma(\varepsilon - K)}},$$

where the right-hand side of the inequality above converges to 0 as σ goes to infinity. Therefore, since both s_{-i} and u_{-i} cannot be s^*_{-i}, for the case (i) in which $s_{-i} \neq s^*_{-i}$ and case (ii), $\frac{\sigma}{e^{\sigma\varepsilon} + (|S_i| - 1)e^{\sigma(\varepsilon - K)}}$ is the upper bound.

Next, for the case (i) in which $s_{-i} = s^*_{-i}$, we have

$$\begin{aligned}\sigma(1 - f_{-i,s^*_i}(Q_i)) =& \sum_{t_{-i} \neq s^*_{-i}} \sigma f_{\sigma,-i,t_{-i}}(Q_{-i}) \\ <& (|S_i| - 1) \frac{\sigma}{e^{\sigma\varepsilon} + (|S_i| - 1)e^{\sigma(\varepsilon - K)}},\end{aligned}$$

where the right-hand side of the inequality above converges to zero as σ goes to infinity. Therefore, for the case (i) in which $s_{-i} = s^*_{-i}$, $(|S_i| - 1)\frac{\sigma}{e^{\sigma\varepsilon} + (|S_i| - 1)e^{\sigma(\varepsilon - K)}}$ is the upper bound.

Appendix J Proof for Proposition 8

By Proposition 2, there exists σ such that on $\mathcal{Q}$, V becomes a Lyapunov function. To utilise Corollary 12 of Borkar (2008), we also need to show that (i) we can pick n_0 such that $(C\lambda_{n_0} + cK_1CLb(n_0))c$, where K_1 is some constant, is small enough and (ii) we can pick some constant K_2 such that $||M_m||_\infty \leq K_2(1 + ||Q_{m-1}||_\infty)$ for $m \geq 1$. Regarding (i), since λ_n and $b(n)$ converge to zero as n diverges, there exists such n_0. Regarding (ii), since M_n is bounded for each n, we can pick K_2 large enough so that the inequality holds.

Appendix K Proof for Lemma 3

Consider the event $\mathcal{E}$ defined by the following steps.

- Step 1: each player i except player 1 plays s_i^* and player 1 plays all of her actions except s_1^* until some period $n_1 > \overline{n}$ such that $Q_{n_1,1,s_1} < \pi^*_{1,s_1^*} - \frac{1}{2}\varepsilon'$ for each $s_1 \neq s_1^*$.
- Step 2: each player i except player 2 plays s_i^* and player 2 plays all of her actions except s_2^* until some period $n_2 > n_1$ such that $Q_{n_2,2,s_2} < \pi^*_{2,s_2^*} - \frac{1}{2}\varepsilon'$ for each $s_2 \neq s_2^*$.
- Step 3 to N: follow the same procedure as Step 1 and Step 2 for all the remaining players.
- Step $N+1$: players play $s^* = (s_i^*)_i$ until some period $n_0 > n_N$ such that (i) $Q_{n_0,i,s_i^*} \in (\pi^*_{i,s_i^*} - \frac{1}{2}\varepsilon', \pi^*_{i,s_i^*} + \frac{1}{2}\varepsilon')$ for each i and s_i and (ii) $Q_{n_0,i,s_i} < \pi^*_{i,s_i^*} - \frac{1}{2}\varepsilon'$ for each i and $s_i \neq s_i^*$.

Note that if $\gamma'_{i,s_i,s_i^*} = 0$ for any i and $s_i \neq s_i^*$, then at Step $N+1$, s^* is repeatedly played until Q_{n,i,s_i^*} reaches the interval for each i. If $\gamma'_{i,s_i,s_i^*} \neq 0$ for some i and s_i, Q_{n,i,s_i} may be adjusted and become greater than or equal to $\pi^*_{i,s_i^*} - \frac{1}{2}\varepsilon'$ when player i chooses s_i^* and some player $j > i$ chooses $s_j \neq s_j^*$ after Step i. In this case, at Step $N+1$, s^* is repeatedly played until both conditions (i) and (ii) at Step $N+1$ are satisfied. Note that for any assessment profile in period $\overline{n}$, this event happens in some finite period as the sets of players and actions are finite. Note also that since the probability of any action profile $s = (s_i)_i$ being chosen in each period given the assessment profile is greater than $\prod_i f_{\sigma,i,s_i}(\underline{Q}_i) > 0$, where for $\underline{Q}_i = (\underline{Q}_{i,u_i})_{u_i}$, $\underline{Q}_{i,s_i} := \min\{Q_{0,i,s_i}, \min_{t_{-i}} \pi_i(s_i, t_{-i})\}$ and $\underline{Q}_{i,t_i} := \max\{Q_{0,i,t_i}, \max_{t_{-i}} \pi_i(t_i, t_{-i})\}$ for $t_i \neq s_i$, we have $P(\mathcal{E}) > 0$. Lastly, since $\mathcal{E} \subset \{Q_{n_0} \in B_{\varepsilon'}\}$, $P(Q_{n_0} \in B_{\varepsilon'}) > 0$.

Appendix L Proof for Proposition 10

Note that the updating rule can be expressed as follows:

$$\begin{aligned} Q_{n+1,i,s_i} =& Q_{n,i,s_i} + \lambda_n \sum_{j\in\mathcal{N}} \mathbb{1}_{n,i,j}\big(\gamma_{n,i,s_i}(\sum_{s_{-i}\in S_{-i}} \pi_i(s_i,s_{-i})\mathbb{1}_{n,j,s_{-i}} - Q_{n,i,s_i})\big) \\ =& Q_{n,i,s_i} + \lambda_n p_i\overline{\gamma}_{n,i,s_i}(\sum_{s_{-i}\in S_{-i}} \pi_i'(s_i,s_{-i})x_{n,\sigma,-i,s_{-i}} - Q_{n,i,s_i} + M'_{n,i,s_i}), \end{aligned}$$

where $x_{n,\sigma,-i,s_i}$ denotes the probability that player i's adaptive opponent chooses action s_{-i} in period n and

$$\begin{aligned} M'_{n,i,s_i} :=& \frac{1}{p_i\overline{\gamma}_{n,i,s_i}}\Big(\sum_{j\in\mathcal{N}} \mathbb{1}_{n,i,j}\big(\gamma_{n,i,s_i}(\sum_{s_{-i}\in S_{-i}} \pi_i(s_i,s_{-i})\mathbb{1}_{n,j,s_{-i}} - Q_{n,i,s_i})\big) \\ & - \mathbb{E}[\sum_j \mathbb{1}_{n,i,j}\big(\gamma_{n,i,s_i}(\sum_{s_{-i}\in S_{-i}} \pi_i(s_i,s_{-i})\mathbb{1}_{n,j,s_{-i}} - Q_{n,i,s_i})\big) \mid \mathcal{F}_n]\Big). \end{aligned}$$

The remaining proof follows the argument in Proposition 9.

Acknowledgements

I am grateful to Dai Zusai, Ryoji Sawa, Tadashi Sekiguchi, Jonathan Newton and seminar participants at the 27th Decentralization Conference in Japan, 2022 Japanese Economic Association Autumn Meeting, SING 17, the Lisbon Meetings 2023 and Kyoto University for suggestions and helpful comments. I also thank Ryutsu Keizai University for internal research fund and an opportunity to conduct a part of the analysis. I would like to thank JSPS KAKENHI Grant (# 22H00826) and the institute for economic and business research, Shiga University, for providing research support. I also thank Caroline Orr for proofreading the paper. All remaining errors are mine.

References

[1] Andreoni, J., 1995. Cooperation in public-goods experiments: kindness or confusion? Amer. Econ. Rev. 85, 891–904. http://www.jstor.org/stable/2118238.

[2] Andreoni, J., Miller, J. H., 1993. Rational cooperation in the finitely repeated prisoner's dilemma: experimental evidence. Econ. J. 103, 570–585. https://doi.org/10.2307/2234532.

[3] Benaïm, M., 1999. Dynamics of stochastic approximation algorithms. In: Séminaire de Probabilités, XXXIII, Lecture Notes in Mathematics, vol. 1709, Springer, Berlin, pp. 1–68. https://doi.org/10.1007/BFb0096509

[4] Benaïm, M., Hirsch, M., 1999. Mixed equilibria and dynamical systems arising from fictitious play in perturbed games. Games Econ. Behav. 29, 36–72. https://doi.org/10.1006/game.1999.0717.

[5] Block, J. I., Fudenberg, D., Levine, D. K., 2019. Learning dynamics with social comparisons and limited memory. Theor. Econ. 14, 135–172. https://doi.org/10.3982/TE2626.

[6] Borkar, V. S., 2008. Stochastic approximation: a dynamical systems viewpoint. Cambridge, UK: Cambridge University Press.

[7] Camerer, C., Ho, T. H., 1999. Experience-weighted attraction learning in normal form games. Econometrica 67, 827–874. https://doi.org/10.1111/1468-0262.00054

[8] Carvalho, J. P., 2017. Coordination and culture. Econ. Theory 64, 449–475. https://doi.org/10.1007/s00199-016-0990-3.

[9] Chmura, T., Goerg, J. S., Selten, R., 2012. Learning in experimental 2×2 games. Games Econ. Behav. 76, 44–73. https://doi.org/10.1016/j.geb.2012.06.007.

[10] Cominetti, R., Melo, E., Sorin, S., 2010. A payoff-based learning and its application to traffic games. Games Econ. Behav. 70, 71–83. https://doi.org/10.1016/j.geb.2008.11.012.

[11] Erev, I., Roth, A. E., 1998. Predicting how people play games: reinforcement learning in experimental games with unique mixed strategy equilibria. Amer. Econ. Rev. 88, 848–881. http://www.jstor.org/stable/117009.

[12] Fudenberg, D., Kreps, D. M., 1993. Learning mixed equilibria. Games Econ. Behav. 5, 320–367. https://doi.org/10.1006/game.1993.1021.

[13] Fudenberg, D., Takahashi, S., 2011. Heterogeneous beliefs and local information in stochastic fictitious play. Games Econ. Behav. 71, 100–120. https://doi.org/10.1016/j.geb.2008.11.014.

[14] Funai, N., 2014. An adaptive learning model with foregone payoff information. B.E. J. Theor. Econ. 14, 149–176. https://doi.org/10.1515/bejte-2013-0043.

[15] Funai, N., 2019. Convergence results on stochastic adaptive learning. Econ. Theory. 68, 907–934. https://doi.org/10.1007/s00199-018-1150-8.

[16] Funai, N., 2022. Reinforcement learning with foregone payoff information in normal form games, J. Econ. Behav. Organ. 200, 638–660. https://doi.org/10.1016/j.jebo.2022.06.021.

[17] Goeree, J. Y., Holt, C. A., Palfrey, T. R., 2016. Quantal response equilibrium. Princeton University Press.

[18] Heller, Y., Mohlin, E., 2018. Observations on cooperation. Rev. Econ. Stud. 85, 2253–2282. https://doi.org/10.1093/restud/rdx076.

[19] Ianni, A., 2014. Learning strict Nash equilibria through reinforcement. J. Math. Econ. 50, 148–155. https://doi.org/10.1016/j.jmateco.2013.04.005.

[20] Kreps, D. M., Milgrom, P., Roberts, J., Wilson, R., 1982. Rational cooperation in the finitely repeated prisoners' dilemma. J. Econ. Theory 27, 245–252. https://doi.org/10.1016/0022-0531(82)90029-1.

[21] Leslie, D. S., Collins, E. J., 2005. Individual q-learning in normal form games. SIAM J. Control Optim. 44, 495–514. https://doi.org/10.1137/S0363012903437976.

[22] McKelvey, R. D., Palfrey, T. R., 1995. Quantal response equilibria for normal form games. Games Econ. Behav. 10, 6–38. https://doi.org/10.1006/game.1995.1023.

[23] Roth, A. E., Erev, I., 1995. Learning in extensive-form games: experimental data and simple dynamic models in the intermediate term. Games Econ. Behav. 8, 164–212. https://doi.org/10.1016/S0899-8256(05)80020-X.

[24] Sandholm, W. H., 2012. Stochastic imitative game dynamics with committed agents. J. Econ. Theory 147, 2056–2071. https://doi.org/10.1016/j.jet.2012.05.018.

[25] Sarin, R., Vahid, F., 1999. Payoff assessments without probabilities: a simple dynamic model of choice. Games Econ. Behav. 28, 294–309. https://doi.org/10.1006/game.1998.0702.

[26] Sarin, R., Vahid, F., 2004. Strategy similarity and coordination. Econ. J. 114, 506–527. http://www.jstor.org/stable/3590293.

[27] Sawa, R., Zusai, D., 2019. Evolutionary dynamics in multitasking environments. J. Econ. Behav. Organ. 166, 288–308. https://doi.org/10.1016/j.jebo.2019.06.021.

[28] Singh, P., Sreenivasan, S., Szymanski, B. K., Korniss, G., 2012. Accelerating consensus on coevolving networks: The effect of committed individuals. Phys. Rev. E 85, 046104. https://doi.org/10.1103/PhysRevE.85.046104.

[29] Skyrms, B., Pemantle, R., 2000. A dynamic model of social network formation. Proc. Natl. Acad. Sci. U.S.A. 97, 9340–9346. https://doi.org/10.1073/pnas.97.16.9340.

[30] Tsitsiklis, J. N., 1994. Asynchronous stochastic approximation and q-learning. Mach. Learn. 16, 185–202. https://doi.org/10.1023/A:1022689125041.

[31] Yechiam, E., Busemeyer, J. R., 2005. Comparison of basic assumptions embedded in learning models for experience-based decision making. Psychon. Bull. & Rev. 12, 387–402. https://doi.org/10.3758/BF03193783.

[32] Yechiam, E., Busemeyer, J. R., 2006. The effect of foregone payoffs on underweighting small probability events. J. Behav. Dec. Making. 19, 1–16. https://doi.org/10.1002/bdm.509.