Stochastic Adaptive Learning in Dominance Solvable Games

Naoki Funai

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

March 5, 2026

Abstract

We investigate the convergence properties of a stochastic adaptive learning model that overlaps with the models of experience-weighted attraction learning and stochastic fictitious play learning. In the model, each adaptive player assigns a payoff assessment to each of her strategies, chooses a strategy with the highest randomly perturbed assessment, observes payoffs for both chosen and unchosen strategies, and updates her assessments towards the observed payoffs. In particular, we focus on the case in which adaptive players repeatedly face a fixed dominance-solvable game in pure strategies, that is, a game in which a unique strategy profile survives iterated elimination of strategies strictly dominated by pure strategies. We provide conditions under which the stochastic adaptive learning process, the sequence of players' choice probability profiles, almost surely converges to a logit quantal response equilibrium (LQRE) corresponding to the uniquely surviving strategy profile. In contrast, we show that, as in the rational-players argument, heterogeneity among adaptive players affects the almost sure convergence result. In particular, players' random choice behaviour and the rates at which they incorporate new payoff information into payoff assessments affect almost sure convergence to the surviving strategy profile. Lastly, we consider a more general learning model overlapping with payoff-assessment learning and delta learning, in which adaptive players may not observe foregone (counterfactual) payoffs, and show that the process converges to the LQRE with positive probability.
Keywords: Stochastic adaptive learning; stochastic fictitious play; experience-weighted attraction learning; payoff-assessment learning; dominance-solvable games; iterated elimination of dominated strategies; logit quantal response equilibrium

JEL classification: C72; D83

1 Introduction

If rational players face games with dominated strategies, we predict that they do not choose dominated strategies and do not expect other rational players to choose them. In addition, if "new" dominated strategies emerge after eliminating the "original" dominated strategies, rational players neither choose these new dominated strategies nor expect others to choose them. If only one strategy profile survives after the iterated elimination of dominated strategies, we use the surviving strategy profile to predict the behaviour of rational players.

For the emergence of such a strategy profile, however, we require each rational player to possess a high level of knowledge and reasoning ability. In particular, (i) players must know the game and the rationality of the other players, so that they recognise the existence of dominated strategies and understand that rational players do not choose them; (ii) given this, players expect to face the reduced game obtained by eliminating those dominated strategies, and if dominated strategies exist in the reduced game, they know that those strategies will also not be played by rational players; (iii) if the iterative procedure yields a single surviving strategy profile, rational players will choose that profile. If, for instance, (i) some player chooses a dominated strategy, (ii) some players do not know the payoff functions, or (iii) players cannot reason iteratively, then the surviving strategy profile may not provide an accurate prediction of behaviour, and we may instead require a different model or equilibrium concept.[1]

There is another perspective on the surviving strategy profile as a prediction of players' behaviour, namely the evolutionary/learning-in-games viewpoint, which does not require players to possess the high level of knowledge and reasoning power described above. Instead, through experience, players learn about the environment that they face and eventually choose the surviving strategy profile. In particular, by repeatedly playing the game over periods, they observe payoffs and/or the behaviour of other players, learn that dominated strategies are unfavourable, and choose them less often. Once dominated strategies are played less often, players face the reduced game in later periods. Again, through experience, they learn not to choose new dominated strategies in this reduced game. In the end, if only one strategy profile survives the iterated elimination of dominated strategies, players learn and end up choosing the surviving strategy profile.

One intriguing question in the evolutionary/learning-in-games viewpoint, which is not well addressed comparing with the rational players' viewpoint, concerns how heterogeneities on players' learning and choice behaviour affect the surviving strategy profile as a prediction for their long-run behaviour. For instance, in the rational players' viewpoint, if one player is not fully rational and chooses a dominated strategy, then the surviving strategy profile may not be played and may fail to provide an accurate prediction of their behaviour. In such a case, if rational players know that a bounded rational player chooses a dominated strategy, then rational players will respond to this fact and may choose a strategy profile

[1] For instance, Kreps et al. (1982) consider a player who commits to a dominated strategy and show that behaviour deviates from the surviving strategy profile in finitely repeated prisoner's dilemma games.

that differs from the surviving strategy profile. That is, rational players' viewpoint requires all players to be rational and possess high levels of knowledge and reasoning ability. In the evolutionary/learning-in-games viewpoint, we may obtain a similar consequence if we introduce heterogeneity among players or differences in how they respond to new experience. However, there exists little discussion of what types of heterogeneity among learners may affect the result. In particular, we may not know how learners' perspective on the game, the randomness in their choices, or their attitudes toward new experience affect the outcome.

In this paper, we consider the case in which adaptive players repeatedly face a fixed dominance solvable game in pure strategies, in which a unique strategy profile survives the iterated elimination of strategies strictly dominated by other pure strategies. This serves as a first step toward investigating how heterogeneity among adaptive players affects their long-run behaviour. We focus on a stochastic adaptive learning model that overlaps with experience-weighted attraction learning (EWAL hereafter; Camerer and Ho, 1999) and stochastic fictitious play learning (SFPL hereafter; Fudenberg and Kreps, 1993). In the model, in each period, each adaptive player assigns a payoff assessment to each of her actions, chooses an action with the highest randomly perturbed assessment, observes the payoffs for chosen and unchosen actions, and adjusts the assessments toward the observed payoffs. We then provide conditions under which the stochastic adaptive learning process, that is, the sequence of the choice probability profiles of adaptive players, almost surely converges to the uniquely surviving strategy profile. In particular, when each player follows the logit choice rule and each player's precision parameter is sufficiently high, we obtain almost sure convergence to the logit quantal response equilibrium (McKelvey and Palfrey, 1995) corresponding to the surviving strategy profile.

In contrast, we also investigate the case in which the conditions above are not satisfied and show that the stochastic adaptive process almost surely converges to a strategy profile that differs from the uniquely surviving strategy profile. In particular, we consider the case in which there exist three players and the precision parameter of one adaptive player is not sufficiently high and the parameters of the remaining adaptive players are sufficiently high. In this case, although the stochastic adaptive learning process almost surely converges to a logit quantal response equilibrium, the corresponding strategy profile differs from the surviving strategy profile. In particular, the adaptive players with high precision parameters choose dominated strategies with probabilities close to one, even though the adaptive player with a low precision parameter ends up choosing the dominating strategy with the highest probability.

We also consider the cases in which (i) the weight on new payoff information remains constant in each period and (ii) the randomness in players' choice decrease over periods. Case (i) represents a situation in which adaptive players put more weight on recent experience, which may be reasonable when players expect that the environment that they face

is not stationary.[2] If such a player exists, we show that the adaptive players' behaviour may fail to converge almost surely. In Case (ii), even when players behave as in Case (i), we obtain almost sure convergence to the surviving strategy profile, rather than the corresponding LQRE, as the randomness vanishes. Therefore, almost sure convergence depends on how players perceive the environment and on the level of randomness in their choices.

We also consider the case in which players may not utilise foregone (counterfactual) payoff information. In particular, we investigate a more general adaptive learning model that overlaps with payoff assessment learning and delta learning. By utilising the argument of Funai (2025), we show that the probability that the stochastic adaptive learning process converges to the LQRE corresponding to the surviving strategy profile is positive. In particular, if the assessment profile becomes close to the payoff profile of the surviving strategy profile in later periods, then the probability of convergence becomes close to one. We also obtain a convergence result when adaptive players' randomness is sufficiently high. By applying the argument in Funai (2019), we show that the stochastic adaptive learning process of the general model converges to a unique LQRE.

The argument in the evolutionary/learning-in-games literature suggests that evolutionary and learning models can substitute the high level of knowledge and deductive reasoning traditionally assumed for players to obtain the result.[3] That is, without assuming that each player knows her opponents' payoff functions, their rationality, or their knowledge of her rationality, we can show that adaptive players end up playing the surviving strategy profile. However, if there exist heterogeneities among adaptive players, we may not obtain the result even when the learning process itself converges to an equilibrium. Just as traditional game theory requires all players to possess high level of knowledge and reasoning ability, learning models also require certain behavioural properties of players in order to ensure convergence to the surviving strategy profile.

The rest of the paper is organised as follows. In Section 2, we provide a formal description of the game that adaptive players repeatedly face and specify their updating and choice rules. In Section 3, we provide conditions for almost sure convergence to the LQRE corresponding to the surviving strategy profile. In Section 4, we provide an example in which some of the conditions above are violated. In this case, we obtain almost sure convergence to an LQRE corresponding to a strategy profile at which adaptive players with high precision parameters choose dominated strategies. In Section 5, we consider the case in which players' precision parameters increase over periods and provide conditions for almost sure convergence to the surviving strategy profile. In Section 6, we consider the case in which the weight on new payoff information in the updating rule stays constant and show that the process does not almost surely converge when the precision parameters are fixed over periods but does almost surely converge when they increase. In Section 7, we investigate the convergence properties of a more general adaptive learning process in

[2] See Sarin and Vahid (1999) and Sutton and Barto (2018).

[3] See Young (1998) for instance.

which adaptive players may not obtain foregone (counterfactual) payoff information. In Section 8, we provide a brief literature review, and in Section 9, we conclude.

2 Model

2.1 Base game

We consider the situation in which, in each period $n \in \mathbb{N} = \{0, 1, 2, ...\}$, adaptive players face a fixed finite normal form game $\mathcal{G} = (\mathcal{I}, S, \pi)$, where (i) $\mathcal{I} := \{1, ..., I\}$ denotes the set of I adaptive players; (ii) $S = \times_{i \in \mathcal{I}} S_i$ denotes the set of strategy profiles, where S_i is the set of player i's strategies; and (iii) $\pi = (\pi_i)_{i \in \mathcal{I}} : S \to \mathbb{R}^I$ denotes the payoff function, where π_i is player i's payoff function, and $\pi_i(s)$ denotes player i's payoff for strategy profile $s \in S$. In particular, we also write the payoff of player i for strategy profile $s = (s_1, ..., s_I)$ as $\pi_i(s) = \pi_i(s_i, s_{-i})$ to emphasise that player i receives the payoff when she chooses strategy $s_i \in S_i$ and the other players choose $s_{-i} = (s_1, ..., s_{i-1}, s_{i+1}, ..., s_I) \in S_{-i} := \times_{j \neq i} S_j$, where S_{-i} is the set of strategy profiles of all players except player i.

As in the standard approach, we extend the domain of the payoff function to the set of mixed strategies. Let $\pi_i(x)$ denote player i's expected payoff given players' mixed strategy profile $x = (x_1, ..., x_I) \in \Delta := \times_{i \in \mathcal{I}} \Delta(S_i)$, where $\Delta(S_i) := \{x_i \in [0, 1]^{|S_i|} : \sum_{s_i \in S_i} x_{i,s_i} = 1\}$ denotes the set of mixed strategies of player i. Here, $x_i = (x_{i,s_i})_{s_i}$ denotes player i's mixed strategy, and x_{i,s_i} denotes the probability that the mixed strategy x_i assigns to strategy s_i. In particular, $\pi_i(x)$ is defined as follows: for each $i \in \mathcal{I}$ and $x \in \Delta$,

$$\pi_i(x) = \sum_{s=(s_1,...,s_I) \in S} \pi_i(s) \prod_{j \in \mathcal{I}} x_{j,s_j}.$$

This definition reflects the assumption that players' mixed strategies are independent. Let $\pi_i(s_i, x_{-i})$ denote the expected payoff of player i when she chooses s_i with probability one and her opponents follow the mixed strategy profile $x_{-i} = (x_1, ..., x_{i-1}, x_{i+1}, ..., x_I) \in \Delta_{-i} := \times_{j \neq i} \Delta(S_j)$. That is,

$$\pi_i(s_i, x_{-i}) = \sum_{s_{-i} \in S_{-i}} \pi_i(s_i, s_{-i}) \prod_{j \neq i} x_{j,s_j}.$$

In this paper, we investigate the case in which adaptive players repeatedly face a fixed dominance solvable games. In particular, we focus on games in which a unique strategy profile survives the iterated elimination of strategies that are strictly dominated by *pure* strategies, but not by non-degenerate mixed strategies. For player i, we say that $s_i \in S_i$ strictly dominates $t_i \in S_i$ (or equivalently, that t_i is strictly dominated by s_i) if $\pi_i(s_i, s_{-i}) > \pi_i(t_i, s_{-i})$ for any $s_{-i} \in S_{-i}$. We then define dominance solvable games as follows.

First, let $\mathcal{G}^0 = (\mathcal{I}, S^0, \pi^0) = (\mathcal{I}, S, \pi)$ denote the original normal form game. Suppose that in $\mathcal{G}^0$, there exists a player $i_1 \in \mathcal{I}$ who has strictly dominated strategies (by pure

strategies). Then define $\mathcal{G}^1 = (\mathcal{I}, S^1, \pi^1)$ as the game obtained by eliminating all strictly dominated strategies of player i_1. More precisely, in the game $\mathcal{G}^1$, (i) the set of players in $\mathcal{G}^1$ is the same as in $\mathcal{G}^0$; (ii) the reduced strategy profile set is given by $S^1 := \times_{i\in\mathcal{I}} S_i^1$, where S_i^1 denotes the set of strategies of player i after eliminating the strictly dominated strategies of player i_1, that is,

$$S_i^1 = \begin{cases} S_i^0 \text{ for any } i \neq i_1 \text{ and} \\ S_{i_1}^1 = S_{i_1}^0 \setminus \{t_{i_1} \in S_{i_1}^0 : t_{i_1} \text{ is strictly dominated by some } s_{i_1} \in S_{i_1}^0 \text{ in } \mathcal{G}^0\} \text{ for } i = i_1; \end{cases}$$

and (iii) the payoff function $\pi^1 : S^1 \to \mathbb{R}^I$ is a restriction of π^0 to S^1, that is, $\pi^1(s) = \pi^0(s)$ for each $s \in S^1$.

Next, for $k \geq 2$, we define the game $\mathcal{G}^k = (\mathcal{I}, S^k, \pi^k)$ recursively. Suppose that in the game $\mathcal{G}^{k-1}$, there exists a player $i_k \in \mathcal{I}$ who has strictly dominated strategies. Then let $\mathcal{G}^k = (\mathcal{I}, S^k, \pi^k)$ be the game which is obtained by eliminating all strictly dominated strategies of player i_k in the game $\mathcal{G}^{k-1}$. More precisely, in the game $\mathcal{G}^k$, (i) the set of players in $\mathcal{G}^k$ is the same as in $\mathcal{G}^{k-1}$; (ii) $S^k := \times_{i\in\mathcal{I}} S_i^k$, where S_i^k denotes the set of strategies of player i after eliminating strictly dominated strategies of player i_k in game $\mathcal{G}^{k-1}$, that is,

$$S_i^k = \begin{cases} S_i^{k-1} \text{ for any } i \neq i_k \text{ and} \\ S_{i_k}^k = S_{i_k}^{k-1} \setminus \{t_{i_k} \in S_{i_k}^{k-1} : t_{i_k} \text{ is strictly dominated by some } s_{i_k} \in S_{i_k}^{k-1} \text{ in } \mathcal{G}^{k-1}\} \\ \text{for } i = i_k; \end{cases}$$

and (iii) $\pi^k : S^k \to \mathbb{R}^I$ is a restriction of π^{k-1} to S^k, that is, $\pi^k(s) = \pi^{k-1}(s)$ for each $s \in S^k$.

Finally, for some k, if there exists no player who has strictly dominated strategies in game $\mathcal{G}^k$, then $\mathcal{G}^{k+1} = \mathcal{G}^k$ and let $S^\infty := \cap_{k=0}^\infty S^k$. If S^∞ is a singleton, then we say that game $\mathcal{G}$ is dominance solvable.

2.2 Nash equilibrium and logit quantal response equilibrium

In this section, we introduce two equilibrium concepts, Nash equilibrium and logit quantal response equilibrium (McKelvey and Palfrey, 1995; hereafter LQRE), as potential convergence targets of the stochastic adaptive learning process.

We say that $s^* = (s_i^*)_{i\in\mathcal{I}} \in S$ is a Nash equilibrium if the following condition holds: for each i and $s_i \in S_i$,

$$\pi_i(s_i^*, s_{-i}^*) \geq \pi_i(s_i, s_{-i}^*).$$

In particular, if the above inequality holds strictly, that is, if

$$\pi_i(s_i^*, s_{-i}^*) > \pi_i(s_i, s_{-i}^*)$$

for each i and $s_i \neq s_i^*$, we call $s^* \in S$ a strict Nash equilibrium.

We say that $x^* = (x^*_{i,s_i})_{i \in \mathcal{I}, s_i \in S_i} \in \Delta$ is an LQRE if the following condition holds: for each i and $s_i \in S_i$,

$$x^*_{i,s_i} = \frac{\exp(\sigma_i \pi^*_{i,s_i})}{\sum_{t_i \in S_i} \exp(\sigma_i \pi^*_{i,t_i})},$$

where (i) σ_i denotes player i's precision parameter for the choice rule and (ii) π^*_{i,s_i} denotes player i's equilibrium payoff for strategy s_i, that is,

$$\begin{aligned}\pi^*_{i,s_i} =& \pi_i(s_i, x^*_{-i}) = \sum_{s_{-i} \in S_{-i}} \pi_i(s_i, s_{-i}) x^*_{-i,s_{-i}} \\ =& \sum_{s_{-i} \in S_{-i}} \pi_i(s_i, s_{-i}) \prod_{j \neq i} \frac{\exp(\sigma_j \pi^*_{j,s_j})}{\sum_{t_j} \exp(\sigma_j \pi^*_{j,t_j})},\end{aligned}$$

where (a) $x^*_{-i} = (x^*_j)_{j \neq i}$ denotes the LQRE strategy profile except player i and (b) $x^*_{-i,s_{-i}} := \prod_{j \neq i} \frac{\exp(\sigma_j \pi^*_{j,s_j})}{\sum_{t_j} \exp(\sigma_j \pi^*_{j,t_j})}$ denotes the probability that players except player i choose a strategy profile $s_{-i} = (s_1, ..., s_{i-1}, s_{i+1}, ..., s_I)$. Note that (i) players' choices are assumed to be independent and (ii) players' precision parameters are allowed to differ.[4]

In the following sections, we investigate the case in which, when the precision parameters of adaptive players are sufficiently high, the sequence of their choice probability profiles (stochastic adaptive learning process) converges to the LQRE that corresponds to (or is closest to) a surviving strategy profile. More formally, for $\varepsilon > 0$, we say that an LQRE x^* is ε-close to a strategy profile s^* if $||x^* - s^*||_\infty < \varepsilon$, where, with slight abuse of notation, s^* also denotes the mixed strategy profile which assigns probability one to the pure strategy profile s^*. We then show that, for any $\varepsilon > 0$, the stochastic adaptive learning process almost surely converges to an LQRE which is ε-close to the surviving strategy profile when the precision parameters of adaptive players are sufficiently high.

One question which may arise is whether there exists an LQRE which is ε-close to the surviving strategy profile. Note that, for each dominance-solvable game, the surviving strategy profile is a strict Nash equilibrium. By following the argument of Funai (2025), we can show that there exists an LQRE x^* which is ε-close to the surviving strategy profile s^* for sufficiently high precision parameters.

[4] If each adaptive player chooses a strategy which has the highest perturbed assessment in the sense that for each i, s_i and $y_i \in \mathbb{R}^{|S_i|}$,

$$\begin{aligned}x_{i,s_i}(y_i) &= \mathbb{P}(\arg\max_{t_i \in S_i}(y_{t_i} + \eta_{t_i}) = s_i) \\ &= \mathbb{P}(y_{i,s_i} + \eta_{i,s_i} \geq y_{i,t_i} + \eta_{i,t_i}, \forall t_i \in S_i),\end{aligned}$$

where $\eta_i = (\eta_{i,s_i})_{s_i}$ corresponds to an i.i.d. random perturbation profile for the assessments of player i with the extreme value distribution $F_i(\eta_{i,s_i}) = \exp(-\exp(-\sigma_i \eta_{i,s_i}))$, then we obtain the choice rule.

Proposition 1. For any strict Nash equilibrium s^* and $\varepsilon > 0$, there exists an LQRE x^* which is ε-close to the strict Nash equilibrium s^* for sufficiently high $\sigma = (\sigma_i)$. In particular, for any dominance-solvable game and $\varepsilon > 0$, there exists an LQRE which is ε-close to the surviving strategy profile for sufficiently high $\sigma = (\sigma_i)$.

Proof. Utilising the proof of Proposition 3 in Funai (2025), we can show that for *any* sequence of precision parameters $\{\sigma_n = (\sigma_{n,i}) : n \in \mathbb{N}\}$ with $\sigma_{n,i} \uparrow \infty$ for each i, there exists N such that for any $n > N$, there exists an LQRE under σ_n, $x^*_{\sigma_n}$, such that $x^*_{\sigma_n} \to s^*$ as $n \to \infty$. Therefore, for any $\varepsilon > 0$, there exists $\overline{\sigma}$ such that for every $\sigma = (\sigma_i)$ such that $\sigma_i > \overline{\sigma}$ for each i, there exists an LQRE satisfying $||x^*_\sigma - s^*||_\infty < \varepsilon$.[5] □

2.3 Stochastic adaptive learning

In this section, we provide a description of adaptive players' behaviour. In particular, in each period, (i) each adaptive player assigns a subjective payoff assessment (or propensity) to each of her strategies; (ii) given the assessments, she chooses a strategy which has the highest assessment with some randomness (choice rule); and (iii) after obtaining payoff information, she updates her assessments using that information (updating rule).

Before providing a formal description of the choice and updating rules, we introduce the following notation. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space on which all random variables in this paper are defined. Let $\mathbb{N} = \{0, 1, 2, ...\}$ denote the set of non-negative integers, which represent periods in each of which adaptive players play a fixed normal form game $(\mathcal{I}, S, \pi)$. For each $n \in \mathbb{N}$, let $\mathcal{F}_n$ denote the σ-algebra which is generated by (i) the initial payoff assessment profile Q_0, (ii) the information about all the choices of adaptive players up to, but not including, period $n \in \mathbb{N}$, and (iii) each player's weighting parameters on payoff information up to period n. Formally,

$$\mathcal{F}_n := \sigma(Q_0, \mathbb{1}_{m-1,i,s_i}, \lambda_m : m \leq n, i \in \mathcal{I}, s_i \in S),$$

where for each m, (i) $Q_0 = (Q_{0,i,s_i})_{i,s_i}$ denotes the initial assessment profile, (ii) $\mathbb{1}_{m,i,s_i}$ denotes the indicator function which equals 1 if s_i is chosen in period m and 0 otherwise, and (iii) λ_m is the weighting parameter in period m.[6] Note that the sequence $\{\mathcal{F}_n\}$ forms a filtration, where $\mathcal{F}_m \subset \mathcal{F}_n$ for $m < n$.

[5] Suppose, to the contrary, that there exists $\varepsilon > 0$ such that for every $\overline{\sigma}$ there exists $\sigma > \overline{\sigma}$ ("$\sigma > \overline{\sigma}$" means $\sigma_i > \overline{\sigma}$ for each i) for which there exists no LQRE x^*_σ satisfying $||x^*_\sigma - s^*||_\infty < \varepsilon$. Then we construct a strictly increasing sequence $\{\sigma_n\}$ in the way to obtain a contradiction. First, let σ_1 be such that there is no LQRE $x^*_{\sigma_1}$ with $||x^*_{\sigma_1} - s^*||_\infty < \varepsilon$. Given σ_n, let $\overline{\sigma}_n := \sigma_n$. By assumption, we can find $\sigma_{n+1} > \overline{\sigma}_n = \sigma_n$ such that there does not exist an LQRE $x^*_{\sigma_{n+1}}$ with $||x^*_{\sigma_{n+1}} - s^*||_\infty < \varepsilon$. Thus $\{\sigma_n\}$ is strictly increasing, and for each n, there does not exist an LQRE $x^*_{\sigma_n}$ with $||x^*_{\sigma_n} - s^*||_\infty < \varepsilon$. This contradicts the fact that we can find a sequence of LQREs converging to s^* for any increasing sequence of precision parameters.

[6] The initial payoff assessment profile, the choice rule and the weighting parameter are formally defined in the following argument.

2.3.1 Payoff assessments

In each period $n \in \mathbb{N}$, each adaptive player assigns a payoff assessment to each of her strategies. For each $n \in \mathbb{N}$, $i \in \mathcal{I}$, and $s_i \in S_i$, let (i) $Q_{n,i,s_i} \in \mathbb{R}$ denote player i's assessment on strategy s_i, (ii) $Q_{n,i} = (Q_{n,i,s_i})_{s_i}$ denote player i's assessment profile in period n, and (iii) $Q_n = (Q_{n,i})_i$ denote players' assessment profile in period n. We assume that there exists a constant $M > 0$ such that $||Q_0||_\infty < M$ almost surely, where $||\cdot||_\infty$ denotes the maximum norm.[7]

2.3.2 Choice rule

Given her assessment profile, each player chooses a strategy according to the following choice rule: for each $n \in \mathbb{N}$, i, s_i and $Q_{n,i} = (Q_{n,i,t_i})_{t_i \in S_i}$,

$$x_{n,i,s_i} = \frac{\exp(\sigma_i Q_{n,i,s_i})}{\sum_{t_i \in S_i} \exp(\sigma_i Q_{n,i,t_i})}, \tag{1}$$

where (i) x_{n,i,s_i} denotes player i's choice probability of strategy s_i in period n and (ii) σ_i denotes the precision parameter for player i. We can obtain the choice rule by assuming that each player chooses a strategy with the highest randomly perturbed assessment. In particular,

$$\begin{aligned} x_{n,i,s_i} :=& \mathbb{E}[\mathbb{1}_{n,i,s_i} \mid \mathcal{F}_n] \\ =& \mathbb{P}(\{s_i \text{ is chosen in period } n\} \mid \mathcal{F}_n) \\ =& \mathbb{P}(\{Q_{n,i,s_i} + \eta_{n,i,s_i} \geq Q_{n,i,t_i} + \eta_{n,i,t_i} \text{ for each } t_i\} \mid \mathcal{F}_n), \end{aligned}$$

where η_{n,i,s_i} is a random perturbation to player i's assessment of strategy s_i in period n, which is assumed to be (i) identically distributed across strategies and periods (but not across players) and (ii) independent of all other random variables. Thus, each player chooses a strategy with the highest payoff assessment plus random noise.[8] Moreover, if η_{i,s_i} follows the extreme value distribution $F_i(\eta_{i,s_i}) = \exp\big(-\exp(-\sigma_i \eta_{i,s_i})\big)$, then each player follows the logit choice rule (1).

We also assume that in each period, players choose their strategies independently. That

[7] Under this assumption and the choice rule, each strategy is played with positive probability in each period.

[8] The perturbation η can be interpreted as a random payoff disturbance or emotional shock; see Fudenberg and Kreps (1993) and Sarin and Vahid (1999).

is, for any $\mathcal{J} \subset \mathcal{I}$ and $(s_j)_{j\in\mathcal{J}} \in \times_{j\in\mathcal{J}} S_j$,

$$\begin{aligned}
&\mathbb{P}(\{(s_j)_{j\in\mathcal{J}} \text{ is chosen in period } n\} \mid \mathcal{F}_n) \\
=&\mathbb{E}[\prod_{j\in\mathcal{J}} \mathbb{1}_{n,j,s_j} \mid \mathcal{F}_n] \\
=&\prod_{j\in\mathcal{J}} \mathbb{E}[\mathbb{1}_{n,j,s_j} \mid \mathcal{F}_n] \\
=&\prod_{j\in\mathcal{J}} x_{n,j,s_j}.
\end{aligned}$$

In particular, for each player i, the probability that the other players choose $s_{-i} = (s_1, ..., s_{i-1}, s_{i+1}, ..., s_I)$, denoted by $x_{n,-i,s_{-i}}$, is given by

$$x_{n,-i,s_{-i}} := \mathbb{E}[\mathbb{1}_{n,-i,s_{-i}} \mid \mathcal{F}_n] = \mathbb{P}(\{s_{-i} \text{ is chosen in period } n\} \mid \mathcal{F}_n) = \prod_{j\neq i} x_{n,j,s_j}.$$

Let $x_{n,i} := (x_{n,i,s_i})_{s_i\in S_i}$ denote player i's choice probability profile in period n and $x_n := (x_{n,i})_{i\in\mathcal{I}}$ denote the choice probability profile of players in period n. We call a sequence $\{x_n : n \in \mathbb{N}\}$ a stochastic adaptive learning process.

2.3.3 Updating rule

In each period, given her assessments, each player chooses a strategy according to the choice rule described above, observes payoff information, and updates her assessments using the payoff information. In particular, we consider the following updating rule for each assessment: for each n, i and s_i,

$$Q_{n+1,i,s_i} = Q_{n,i,s_i} + \lambda_{n,i}\big(\pi_{n,i,s_i} - Q_{n,i,s_i}\big), \tag{2}$$

where (i) π_{n,i,s_i} denotes the payoff information for strategy s_i toward which the assessment is adjusted and (ii) $\lambda_{n,i} \in [0, 1]$ is the weighting parameter, which describes how much the current assessment is adjusted toward the payoff.

In the following, we first specify the payoff information that each player observes for each strategy. If her opponents choose s_{-i} in period n, player i's payoff information for strategy s_i in period n is her payoff from the strategy profile (s_i, s_{-i}), that is, $\pi_{n,i,s_i} = \pi_i(s_i, s_{-i})$. Therefore,

$$\begin{aligned}
\pi_{n,i,s_i} &= \sum_{s_{-i}\in S_{-i}} \pi_i(s_i, s_{-i}) \mathbb{1}_{n,-i,s_{-i}} \\
&= \sum_{s_{-i}\in S_{-i}} \pi_i(s_i, s_{-i}) \prod_{j\neq i} \mathbb{1}_{n,j,s_j}.
\end{aligned}$$

Note that each player observes not only the payoff that she has obtained but also the payoff that she could have obtained if she had chosen other actions. That is, we consider the case in which players also observe foregone (or counterfactual) payoffs.[9]

Next, we specify the assumptions on the weighting parameters. We assume that $\lambda_{n,i}$ is an $\mathcal{F}_n$-measurable random variable satisfying the following condition:

$$\sum_n \lambda_{n,i} = \infty \quad \text{and} \quad \sum_n (\lambda_{n,i})^2 < \infty \tag{3}$$

with probability one. That is, the weight placed on the latest payoff information decreases over periods, but does not decrease so rapidly that the payoff information continues to influence players' behaviour in later periods. In Section 6, we also consider the case in which players place more weight on the latest payoff information, in particular, the case in which weighting parameter remains constant over periods.

2.3.4 Stochastic fictitious play learning and experience-weighted attraction learning

In this section, we show that the updating rule (2) overlaps the ones of stochastic fictitious play learning and experience-weighted attraction learning (EWAL). We first introduce the stochastic fictitious play learning model. In the model, each player observes the strategies chosen by her opponents and records their past play. Let $\tau_{n,-i,s_{-i}}$ denote the number of the times that strategy profile s_{-i} is chosen up to period n. Then, $\frac{\tau_{n,-i,s_{-i}}}{n}$ represents the empirical frequency of s_{-i} being chosen up to period n. The expected payoff of each strategy given this empirical distribution is therefore expressed as follows: for each n, i and s_i,

$$Q_{n,i,s_i} = \sum_{s_{-i} \in S_{-i}} \pi_i(s_i, s_{-i}) \frac{\tau_{n,-i,s_{-i}}}{n}.$$

[9] In Section 7, we also consider the case in which players may not observe foregone (counterfactual) payoffs and provide a weaker convergence result.

Note that the updating rule for the expected payoff can be expressed by the form of (2) in the following manner:[10]

$$Q_{n+1,i,s_i} = Q_{n,i,s_i} + \frac{1}{n+1}(\pi_{n,i,s_i} - Q_{n,i,s_i}),$$

where $\pi_{n,i,s_i} := \sum_{s_{-i} \in S_{-i}} \pi_i(s_i, s_{-i}) \mathbb{1}_{n,-i,s_{-i}}$.

More generally, if we let $f_{n,-i,s_{-i}}$ denote the estimate of s_{-i} being played in period n and if the expected payoff is updated according to (2), then

$$\begin{aligned}\sum_{s_{-i} \in S_{-i}} \pi_i(s_i, s_{-i}) f_{n+1,-i,s_{-i}} =& Q_{n+1,i,s_i} \\ =& \lambda_{n,i}\pi_{n,i,s_i} + (1-\lambda_{n,i}) \sum_{s_{-i} \in S_{-i}} \pi_i(s_i, s_{-i}) f_{n,-i,s_{-i}} \\ =& \sum_{s_{-i} \in S_{-i}} \pi_i(s_i, s_{-i}) \Big(\lambda_{n,i} \mathbb{1}_{n,s_{-i}} + (1-\lambda_{n,i}) f_{n,-i,s_{-i}} \Big).\end{aligned}$$

Therefore, the model coincides with the one in this paper if the estimate of each strategy is updated in the following convex combination manner:

$$f_{n+1,-i,s_{-i}} = \lambda_{n,i} \mathbb{1}_{n,-i,s_{-i}} + (1-\lambda_{n,i}) f_{n,-i,s_{-i}}.$$

Note also that if the choice probabilities of players are assumed to be (conditionally) independent, then for $s_{-i} = (s_1, ..., s_{i-1}, s_{i+1}, ..., s_N)$,

$$f_{n,-i,s_{-i}} = \prod_{j \neq i} f_{n,j,s_j}.$$

[10] In detail,

$$\begin{aligned}Q_{n+1,i,s_i} =& \sum_{s_{-i} \in S_{-i}} \pi_i(s_i, s_{-i}) \frac{\tau_{n+1,-i,s_{-i}}}{n+1} \\ =& \sum_{s_{-i} \in S_{-i}} \pi_i(s_i, s_{-i}) \Big(\frac{\tau_{n,-i,s_{-i}} + \mathbb{1}_{n,-i,s_{-i}}}{n+1} \Big) \\ =& \sum_{s_{-i} \in S_{-i}} \pi_i(s_i, s_{-i}) \Big(\frac{\tau_{n,-i,s_{-i}}}{n+1} + \frac{\mathbb{1}_{n,-i,s_{-i}}}{n+1} \Big) \\ =& \frac{n}{n+1} \Big(\sum_{s_{-i} \in S_{-i}} \pi_i(s_i, s_{-i}) \frac{\tau_{n,-i,s_{-i}}}{n} \Big) + \frac{1}{n+1} \Big(\sum_{s_{-i} \in S_{-i}} \pi_i(s_i, s_{-i}) \mathbb{1}_{n,-i,s_{-i}} \Big) \\ =& \frac{n}{n+1} Q_{n,i,s_i} + \frac{1}{n+1} \pi_{n,i,s_i} \\ =& Q_{n,i,s_i} + \frac{1}{n+1}(\pi_{n,i,s_i} - Q_{n,i,s_i}),\end{aligned}$$

Consequently,

$$f_{n+1,-i,s_{-i}} = \lambda_n \prod_{j \neq i} \mathbb{1}_{n,j,s_j} + (1-\lambda_n) \prod_{j \neq i} f_{n,j,s_j}.$$

Marginalising over all players except i and j, we obtain[11]

$$f_{n+1,j,s_j} = \lambda_n \mathbb{1}_{n,j,s_j} + (1-\lambda_n) f_{n,j,s_j}.$$

Thus, when players' choices are independent and each player's estimate is updated in the convex combination manner, the model coincide with the one in this paper.[12]

We next recall the updating rule of experience-weighted attraction learning (EWAL)(Camerer and Ho, 1999). In their model, the updating rule of the assessment (attraction) of strategy s_i is given as follows: for each n, i and s_i,

$$Q_{n+1,i,s_i} = \frac{\phi \cdot N_n \cdot Q_{n,i,s_i} + (\delta + (1-\delta)\mathbb{1}_{n,i,s_i})\pi_{n,i,s_i}}{N_{n+1}},$$

where (i) ϕ is the discount factor on the assessment in the previous period; (ii) N_n is the (discounted) number of past experiences and is updated according to $N_{n+1} = \rho N_n + 1$, where ρ is the discount factor on past experience; and (iii) δ is the discount factor on foregone (counterfactual) payoffs. Note that when we focus on the belief-based part of EWAL, corresponding to the parameter restriction $\phi = \rho = \delta = 1$, the updating rule becomes

$$\begin{aligned} Q_{n+1,i,s_i} =& \frac{n \cdot Q_{n,i,s_i} + \pi_{n,i,s_i}}{n+1} \\ =& Q_{n,i,s_i} + \frac{1}{n+1}(\pi_{n,i,s_i} - Q_{n,i,s_i}). \end{aligned}$$

Therefore, under this parameter restriction, the model in this paper coincides with the belief-based component of EWAL.

[11] In detail,

$$\begin{aligned} f_{n+1,j,s_j} =& \sum_{l \neq i,j} \sum_{s_l} f_{n+1,-i,s_{-i}} \\ =& \lambda_n \sum_{l \neq i,j} \sum_{s_l} \prod_{k \neq i} \mathbb{1}_{n,k,s_k} + (1-\lambda_n) \sum_{l \neq i,j} \sum_{s_l} \prod_{k \neq i} f_{n,k,s_k} \\ =& \lambda_n \mathbb{1}_{n,j,s_j} + (1-\lambda_n) f_{n,j,s_j}. \end{aligned}$$

[12] When $\lambda_{n,i}$ corresponds to the payoff that player i receives, the updating of the estimate coincides with that of Börgers and Sarin (1997).

3 The main result

In this section, we show that the stochastic adaptive learning process $\{x_n\}$ characterised by choice rule (1) and updating rule (2) under condition (3) almost surely converge to an LQRE corresponding to the surviving strategy profile.

Proposition 2. For any small $\varepsilon > 0$, there exists $\overline{\sigma}$ such that, for any $\sigma = (\sigma_i)$ with $\sigma_i > \overline{\sigma}$ for each i, the stochastic adaptive learning process $\{x_n\}$ characterised by choice rule (1) and updating rule (2) under condition (3) almost surely converges to an LQRE which is ε-close to the strategy profile that uniquely survives the iterated elimination of strategies strictly dominated by pure strategies.

Proof. See Appendix A. □

Here, we provide an intuitive description of the proof. First, we show that the assessments of strictly dominated strategies of player i_1 in the original game $\mathcal{G}^0 = (\mathcal{I}, S^0, \pi^0)$ become lower than the assessments of the strategies which strictly dominate them. When her precision parameter of the logit choice rule is sufficiently high, the choice probabilities of strictly dominated strategies become negligible in later periods. As a result, the game that adaptive players face in later periods is close to the reduced game obtained by eliminating strictly dominated strategies, $\mathcal{G}^1 = (\mathcal{I}, S^1, \pi^1)$. In this reduced game, by hypothesis, "new" strictly dominated strategies emerge. Applying the same argument, we obtain that, in later periods, the game that adaptive players face is close to the further reduced game $\mathcal{G}^2 = (\mathcal{I}, S^2, \pi^2)$. Since we focus on a finite game which is dominance solvable, iterating this argument finitely many times implies that, in later periods, the assessments of the surviving strategies are greater than those of the other strategies. Then when players' precision parameters are sufficiently high, the assessment profile almost surely converges to the payoff profile of the LQRE corresponding to the surviving strategy profile. Finally, because the logit choice rule is continuous, the choice probability profile also converges to the equilibrium.

Note that even when σ_i is sufficiently high for each i, strategies other than the surviving one are played with very small but positive probability. Nevertheless, because σ_i is sufficiently high, choice probabilities do no change dramatically comparing to changes in assessments. Moreover, since the surviving strategy profile is chosen much more frequently, with probability close to one and the influence of the new payoff information decreases, the assessment profile of the surviving strategy profile approaches the corresponding payoff profile.

Note also that the fact that each player's weight on new payoff information decreases over periods is important for the almost sure convergence result. If players put more weight on new payoff information even in later periods, then the choice probabilities may fail to converge almost surely.[13]

[13] See Section 6.1 for further details.

Figure 1: The payoff matrix for s_3

	s_2	t_2
s_1	$10, 10, 0$	$0, 12, 0$
t_1	$12, 0, 0$	$2, 2, 0$

Figure 2: The payoff matrix for t_3

	s_2	t_2
s_1	$10, 10, -1$	$6, 6, -1$
t_1	$6, 6, -1$	$2, 2, -1$

4 Almost sure convergence to dominated strategies

In this section, we provide an example in which adaptive players play eliminated strategies with probability close to one in the end. In particular, we consider the case in which three adaptive players repeatedly face the game with the payoff matrices given in Figures 1 and 2. The game represents the following situation. There exist two players, player 1 and player 2, who face a prisoner's dilemma game. For each $i \in \{1, 2\}$, let s_i represent cooperation and t_i represent defection. In addition, there exists another player, player 3, who has an option to intervene in the situation that players 1 and 2 face. If player 3 does not intervene in the situation, which corresponds to strategy s_3, players 1 and 2 face the prisoner's dilemma game, and player 3 neither receives a benefit nor incurs a cost, which is expressed by making her payoff 0. If player 3 decides to intervene, which is represented by strategy t_3, player 3 incurs a cost of 1 and makes the payoffs of both players 1 and player 2 equal for each case.

We first solve the problem by the iterated elimination of strategies strictly dominated by pure strategies. First, note that the payoff of s_3 is always 0 while the payoff of t_3 is always -1. Therefore, t_3 is strictly dominated by s_3. After eliminating t_3, the game that player 1 and player 2 face is the prisoner's dilemma game, and after the elimination of dominated strategies, we have a unique strategy profile (t_1, t_2, s_3). That is, player 3 does not intervene, and player 1 and player 2 face the prisoner's dilemma game, in which they both choose defection.

Now, we consider the case in which adaptive players 1, 2 and 3 face the game repeatedly and the condition in Proposition 2 does not satisfy. In particular, the precision parameter of player 3 does not satisfy the condition. Then we obtain the following result.

Proposition 3. For any small ε, there exist $\overline{\sigma}$ such that for any $\sigma_i > \overline{\sigma}$ for $i \in \{1, 2\}$, the choice probability profile of player 1 and player 2 almost surely converges to a mixed strategy profile (x_1^*, x_2^*) which is ε-close to (s_1, s_2) if $x_{3,s_3}^* < \frac{2}{3}$, where x_{3,s_3}^* is the LQRE choice probability for s_3 of player 3. In particular, the stochastic adaptive learning process converges to the LQRE (x_1^*, x_2^*, x_3^*).

Proof. See Appendix B. □

Note that since $x_{3,s_3}^* = \frac{1}{1+\exp(-\sigma_3)}$, we obtain the result above if $\sigma_3 < -\ln(\frac{1}{2}) \approx 0.69$. Since $x_{i,s_i}^* \to 1$ as $\sigma_i \to \infty$ for players 1 and 2, if the choice rule of player 3 is noisy but the ones of players 1 and 2 are not, then the stochastic adaptive learning process almost

converges to an LQRE, which is far from the one that survives the iterative elimination of strictly dominated strategies. In particular, if player 3's choice is noisy enough, players 1 and 2 end up choosing "cooperation" in the prisoner's dilemma game.

5 Almost sure convergence with increasing precision parameter

In this section, we consider the case in which each adaptive player follows the updating rule (2), but follows the logit choice rule with increasing precision parameters. That is, for each n, i and s_i,

$$Q_{n+1,i,s_i} = Q_{n,i,s_i} + \lambda_{n,i}\big(\pi_{n,i,s_i} - Q_{n,i,s_i}\big),$$

where $\{\lambda_{n,i}\}$ satisfy condition (3), and

$$x_{n,i,s_i} = \frac{\exp(\sigma_{n,i} Q_{n,i,s_i})}{\sum_{t_i \in S_i} \exp(\sigma_{n,i} Q_{n,i,t_i})}, \tag{4}$$

where $\sigma_{n,i} \to \infty$ as $n \to \infty$. Then we obtain the almost sure convergence to the surviving strategy profile, not the corresponding LQRE. That is, x_n almost surely converges to the surviving strategy profile s^* with probability one.

Proposition 4. Given the the updating rule (2) with $\{\lambda_{n,i}\}$ satisfying condition (3) and the choice rule (4), the stochastic adaptive learning process $\{x_n\}$ almost surely converges to the surviving strategy profile s^*.

Proof. See Appendix C. □

Remark. Milgrom and Roberts (1991) show a related result, but they focus on the convergence of a subsequence of a stochastic adaptive learning process which is characterised as follows. In each period, (i) with a small probability that vanishes over periods, each player experiments and chooses each of her strategies with equal probability; and (ii) when not experimenting, she picks the strategy with the highest average payoff, using only the payoffs obtained during experimentation periods. By contrast, in this paper, adaptive players can take into account payoff information in each period, and we show that the sequence of their choice probability profiles almost surely converges, which also implies that any subsequence of the process converges. Leslie and Collins (2006) obtain a similar result for potential games, which include some, but not all, dominance solvable games.

6 Almost sure convergence with constant weighting parameters

In this section, we consider cases in which the weighting parameters do not satisfy condition (3); that is, adaptive players put more weight on recent experience. In particular, we assume that there exists a scalar $\lambda \in (0,1)$ such that $\lambda_{n,i} = \lambda$ for each n and i.

In Section 6.1, we first consider the case in which each player's precision parameter in the logit choice rule is constant, as in Sections 3 and 4. We then show that almost sure convergence cannot be obtained for non-trivial 2×2 games. In Section 6.2, we next consider the case in which the precision parameter of the logit choice rule increases over periods, as in Section 5. We then show that the stochastic adaptive learning process almost surely converges to the surviving strategy profile.

6.1 Constant weighting and precision parameters

We first consider the case in which (i) there exist only two adaptive players, who face a fixed 2×2 game over periods, and (ii) the stochastic adaptive learning process $\{x_n\}$ is characterised by the choice rule (1) and the following updating rule: for each n, i and s_i,

$$Q_{n+1,i,s_i} = Q_{n,i,s_i} + \lambda\big(\pi_{n,i,s_i} - Q_{n,i,s_i}\big), \tag{5}$$

where $\lambda \in (0,1)$ is constant and fixed over periods. That is, in this section, the weighting parameters $\{\lambda_{n,i}\}$ do not satisfy condition (3).

Note that if for each player i, there exists a constant k_i such that $\pi_i(s_i, s_{-i}) - \pi_i(t_i, s_{-i}) = k_i$ for each $s_{-i} \in S_{-i}$, then we obtain

$$\begin{aligned} x_{n,i,s_i} =& \frac{1}{1 + \exp(\sigma_i(Q_{n,i,s_i} - Q_{n,i,t_i}))} \\ \to& \frac{1}{1 + \exp(\sigma_i(k_i))} \end{aligned}$$

as $n \to \infty$.[14] That is, the stochastic adaptive learning process almost surely converges.

[14] Note that

$$\begin{aligned} Q_{n+1,i,s_i} - Q_{n+1,i,t_i} =& (1-\lambda)(Q_{n,i,s_i} - Q_{n,i,t_i}) + \lambda(\pi_{n,i,s_i} - \pi_{n,i,t_i}) \\ =& (1-\lambda)^2(Q_{n-1,i,s_i} - Q_{n-1,i,t_i}) + (1-\lambda)\lambda(\pi_{n-1,i,s_i} - \pi_{n-1,i,t_i}) + \lambda(\pi_{n,i,s_i} - \pi_{n,i,t_i}) \\ =& \cdots \\ =& (1-\lambda)^{n+1}(Q_{0,i,s_i} - Q_{0,i,t_i}) + \sum_{m=1}^{n} \lambda(1-\lambda)^{n-m}(\pi_{m,i,s_i} - \pi_{m,i,t_i}) \\ =& (1-\lambda)^{n+1}(Q_{0,i,s_i} - Q_{0,i,t_i}) + (k_i)\sum_{m=1}^{n} \lambda(1-\lambda)^{n-m}, \end{aligned}$$

so $Q_{n+1,i,s_i} - Q_{n+1,i,t_i}$ almost surely converges to k_i as $n \to \infty$, since $\sum_{m=1}^{n} \lambda(1-\lambda)^{n-m} = 1 - (1-\lambda)^n$.

Except for such trivial cases, the almost sure convergence result does not hold.

Proposition 5. In 2×2 games, if there exists a player i such that $\pi_i(s_i, s_{-i}) - \pi_i(t_i, s_{-i}) \neq \pi_i(s_i, t_{-i}) - \pi_i(t_i, t_{-i})$ for strategies s_{-i} and t_{-i} of player $-i$, then the stochastic adaptive learning process $\{x_n\}$ characterised by the choice rule (1) and the updating rule (5) does not converge almost surely.

Proof. See Appendix D. □

Remark. This result shows that when players put more weight on recent experience, players' behaviour does not converge almost surely, except in trivial 2×2 games. Note that if we instead rule out randomness in adaptive players' behaviour, so that each player always chooses the strategy with the highest assessment in each period, then, as she observes foregone (counterfactual) payoffs, her assessments of dominated strategies eventually become lower than those of dominating strategies, and she stops choosing dominated strategies with probability one.[15] This observation implies that introducing randomness into adaptive players' choice behaviour affects almost sure convergence. That is, Proposition 2 may not be an immediate consequence once randomness is introduced. For almost sure convergence to the surviving strategy profile, both the way adaptive players weight recent payoff information and the degree of randomness in their choice behaviour matter.

6.2 Constant weighting parameters and increasing precision parameters

We next consider the case in which the stochastic adaptive learning process $\{x_n\}$ is characterised by the choice rule (4) and the updating rule (5). Here, we do not restrict our attention to 2×2 games, but instead focus on dominance solvable games. We then obtain the following result.

Proposition 6. The stochastic adaptive learning process characterised by the choice rule (4) and the updating rule (5) converges to the surviving strategy profile almost surely.

Proof. See Appendix E. □

7 Convergence with partial foregone payoff information

In this section, we consider the case in which players obtain partial foregone payoff information and show a convergence result. In particular, we focus on the updating rule given

[15] Note that if s_i strictly dominates t_i and $Q_{n,i,s_i} > Q_{n,i,t_i}$ in some period n, then

$$\begin{aligned} Q_{n+1,i,s_i} =& (1-\lambda) Q_{n,i,s_i} + \lambda \pi_{n,i,s_i} \\ >& (1-\lambda) Q_{n,i,t_i} + \lambda \pi_{n,i,t_i} \\ =& Q_{n+1,i,t_i}. \end{aligned}$$

by

$$Q_{n+1,i,s_i} = Q_{n,i,s_i} + \lambda_n \gamma_{n,i,s_i,t_i} \big(\pi_{n,i,s_i} - Q_{n,i,s_i}\big), \tag{6}$$

where (i) γ_{n,i,s_i,t_i} represents the discount factor on the payoff information of s_i when t_i is chosen in period n and (ii) $\lambda_n \in [0,1]$ is an $\mathcal{F}_n$-measurable random variable which satisfies condition (3) almost surely. In particular, we consider the following two cases for γ_{n,i,s_i}: (a) $\gamma_{n,i,s_i,t_i} = 1$ if $t_i = s_i$ and $\gamma_{n,i,s_i,t_i} = \gamma \in [0,1)$ otherwise; and (b) $\gamma_{n,i,s_i,t_i} = 1$ for each s_i and t_i. Case (a) represents the situation in which the foregone payoff information is discounted by γ. If $\gamma = 0$, this case corresponds to the payoff assessment learning model of Cominetti et al. (2010), Funai (2014, 2019, 2025) and Sarin and Vahid (1999). If $\gamma \in (0,1)$, it corresponds to the delta learning model of Yechiam and Busemeyer (2005, 2006). Case (b) represents the situation in which players do not discount foregone payoff information; in this case, the model corresponds to the belief-based component of EWAL and to the stochastic fictitious play learning model. Utilising the result of Funai (2025), we obtain the following result.

Proposition 7. For any small $\varepsilon > 0$, there exists $\overline{\sigma}$ such that, for $\sigma = (\sigma_i)$ with $\sigma_i > \overline{\sigma}$ for each i, the stochastic adaptive learning process characterised by the choice rule (1) and the updating rule (6) converges to the LQRE which is ε-close to the surviving strategy profile with positive probability. In particular, given that the assessment profile is close to the LQRE payoff profile, the probability of the process converging to the LQRE converges to 1 as n converges to infinity.

Proof. Note that the surviving strategy profile is a strict Nash equilibrium, and the proof follows those of Proposition 7 and Corollary 1 in Funai (2025). □

Lastly, we consider the case in which σ_i is sufficiently low for each i. Up to this point, we have obtained convergence results under the assumption that the precision parameters of players are sufficiently high. A natural question that then arises is whether we obtain a convergence result when the randomness in players' choices is relatively high. Funai (2019) provides the following answer.

Proposition 8. When σ_i is sufficiently low for each i, the stochastic adaptive learning process characterised by the choice rule (1) and the updating rule (6) converges to a unique LQRE almost surely.

Proof. See Theorem 1, Corollary 1, and Proposition 4 of Funai (2019). □

8 Brief Literature Review

The model we consider in this paper overlaps with the stochastic fictitious play learning model of Fudenberg and Kreps (1993), who show global convergence of the process in games

with a unique mixed Nash equilibrium. There also exist other global convergence results for stochastic fictitious play.[16] Benaïm and Hirsch (1999) show convergence in 2×2 games with countably many Nash equilibria. Hofbauer and Hopkins (2005) show convergence in two-player zero-sum games and partnership games. Hofbauer and Sandholm (2002) show convergence in games with an interior ESS, zero-sum games, potential games and supermodular games. In particular, the potential games that they focus on are such that, at any strategy profile, each player receives the same payoff; they show convergence for partnership games. Although the class of partnership games includes some dominance-solvable games, it does not include all dominance-solvable games.[17]

In the reinforcement learning model of Erev and Roth (1998) and Roth and Erev (1995), which the EWAL model overlaps with, Beggs (2005) shows convergence in dominance-solvable games. To show the convergence, he first shows that the choice probabilities of strictly dominated strategies converge to zero. As a result, in later periods, these strategies become negligible, and players face an almost reduced game. Applying the argument iteratively to the reduced game, and using the fact that the game is finite, the learning process converges to the surviving strategy profile. Although the argument for the main result of this paper is similar to his, an important difference is that the choice probabilities of strictly dominated strategies never become zero. That is, even in later periods when the probability of choosing the surviving strategy profile is high, the probabilities of choosing other strategies remain strictly positive and bounded away from zero. This aspect implies that the process may, in principle, move away from the surviving strategy profile. One contribution of this paper is to show that, despite the fact that the choice probabilities of strictly dominated strategies never vanish and are bounded away from zero, the learning process nevertheless converges to the LQRE corresponding to the surviving strategy profile. Moreover, Beggs (2005) does not consider heterogeneity in players' choice and learning behaviour, as in this paper, which may lead the learning process to converge to a different strategy profile in the long run.

Lastly, the evolutionary game theory literature provides several arguments for convergence to the surviving strategy profile. For instance, Hofbauer and Weibull (1996), Laraki and Mertikopoulos (2013), Nachbar (1990), and Samuelson and Zhang (1992) show that, under evolutionary dynamics, populations playing strategies which are iteratively strictly

[16] For discrete-time non-stochastic fictitious play learning, Monderer and Shapley (1996) show convergence in potential games and Milgrom and Roberts (1991) show convergence to the surviving strategy profile in finite dominance-solvable games. For continuous-time non-stochastic fictitious play learning, Viossat and Zapechelnyukshow (2013) show convergence to the surviving strategy profile in two-player dominance-solvable games.

[17] They also show convergence in the following cases: (i) for each player and each opponents' strategy profiles, the payoffs from all of her strategies are shifted in the same direction by the same amount; and (ii) for each player and each of her strategies, the payoffs for that strategy against all strategy profiles of her opponents are shifted in the same direction by the same amount. However, there exist dominance-solvable games that cannot be obtained by transforming a partnership game using the payoff transformations (i) and (ii). Moreover, they do not consider heterogeneity among players.

dominated become extinct in the long run.[18] In contrast, Hofbauer and Sandholm (2011) provide four mild conditions for deterministic evolutionary dynamics under which strictly dominated strategies fail to be eliminated in extended rock-scissors-paper games. However, these arguments do not adequately address how heterogeneity among players affects almost sure convergence. In particular, it remains unclear which conditions on (i) players' noise levels in their choice behaviour and (ii) players' weights on recent experience are sufficient to ensure convergence to the surviving strategy profile. One advantage of learning-in-games models is that they provide a more direct framework for analysing how heterogeneity among players affects their long-run behaviour: the model in this paper provides a clear threshold, which is determined by the payoff functions and the noise level of players' choice rules, under which stochastic adaptive learning process almost surely converges to (the LQRE corresponding to) the surviving strategy profile.

9 Conclusion and discussion

In this paper, we investigate the convergence properties of a stochastic adaptive learning model which overlaps with those of stochastic fictitious play and experience-weighted attraction learning in dominance-solvable games, in which a unique strategy profile survives through the iterated elimination of strategies which are strictly dominated by pure strategies.

We show that when (i) each player's precision parameter in the logit choice rule is constant over periods and sufficiently high, and (ii) the weight on payoff information in the updating rule decreases over periods, the stochastic adaptive learning process, that is, the players' choice probability profile, almost surely converges to a logit quantal response equilibrium (LQRE) corresponding to the uniquely surviving strategy profile.

When condition (i) does not hold, that is, when the precision parameter of at least one player is not sufficiently high, we still obtain almost sure convergence to an LQRE; however, the long-run behaviour may not coincide with the surviving strategy profile. In particular, in the long run, players may play dominated strategies with probability close to one.

We also consider the case in which condition (i) holds but condition (ii) does not hold. In particular, we consider the case in which the weight on payoff information in the updating rule is constant over periods. In this case, we show that the stochastic adaptive learning process does not converge almost surely in non-trivial 2×2 games. By contrast, if instead

[18] See Viossat (2015) for further discussion. Laraki and Mertikopoulos (2013) study higher-order replicator dynamics, which they justify via a learning-in-games model. In particular, they utilise a continuous-time reinforcement learning model of Erev and Roth (1998) and Roth and Erev (1995) with the following modifications: (i) players put heavier weights on earlier realised payoffs; (ii) they observe foregone payoffs; and (iii) they follow the logit choice rule. For convergence results for the discrete-time original reinforcement learning model with foregone payoffs and the logit choice rule, see Funai (2022).

of condition (i) we assume that the precision parameter increases over periods, we again obtain almost sure convergence.

Lastly, we consider the case in which adaptive players may not obtain payoff information for unchosen strategies, that is, they observe only partial foregone (counterfactual) payoff information. By utilising the results of Funai (2025), we show that the stochastic adaptive learning process converges to an LQRE corresponding to the surviving strategy profile with positive probability.

One aspect that is not fully addressed in the existing literature is that, in the presence of heterogeneity in players' choice and learning behaviour, the surviving strategy profile may not emerge as the outcome of learning. That is, just as standard game theory requires all players to possess high levels of knowledge and deductive reasoning for the prediction of a surviving strategy profile, our results require all adaptive players to satisfy certain behavioural conditions. In particular, if one player's choice behaviour is noisy or if one player puts more weight on recent experience than other players, we may not obtain almost sure convergence to the surviving strategy profile.

Acknowledgements

I would like to thank Heinrich H. Nax and audiences at Kobe University 2024 Rokko Forum and Game Theory Workshop 2025 for valuable comments. This work has been supported by JSPS KAKENHI Grants # 22H00826 and # 25K05001. I have used Chat GPT for improving readability, but have not used it for any analytical work. All remaining errors are mine.

Appendix A Proof of Proposition 2

First, we rewrite updating rule (2) in the following manner: for each i and $s_i \in S_i$,

$$\begin{aligned} Q_{n+1,i,s_i} =& Q_{n,i,s_i} + \lambda_{n,i}\Big(\sum_{s_{-i}} \pi_i(s_i, s_{-i}) \mathbb{1}_{n,-i,s_{-i}} - Q_{n,i,s_i} \Big) \\ =& Q_{n,i,s_i} + \lambda_{n,i}\Big(\sum_{s_{-i}} \pi_i(s_i, s_{-i}) x_{n,-i,s_{-i}} - Q_{n,i,s_i} + M_{n,i,s_i} \Big) \end{aligned}$$

where (i) $\mathbb{1}_{n,-i,s_{-i}}$ is the indicator function which is equal to 1 when s_{-i} is chosen in period n and 0 otherwise, (ii) M_{n,i,s_i} is a martingale difference noise such that

$$M_{n,i,s_i} := \sum_{s_{-i}} \pi_i(s_i, s_{-i}) \mathbb{1}_{n,-i,s_{-i}} - \sum_{s_{-i}} \pi_i(s_i, s_{-i}) x_{n,-i,s_{-i}}$$

and (iii) $\lambda_{n,i} \in [0,1]$ is a $\mathcal{F}_n$-measurable random variable satisfying condition (3) almost surely, that is,

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

with probability one.

Next, we consider the payoff assessment differences between dominated and dominating strategies in each step of the iterated elimination of strictly dominated strategies. In particular, without loss of generality, we consider the case in which only one dominated strategy of a player is eliminated in each step.[19]

First, we consider the original game $\mathcal{G}^0 = (\mathcal{I}, S, \pi)$, in which there exists player i_0 such that s_{i_0} strictly dominates t_{i_0}. That is, for each $s_{-i_0} \in S_{-i_0}$,

$$\pi_{i_0}(s_{i_0}, s_{-i_0}) > \pi_{i_0}(t_{i_0}, s_{-i_0}).$$

Then the updating rule of the payoff assessment difference of dominated and dominating strategies, $DQ_{n,i_0} := Q_{n,i_0,t_{i_0}} - Q_{n,i_0,s_{i_0}}$, can be expressed as follows:

$$DQ_{n+1,i_0} = DQ_{n,i_0} + \lambda_{n,i}\Big(\sum_{s_{-i_0} \in S_{-i_0}} \big(\pi_{i_0}(t_{i_0}, s_{-i_0}) - \pi_i(s_{i_0}, s_{-i_0})\big) x_{n,-i_0,s_{-i_0}} - DQ_{n,i_0} + DM_{n,i_0} \Big), \tag{7}$$

where $DM_{n,i_0} := M_{n,i_0,t_{i_0}} - M_{n,i_0,s_{i_0}}$ is still a martingale difference noise. Note that

$$\pi_{i_0}(t_{i_0}, s_{-i_0}) - \pi_{i_0}(s_{i_0}, s_{-i_0}) \le \max_{t_{-i_0}} \big(\pi_{i_0}(t_{i_0}, t_{-i_0}) - \pi_{i_0}(s_{i_0}, t_{-i_0})\big) =: -k_0,$$

where $k_0 > 0$. We can show that in later periods, the payoff assessment difference becomes smaller than $-k_0 + \varepsilon_0$ for any $\varepsilon_0 > 0$. Here, we fix $\sigma_i > 0$ for each i, but we can pick σ_i arbitrarily.

Lemma 1. For any $\varepsilon_0 > 0$, there exist N_0 such that for any $n > N_0$,

$$DQ_{n,i_0} < -k_0 + \varepsilon_0$$

almost surely.

[19] We can allow the dominated strategies of one player to be chosen in consecutive steps.

Proof. This proof follows the proof of Lemma 1 in Funai (2025). Note that

$$
\begin{aligned}
&DQ_{n+1,i_0} \\
=&(1-\lambda_{n,i_0})DQ_{n,i_0} + \lambda_{n,i_0}\Big(\sum_{s_{-i_0}}\big(\pi_{i_0}(t_{i_0},s_{-i_0})-\pi_{i_0}(s_{i_0},s_{-i_0})\big)x_{n,-i_0 s_{-i_0}}\Big) + \lambda_{n,i_0}DM_{n,i_0} \\
=&(1-\lambda_{n,i_0})\Big((1-\lambda_{n-1,i_0})DQ_{n-1,i_0} + \lambda_{n-1,i_0}\big(\sum_{s_{-i_0}}\big(\pi_{i_0}(t_{i_0},s_{-i_0})-\pi_{i_0}(s_{i_0},s_{-i_0})\big)x_{n-1,-i_0,s_{-i_0}}\big) \\
&+\lambda_{n-1,i_0}DM_{n-1,i_0}\Big) + \lambda_{n,i_0}\Big(\sum_{s_{-i}}\big(\pi_{i_0}(t_{i_0},s_{-i_0})-\pi_{i_0}(s_{i_0},s_{-i_0})\big)x_{n,-i_0,s_{-i_0}}\Big) + \lambda_{n,i_0}DM_{n,i_0} \\
=&(1-\lambda_{n,i_0})(1-\lambda_{n-1,i_0})DQ_{n-1,i_0} \\
&+\lambda_{n-1,i_0}(1-\lambda_{n,i_0})\big(\sum_{s_{-i_0}}\big(\pi_{i_0}(t_{i_0},s_{-i_0})-\pi_{i_0}(s_{i_0},s_{-i_0})\big)x_{n-1,-i_0,s_{-i_0}} \\
&+DM_{n-1,i_0}\big) + \lambda_{n,i_0}\Big(\sum_{s_{-i}}\big(\pi_{i_0}(t_{i_0},s_{-i_0})-\pi_{i_0}(s_{i_0},s_{-i_0})\big)x_{n,-i_0,s_{-i_0}} + DM_{n,i_0}\Big) \\
=&\dots \\
=&\prod_{m=0}^{n}(1-\lambda_{m,i_0})DQ_{0,i_0} \\
&+\sum_{m=1}^{n}\lambda_{m,i_0}\big(\prod_{l=m+1}^{n}(1-\lambda_{l,i_0})\big)\big(\sum_{s_{-i_0}}\big(\pi_{i_0}(t_{i_0},s_{-i_0})-\pi_i(s_{i_0},s_{-i_0})\big)x_{m,-i_0,s_{-i_0}} + DM_{m,i_0}\big),
\end{aligned}
$$

where for $m=n$, $\prod_{l=m+1}^{n}(1-\lambda_{l,i_0}) := 1$. Note that $\prod_{m=0}^{n}(1-\lambda_{m,i_0})$ converges to zero as $n\to\infty$.[20] Therefore, for any ε_0, we can take N_{01} such that for any $n > N_{01}$,

$$\prod_{m=0}^{n}(1-\lambda_{m,i_0})|DQ_{0,i_0}| < \frac{\varepsilon_0}{3}.$$

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

$$\sum_{m=1}^{n}\lambda_{m,i_0}\big(\prod_{l=m+1}^{n}(1-\lambda_{l,i_0})\big)DM_{m,i_0} \to 0$$

[20] See footnote 43 of Funai (2025).

almost surely.[21] Thus, we can take N_{02} such that for $n > N_{02}$,

$$|\sum_{m=1}^{n} \lambda_{m,i_0}(\prod_{l=m+1}^{n}(1-\lambda_{l,i_0}))DM_{m,i_0}| < \frac{\varepsilon_0}{3}.$$

Since $\pi_{i_0}(t_{i_0}, s_{-i_0}) - \pi_{i_0}(s_{i_0}, s_{-i_0}) \leq -k_0$ for each s_{-i_0},

$$\sum_{s_{-i_0}} \big(\pi_{i_0}(t_{i_0}, s_{-i_0}) - \pi_{i_0}(s_{i_0}, s_{-i_0})\big)x_{n,-i_0,s_{-i_0}} \leq -k_0$$

and thus, we can take N_{03} such that for $n > N_{03}$,

$$\sum_{m=1}^{n} \lambda_{m,i_0}(\prod_{l=m+1}^{n}(1-\lambda_{l,i_0}))\sum_{s_{-i_0}} \big(\pi_{i_0}(t_{i_0}, s_{-i_0}) - \pi_{i_0}(s_{i_0}, s_{-i_0})\big)x_{m,-i_0,s_{-i_0}} < -k_0 + \frac{\varepsilon_0}{3}.$$

Therefore, for $n > N_0 := \max\{N_{01}, N_{02}, N_{03}\}$,

$$DQ_{n,i_0} < -k_0 + \varepsilon_0.$$

□

Next, given the result above, we show that for any ε_0', we can pick σ_{0,i_0} such that the choice probability for the strictly dominated strategy becomes less than ε_0'.

Lemma 2. For strictly dominated strategy t_{i_0}, for any $\varepsilon_0' > 0$, there exist $\overline{\sigma}_0$ and N_0 such that for any $\sigma_{i_0} > \overline{\sigma}_0$ and $n > N_0$, $x_{n,i_0,t_{i_0}} < \varepsilon_0'$ almost surely.

Proof. Note that

$$\begin{aligned} x_{n,i_0,t_{i_0}} =& \frac{\exp(\sigma_{i_0}Q_{n,i_0,t_{i_0}})}{\sum_{u_{i_0}}\exp(\sigma_{i_0}Q_{n,i_0,u_{i_0}})} \\ =& \frac{1}{1+\exp(\sigma_{i_0}(Q_{n,i_0,s_{i_0}} - Q_{n,i_0,t_{i_0}})) + \sum_{u_{i_0}\neq s_{i_0},t_{i_0}}\exp(\sigma_{i_0}(Q_{n,i_0,u_{i_0}} - Q_{n,i_0,t_{i_0}}))}. \end{aligned}$$

[21] Note that we can express as follows:

$$\begin{aligned} \mathcal{DM}_{n,i_0} :=& \sum_{m=1}^{n} \lambda_{m,i_0}(\prod_{l=m+1}^{n}(1-\lambda_{l,i_0}))DM_{m,i_0} \\ =& \sum_{m=1}^{n-1} \lambda_{m,i_0}(\prod_{l=m+1}^{n}(1-\lambda_{l,i_0}))DM_{m,i_0} + \lambda_{n,i_0}DM_{n,i_0} \\ =& (1-\lambda_{n,i_0})\sum_{m=1}^{n-1} \lambda_{m,i_0}(\prod_{l=m+1}^{n-1}(1-\lambda_{l,i_0}))DM_{m,i_0} + \lambda_{n,i_0}DM_{n,i_0} \\ =& (1-\lambda_{n,i_0})\mathcal{DM}_{n-1,i_0} + \lambda_{n,i_0}DM_{n,i_0}. \end{aligned}$$

Then by Lemma 1 of Tsitsiklis (1994), we obtain the result.

By Lemma 1, we know that for randomly picked and fixed $\sigma_{i_0} > 0$, for any ε_0, there exists N_0 such that $Q_{n,i_0,s_{i_0}} - Q_{n,i_0,t_{i_0}} > k_0 - \varepsilon_0$ for $n > N_0$. We fix ε_0 such that $k_0 - \varepsilon_0 > 0$. Since for any $\sigma_{i_0} > 0$, there exist such N_0 that the payoff assessment difference is greater than $k_0 - \varepsilon_0$, if σ_{i_0} becomes greater, then x_{n,i,t_i} becomes smaller. Note that for each σ_{i_0}, there exists N_0 such that for each $n > N_0$,

$$\begin{aligned} x_{n,i_0,t_{i_0}} =& \frac{1}{1 + \exp(\sigma_{i_0}(Q_{n,i_0,s_{i_0}} - Q_{n,i_0,t_{i_0}})) + \sum_{u_{i_0} \neq s_{i_0}, t_{i_0}} \exp(\sigma_{i_0}(Q_{n,i_0,u_{i_0}} - Q_{n,i_0,t_{i_0}}))} \\ \leq& \frac{1}{1 + \exp(\sigma_{i_0}(k_0 - \varepsilon_0))}, \end{aligned}$$

where the last part converges to 0 as $\sigma_{i_0} \to \infty$. In particular, we have $x_{n,i,t_i} \to 0$ as $\sigma_{i_0} \to \infty$, that is, for any ε_0', there exists $\overline{\sigma}_0$ and N_0 such that for any $\sigma_{i_0} > \overline{\sigma}_0$ and $n > N_0$, $Q_{n,i_0,s_{i_0}} - Q_{n,i_0,t_{i_0}} > k_0 - \varepsilon_0$ and $x_{n,i,t_i} < \varepsilon_0'$. □

Next, we consider the game $\mathcal{G}^1 = (\mathcal{I}, S^1 := \times_i S_i^1, \pi^1)$, where S_i^1 denotes the set of strategies of player i after eliminating the strictly dominated strategy of i_0. That is, $S_{i_0}^1 = S_{i_0} \setminus \{t_{i_0} : t_{i_0} \text{ is strictly dominated by a pure strategy } s_{i_0}\}$ and $S_j^1 = S_j$ for $j \neq i_0$. Let π^1 denote the restriction of the payoff function π to S^1.

Then, since we focus on a dominance solvable game in pure strategy, there exists a player i_1 who has strictly dominated strategies: there exist $t_{i_1} \in S_{i_1}^1$ and $s_{i_1} \in S_{i_1}^1$ such that t_{i_1} is strictly dominated by s_{i_1}:

$$\pi_{i_1}^1(s_{i_1}, s_{-i_1}) > \pi_{i_1}^1(t_{i_1}, s_{-i_1})$$

for any $s_{-i_1} \in S_{-i_1}^1$. In particular, let $k_1 > 0$ be such that

$$-k_1 := \max_{s_{-i_1} \in S_{-i_1}^1} (\pi_{i_1}^1(t_{i_1}, s_{-i_1}) - \pi_{i_1}^1(s_{i_1}, s_{-i_1})).$$

Now for any ε_1, we take small enough ε_0' such that[22]

$$\sum_{s_{-i_1} \in S_{-i_1}} (\pi_{i_1}(t_{i_1}, s_{-i_1}) - \pi_{i_1}(s_{i_1}, s_{-i_1})) x_{n,-i_1,s_{-i_1}} < -k_1 + \frac{\varepsilon_1}{4}.$$

[22]Let $S_{-i} = \times_{j \neq i} S_j$ and $S_{-i}^1 = \times_{j \neq i} S_j^1$. Note that

$$\begin{aligned} &\sum_{s_{-i_1} \in S_{-i_1}} (\pi_{i_1}(t_{i_1}, s_{-i_1}) - \pi_{i_1}(s_{i_1}, s_{-i_1})) x_{n,-i_1,s_{-i_1}} \\ =& \sum_{s_{-i_1} \in S_{-i_1}^1} (\pi_{i_1}(t_{i_1}, s_{-i_1}) - \pi_{i_1}(s_{i_1}, s_{-i_1})) x_{n,-i_1,s_{-i_1}} + \sum_{s_{-i_1} \in S_{-i_1} \setminus S_{-i_1}^1} (\pi_{i_1}(t_{i_1}, s_{-i_1}) - \pi_{i_1}(s_{i_1}, s_{-i_1})) x_{n,-i_1,s_{-i_1}} \\ \leq& -k_1 (1 - \sum_{s_{-i_1} \in S_{-i_1} \setminus S_{-i_1}^1} x_{n,-i_1,s_{-i_1}}) + \sum_{s_{-i_1} \in S_{-i_1} \setminus S_{-i_1}^1} (\pi_{i_1}(t_{i_1}, s_{-i_1}) - \pi_{i_1}(s_{i_1}, s_{-i_1})) x_{n,-i_1,s_{-i_1}} \\ =& -k_1 + \sum_{s_{-i_1} \in S_{-i_1} \setminus S_{-i_1}^1} (k_1 + \pi_{i_1}(t_{i_1}, s_{-i_1}) - \pi_{i_1}(s_{i_1}, s_{-i_1})) x_{n,-i_1,s_{-i_1}} \end{aligned}$$

Then, letting $DQ_{n,i_1} := Q_{n,i_1,t_{i_1}} - Q_{n,i_1,s_{i_1}}$, we obtain the following result:

Lemma 3. For any $\varepsilon_1 > 0$, there exists N_1 such that for any $n > N_1$,

$$DQ_{n,i_1} < -k_1 + \varepsilon_1$$

almost surely.

Proof. Follow the proof of Lemma 1 except for the last part. Note that since we take small enough ε_0' such that

$$\sum_{s_{-i_1} \in S_{-i_1}} (\pi_{i_1}(t_{i_1}, s_{-i_1}) - \pi_{i_1}(s_{i_1}, s_{-i_1}))x_{n,-i_1,s_{-i_1}} < -k_1 + \frac{\varepsilon_1}{4},$$

we can take N_{13} such that for $n > N_{13}$,[23]

$$\sum_{m=1}^{n} \lambda_{m,i_1}\Big(\prod_{l=m+1}^{n}(1-\lambda_{l,i_1})\Big)\sum_{s_{-i_1}}\big(\pi_{i_1}(t_{i_1}, s_{-i_1}) - \pi_{i_1}(s_{i_1}, s_{-i_1})\big)x_{m,-i_1,s_{-i_1}} < -k_1 + \frac{\varepsilon_1}{3}.$$

Therefore, for $n > N_1 := \max\{N_{11}, N_{12}, N_{13}\}$,

$$DQ_{n,i_1} < -k_1 + \varepsilon_1.$$

□

Then we obtain the following result corresponding to Lemma 2.

Lemma 4. For strictly dominated strategy t_{i_1}, for any $\varepsilon_1' > 0$, there exist $\overline{\sigma}_1$ and N_1 such that for any $\sigma_{i_1} > \overline{\sigma}_1$ and $n > N_1$, $x_{n,i_1,t_{i_1}} < \varepsilon_1'$ almost surely.

Proof. Note that by Lemma 3, for $n > N_1$,

$$\begin{aligned} x_{n,i_1,t_{i_1}} =& \frac{1}{1 + \exp(\sigma_{i_1}(Q_{n,i_1,s_{i_1}} - Q_{n,i_1,t_{i_1}})) + \sum_{u_{i_1} \neq s_{i_1}, t_{i_1}} \exp(\sigma_{i_1}(Q_{n,i_1,u_{i_1}} - Q_{n,i_1,t_{i_1}}))} \\ \leq& \frac{1}{1 + \exp(\sigma_{i_1}(k_1 - \varepsilon_1))}. \end{aligned}$$

Note also that for $s_{-i_1} \in S_{-i_1} \setminus S^1_{-i_1}$,

$$x_{n,-i_1,s_{-i_1}} = x_{n,1,s_1} \cdot x_{n,2,s_2} \cdots x_{n,i_0,s_{i_0}} \cdots x_{n,i_1-1,s_{i_1-1}} \cdot x_{n,i_1+1,s_{i_1+1}} \cdots x_{n,N,s_N}$$

and $x_{n,i_0,s_{i_0}}$ is smaller than ε_0'. Note also that when $S_{-i_1} \setminus S^1_{-i_1}$ is empty, that is, $i_1 = i_0$, the inequality also holds.

[23] For m' such that $x_{m',i_0,s_{i_0}} \geq \varepsilon_0'$, we have

$$\lambda_{m',i_1}\Big(\prod_{l=m'+1}^{n}(1-\lambda_{l,i_1})\Big)\sum_{s_{-i_1}}\big(\pi_{i_1}(t_{i_1}, s_{-i_1}) - \pi_{i_1}(s_{i_1}, s_{-i_1})\big)x_{m',-i_1,s_{-i_1}} \to 0$$

as $n \to \infty$. Note also that by Lemma 2, $x_{n,i_0,s_{i_0}} < \varepsilon_0'$ for any $n > N_0$.

If we pick ε_1 small enough so that $(k_1 - \varepsilon_1)$ is positive, then the denominator in the last inequality becomes smaller as σ_{i_1} increases. □

For the remaining iterative steps, we argue by mathematical induction. Suppose that the claim holds at the K-th step of the iterated elimination of strictly dominated strategies. We then consider the $(K+1)$-th elimination step. In particular, we maintain the following induction hypotheses. Let $k_K > 0$ be such that

$$-k_K = \max_{s_{-i_K} \in S^K_{-i_K}} \big(\pi^K_{i_K}(t_{i_K}, s_{-i_K}) - \pi^K_{i_K}(s_{i_K}, s_{-i_K})\big).$$

Assumption 1. For any $\varepsilon_K > 0$, there exists N_K such that for any $n > N_K$,

$$DQ_{n,i_K} < -k_K + \varepsilon_K$$

almost surely.

Assumption 2. For strictly dominated strategy t_{i_K}, for any $\varepsilon'_K > 0$, there exist $\overline{\sigma}_K$ and N_K such that for any $\sigma_{i_K} > \overline{\sigma}_K$ and $n > N_K$,

$$x_{n,i_K,t_{i_K}} < \varepsilon'_K$$

almost surely.

We consider the game $\mathcal{G}^{K+1} = (\mathcal{I}, S^{K+1} = \times_i S_i^{K+1}, \pi^{K+1})$, where $S_i^{K+1} = S_i^K$ for $i \neq i_K$, $S_{i_K}^{K+1} = S_{i_K}^K \setminus \{t_{i_K} : t_{i_K}$ is strictly dominated by a pure strategy $s_{i_K} \in S_{i_K}^K\}$, and π^{K+1} is the restriction of payoff function π^K to S^{K+1}. Now, suppose that there exists a player i_{K+1} who has a strictly dominated strategy: there exist $t_{i_{K+1}} \in S_{i_{K+1}}^{K+1}$ and $s_{i_{K+1}} \in S_{i_{K+1}}^{K+1}$ such that $t_{i_{K+1}}$ is strictly dominated by $s_{i_{K+1}}$ in game $\mathcal{G}^{K+1}$:

$$\pi^{K+1}_{i_{K+1}}(s_{i_{K+1}}, s_{-i_{K+1}}) > \pi^{K+1}_{i_{K+1}}(t_{i_{K+1}}, s_{-i_{K+1}})$$

for any $s_{-i_{K+1}} \in S^{K+1}_{-i_{K+1}}$. In particular, let $k_{K+1} > 0$ be such that

$$-k_{K+1} := \max_{s_{-i_{K+1}} \in S^{K+1}_{-i_{K+1}}} \big(\pi^{K+1}_{i_{K+1}}(t_{i_{K+1}}, s_{-i_{K+1}}) - \pi^{K+1}_{i_{K+1}}(s_{i_{K+1}}, s_{-i_{K+1}})\big).$$

Now, let $\varepsilon'_K > 0$ be an upper bound on the choice probabilities of all strategies eliminated up to and including K-th step, chosen sufficiently small so that

$$\sum_{s_{i_{K+1}} \in S_{i_{K+1}}} (\pi_{i_{K+1}}(t_{i_{K+1}}, s_{-i_{K+1}}) - \pi_{i_{K+1}}(s_{i_{K+1}}, s_{-i_{K+1}}))x_{n,-i_{K+1}} < -k_{K+1} + \frac{\varepsilon_{K+1}}{4}$$

for small enough ε_{K+1}.[24]

Now letting $DQ_{n,i_{K+1}} := Q_{n,i_{K+1},t_{i_{K+1}}} - Q_{n,i_{K+1},s_{i_{K+1}}}$, we show the following result:

Lemma 5. For any $\varepsilon_{K+1} > 0$, there exist N_{K+1} such that for any $n > N_{K+1}$,

$$DQ_{n,i_{K+1}} < -k_{K+1} + \varepsilon_{K+1}$$

almost surely.

Proof. The argument is similar with the one of Lemma 3:

$$\begin{aligned}
DQ_{n,i_{K+1}} &= \prod_{m=0}^{n}(1-\lambda_{m,i_{K+1}})DQ_{0,i_{K+1}} \\
&+ \sum_{m=1}^{n}\lambda_{m,i_{K+1}}\Big(\prod_{l=m+1}^{n}(1-\lambda_{l,i_{K+1}})\Big)\sum_{s_{-i_{K+1}}}\Big(\pi_{i_{K+1}}(t_{i_{K+1}},s_{-i_{K+1}})-\pi_i(s_{i_{K+1}},s_{-i_{K+1}})\Big)x_{m,-i_{K+1},s_{-i_{K+1}}} \\
&+ \sum_{m=1}^{n}\lambda_{m,i_{K+1}}\Big(\prod_{l=m+1}^{n}(1-\lambda_{l,i_{K+1}})\Big)DM_{m,i_{K+1}} \\
&< -k_{K+1} + \varepsilon_{K+1}
\end{aligned}$$

□

[24]Note that

$$\begin{aligned}
&\sum_{s_{-i_{K+1}}\in S_{-i_{K+1}}}(\pi_{i_{K+1}}(t_{i_{K+1}},s_{-i_{K+1}})-\pi_{i_{K+1}}(s_{i_{K+1}},s_{-i_{K+1}}))x_{n,-i_{K+1},s_{-i_{K+1}}} \\
<& -k_{K+1}\Big(1-\sum_{s_{-i_{K+1}}\in S_{-i_{K+1}}\setminus S^{K+1}_{-i_{K+1}}}x_{n,-i_{K+1},s_{-i_{K+1}}}\Big) \\
&+\sum_{s_{-i_{K+1}}\in S_{-i_{K+1}}\setminus S^{K+1}_{-i_{K+1}}}(\pi_{i_{K+1}}(t_{i_{K+1}},s_{-i_{K+1}})-\pi_{i_{K+1}}(s_{i_{K+1}},s_{-i_{K+1}}))x_{n,-i_{K+1},s_{-i_{K+1}}} \\
<& -k_{K+1}+\sum_{s_{-i_{K+1}}\in S_{-i_{K+1}}\setminus S^{K+1}_{-i_{K+1}}}(k_{K+1}+\pi_{i_{K+1}}(t_{i_{K+1}},s_{-i_{K+1}})-\pi_{i_{K+1}}(s_{i_{K+1}},s_{-i_{K+1}}))x_{n,-i_{K+1},s_{-i_{K+1}}}.
\end{aligned}$$

Note that when $S_{-i_{K+1}} \setminus S^{K+1}_{-i_{K+1}} \neq \emptyset$, for each $s_{-i_{K+1}} \in S_{-i_{K+1}} \setminus S^{K+1}_{-i_{K+1}}$, there exists j such that $x_{n,j,s_j} \leq \varepsilon'_K$. Therefore, we take ε'_K small enough such that

$$x_{n,-i_{K+1},s_{-i_{K+1}}} = x_{n,1,s_1}\cdots x_{n,j,s_j}\cdots x_{n,N,s_N}$$

becomes so small that

$$\sum_{s_{-i_{K+1}}\in S_{-i_{K+1}}\setminus S^{K+1}_{-i_{K+1}}}(k_{K+1}+\pi_{i_{K+1}}(t_{i_{K+1}},s_{-i_{K+1}})-\pi_{i_{K+1}}(s_{i_{K+1}},s_{-i_{K+1}}))x_{n,-i_{K+1},s_{-i_{K+1}}} < \frac{\varepsilon_{K+1}}{4}.$$

Given the lemma above, we obtain the following result:

Lemma 6. For strictly dominated strategy $t_{i_{K+1}}$, for any $\varepsilon'_{K+1} > 0$, there exist $\overline{\sigma}_{K+1}$ and N_{K+1} such that for any $\sigma_{i_{K+1}} > \overline{\sigma}_{K+1}$ and $n > N_{K+1}$,

$$x_{n,i_{K+1},t_{i_{K+1}}} < \varepsilon'_{K+1}$$

almost surely.

Proof. Since the argument is the same as the one in Lemma 4, we omit the proof. □

Let s^* be the unique surviving strategy profile, which we obtain by following the iterated elimination of strictly dominated strategies and $k > 0$ be such that

$$\max_{l \in \{1,2,\ldots,K_{\max}\}} (-k_l + \varepsilon_l) < -k < 0$$

for sufficiently small (ε_l), where $K_{\max}$ denotes the total iteration time for s^* to uniquely survive.[25] Then by Lemmas 1 to 6, there exists $\overline{\sigma} := \max_K\{\overline{\sigma}_K\}$ and $N := \max_K\{N_K\}$ such that for any $n > N$ and $\sigma_i > \overline{\sigma}$ for each i,

$$Q_{n,i,s_i} - Q_{n,i,s_i^*} < -k < 0$$

for any i and $s_i \neq s_i^*$.

Let E_k be the event such that the assessment profile ends up aligning with the equilibrium payoff profile with at least k distance:

$$E_k := \{\exists N,\ \forall n > N,\ Q_{n,i,s_i} - Q_{n,i,s_i^*} < -k\ \forall i \text{ and } s_i \neq s_i^*\}.$$

Then, we have the following result.

Lemma 7. $\mathbb{P}(E_k) = 1$.

Proof. Note that if s_i is eliminated at K-th step, then at most $K_{\max} - K$ iterated eliminations of dominated strategies for player i take place for s_i^* to uniquely survive: s_i is strictly dominated by $t_i =: s_{i_K}$, which is strictly dominated by $u_i =: s_{i_{K+1}}$ at $K+1$-th game, which is strictly dominated by $v_i =: s_{i_{K+2}}$ at $K+2$-th game and so forth, and it will continue till $K_{\max}$-th game. Then

$$\begin{aligned}
&Q_{n,i,s_i} - Q_{n,i,s_i^*} \\
&= (Q_{n,i,s_i} - Q_{n,i,s_{i,K}}) + (Q_{n,i,s_{i,K}} - Q_{n,i,s_{i,K+1}}) + \cdots + (Q_{n,i,s_{i,(K_{\max}-1)}} - Q_{n,i,s_i^*}) \\
&< -(K_{\max} - K + 1)k \\
&\leq -k
\end{aligned}$$

for large enough n.[26] □

[25] We can pick such (ε_l), as $-k_l < 0$ for each l.

[26] $K_{\max} - K$ is the maximum number of iterations required for s^* to survive starting from s_i. At the other extreme, if s_i and all other remaining strategies of player i are strictly dominated by s_i^*, then $Q_{n,i,s_i} - Q_{n,i,s_i^*} \leq -k$.

Now, we focus on the event E_k, which occurs with probability one. In particular, we assume that for each player, the assessment of her unique surviving strategy is strictly greater than the assessments of her other strategies in each period. This assumption is without loss of generality: (i) since the mode of convergence we utilise is almost sure convergence, it suffices to argue sample-path-wise; and (ii) with probability one, there exists a (random) period N such that this strict inequality holds for all periods $n \geq N$. Therefore, to investigate the convergence property, it is enough to analyse the process from period N onwards. Under this assumption, by applying the argument of Tsitsiklis (1994), we obtain the almost sure convergence result.

In particular, we utilise Theorem 3 of Tsitsiklis (1994), for which we first introduce the following notation. Let $DQ_{i,s_i} := Q_{i,s_i} - Q_{i,s_i^*}$, which denotes the difference between the assessments of strategies s_i and s_i^*. Similarly, let $D\pi^*_{i,s_i} := \pi_i(s_i, x^*_{-i}) - \pi_i(s_i^*, x^*_{-i})$, which denotes the difference between the LQRE payoffs of strategies s_i and s_i^*, namely $\pi_i(s_i, x^*_{-i})$ and $\pi_i(s_i^*, x^*_{-i})$ respectively. Note that if s_i is eliminated at the l-th iteration, then

$$\begin{aligned} D\pi^*_{i,s_i} &:= \pi_i(s_i, x^*_{-i}) - \pi_i(s_i^*, x^*_{-i}) \\ &\leq (\pi_i(s_{i_l}, x^*_{-i}) - \pi_i(s_{i_{l+1}}, x^*_{-i})) + (\pi_i(s_{i_{l+1}}, x^*_{-i}) - \pi_i(s_{i_{l+2}}, x^*_{-i})) \\ &+ ... + (\pi_i(s_{i_{K_{\max}}}, x^*_{-i}) - \pi_i(s_i^*, x^*_{-i})) \\ &\leq -(K_{\max} - l + 1)k \\ &\leq -k. \end{aligned}$$

Here, s_{i_l} is player i's strategy which is eliminated at the l-th elimination stage and $-k > \max_{l \in \{1,...,K_{\max}\}}(-k_l + \varepsilon_l)$ for sufficiently small (ε_l). Moreover, we assume that $DQ_{i,s_i} \leq -k$, as stated in the preceding paragraph.

Next, to apply Theorem 3 of Tsitsiklis (1994), we find $\beta \in [0, 1)$ such that

$$||F(DQ) - D\pi^*||_\infty \leq \beta ||DQ - D\pi^*||_\infty, \tag{8}$$

where (i) $F = (F_{i,s_i})_{i,s_i \neq s_i^*} : \mathbb{R}^{\prod_i(|S_i|-1)} \to \mathbb{R}^{\prod_i(|S_i|-1)}$ is defined by $F_{i,s_i}(DQ) = \pi_i(s_i, x_{-i}) - \pi_i(s_i^*, x_{-i}) = \sum_{s_{-i}}(\pi_i(s_i, s_{-i}) - \pi_i(s_i^*, s_{-i}))x_{-i,s_{-i}}$ for each i and $s_i \neq s_i^*$; (ii) $x_{-i,s_{-i}} := \prod_{j \neq i} x_{j,s_j}$; and (iii)

$$x_{i,s_i} := \frac{\exp(\sigma_i DQ_{i,s_i})}{1 + \sum_{t_i \neq s_i^*} \exp(\sigma_i DQ_{i,t_i})}.$$

Lemma 8. There exists $\overline{\sigma}$ such that for any $\sigma = (\sigma_i)$ with $\sigma_i > \overline{\sigma}$ for each i, there exists $\beta \in [0, 1)$ such that inequality (8) holds.

Proof. For $||F(DQ) - D\pi^*||_\infty = \max_{i,s_i \neq s_i^*} |F_{i,s_i}(DQ) - D\pi^*_{i,s_i}|$, we focus on $|F_{i,s_i}(DQ) -$

$D\pi^*_{i,s_i}|$ for arbitrary i and s_i. Note that

$$\begin{aligned}|F_{i,s_i}(DQ) - D\pi^*_{i,s_i}| =& |\sum_{s_{-i}}(\pi_i(s_i, s_{-i}) - \pi_i(s^*_i, s_{-i}))(x_{-i,s_{-i}} - x^*_{-i,s_{-i}})| \\ \leq& \sum_{s_{-i}} |\pi_i(s_i, s_{-i}) - \pi_i(s^*_i, s_{-i})| \cdot |x_{-i,s_{-i}} - x^*_{-i,s_{-i}}|,\end{aligned}$$

where the difference $|x_{-i,s_{-i}} - x^*_{-i,s_{-i}}|$ can be expressed as the following telescoping sum.

$$\begin{aligned}|x_{-i,s_{-i}} - x^*_{-i,s_{-i}}| =& |\prod_{j\neq i} x_{j,s_j} - \prod_{j\neq i} x^*_{j,s_j}| \\ =& |(x_0 - x_1) + (x_1 - x_2) + \cdots + (x_{I-2} - x_{I-1})| \\ \leq& \sum_{n=0}^{I-2} |x_n - x_{n+1}|,\end{aligned}$$

where

$$\begin{aligned}x_0 &:= x_{1,s_1} x_{2,s_2} \cdots x_{i-1,s_{i-1}} x_{i+1,s_{i+1}} \cdots x_{I,s_I}, \\ x_1 &:= x^*_{1,s_1} x_{2,s_2} \cdots x_{i-1,s_{i-1}} x_{i+1,s_{i+1}} \cdots x_{I,s_I}, \\ x_2 &:= x^*_{1,s_1} x^*_{2,s_2} \cdots x_{i-1,s_{i-1}} x_{i+1,s_{i+1}} \cdots x_{I,s_I}, \\ &\cdots \\ x_{I-1} &:= x^*_{1,s_1} \cdots x^*_{i-1,s_{i-1}} x^*_{i+1,s_{i+1}} \cdots x^*_{I,s_I}.\end{aligned}$$

Note that for each n,

$$x_n - x_{n+1} = (x_{n+1,s_{n+1}} - x^*_{n+1,s_{n+1}}) \prod_{1\leq l<n+1, l\neq i} x^*_{l,s_l} \prod_{m>n+1, m\neq i} x_{m,s_m}.$$

Next, we express the difference $|x_{n+1,s_{n+1}} - x^*_{n+1,s_{n+1}}|$ as a telescoping sum. For each $k \in \{0, \ldots, |S_{n+1}|\}$, we define DQ_k as follows:

$$\begin{aligned}DQ_0 &:= (DQ_{n+1,s_{n+1}}, DQ_{n+1,t_{n+1}}, \cdots, DQ_{n+1,z_{n+1}}), \\ DQ_1 &:= (D\pi^*_{n+1,s_{n+1}}, DQ_{n+1,t_{n+1}}, \cdots, DQ_{n+1,z_{n+1}}), \\ DQ_2 &:= (D\pi^*_{n+1,s_{n+1}}, D\pi^*_{n+1,t_{n+1}}, \cdots, DQ_{n+1,z_{n+1}}), \\ &\cdots \\ DQ_{|S_{n+1}|} &:= (D\pi^*_{n+1,s_{n+1}}, D\pi^*_{n+1,t_{n+1}}, \cdots, D\pi^*_{n+1,z_{n+1}}).\end{aligned}$$

Let $j := n+1$ and the difference be $|x_{j,s_j} - x^*_{j,s_j}|$. Note that

$$\begin{aligned} x_{j,s_j} =& \frac{\exp(\sigma_j DQ_{j,s_j})}{1+\sum_{t_j \neq s_j^*} \exp(\sigma_j DQ_{j,t_j})} \quad \text{and} \\ x^*_{j,s_j} =& \frac{\exp(\sigma_j D\pi^*_{j,s_j})}{1+\sum_{t_j \neq s_j^*} \exp(\sigma_j D\pi^*_{j,t_j})}. \end{aligned}$$

Then, the difference can be expressed as the following telescoping sum:

$$\begin{aligned} |x_{j,s_j} - x^*_{j,s_j}| =& |(x_{j,s_j}(DQ_0) - x_{j,s_j}(DQ_1)) + \cdots + (x_{j,s_j}(DQ_{|S_j|-1}) - x_{j,s_j}(DQ_{|S_j|}))| \\ \leq& |x_{j,s_j}(DQ_0) - x_{j,s_j}(DQ_1)| + \cdots + |x_{j,s_j}(DQ_{|S_j|-1}) - x_{j,s_j}(DQ_{|S_j|})|, \quad (9) \end{aligned}$$

where $x_{j,s_j}(DQ) := \frac{\exp(\sigma_j DQ_{j,s_j})}{1+\sum_{s_j \neq s_j^*} \exp(\sigma_j) DQ_{j,t_j}}$.

Here, for the first term in the inequality (9), there exist $\lambda \in (0,1)$ and $\overline{DQ} = \lambda DQ_0 + (1-\lambda) DQ_1$ such that

$$|x_{j,s_j}(DQ_0) - x_{j,s_j}(DQ_1)| = |\frac{\partial x_{j,s_j}}{\partial DQ_{j,s_j}}(\overline{DQ})(DQ_{j,s_j} - D\pi^*_{j,s_j})|,$$

where

$$\begin{aligned} \frac{\partial x_{j,s_j}}{\partial DQ_{j,s_j}}(\overline{DQ}) =& \frac{\sigma_j \exp(\sigma_j \overline{DQ}_{j,s_j})(1+\sum_{t_j \neq s_j^*} \exp(\sigma_j \overline{DQ}_{j,t_j})) - \sigma_j \exp(\sigma_j \overline{DQ}_{j,s_j}) \exp(\sigma_j \overline{DQ}_{j,s_j})}{(1+\sum_{t_j \neq s_j^*} \exp(\sigma_j \overline{DQ}_{j,t_j}))^2} \\ =& \sigma_j x_{j,s_j}(\overline{DQ})(1 - x_{j,s_j}(\overline{DQ})). \end{aligned}$$

We now show that $\sigma_j x_{j,s_j}(\overline{DQ})(1-x_{j,s_j}(\overline{DQ}))$ converges to zero as $\sigma_j \to \infty$. Since $\overline{DQ}_{j,t_j} < -k$ for each t_j, we have $x_{j,s_j}(\overline{DQ}) \to 0$ as $\sigma_j \to \infty$. Note that

$$\begin{aligned} & \frac{\partial \frac{1}{x_{j,s_j}(\overline{DQ})}}{\partial \sigma_j} \\ &= \frac{(\sum_{t_j \neq s_j^*} \overline{DQ}_{j,t_j} \exp(\sigma_j \overline{DQ}_{j,t_j})) \exp(\sigma_j \overline{DQ}_{j,s_j}) - \overline{DQ}_{j,s_j} \exp(\sigma_j \overline{DQ}_{j,s_j})(1+\sum_{t_j \neq s_j^*} \exp(\sigma_j \overline{DQ}_{j,t_j}))}{(\exp(\sigma_j \overline{DQ}_{j,s_j}))^2} \\ =& \frac{(\sum_{t_j \neq s_j^*} \overline{DQ}_{j,t_j} \exp(\sigma_j \overline{DQ}_{j,t_j})) - \overline{DQ}_{j,s_j}(1+\sum_{t_j \neq s_j^*} \exp(\sigma_j \overline{DQ}_{j,t_j}))}{\exp(\sigma_j \overline{DQ}_{j,s_j})}, \end{aligned}$$

where, when $\sigma_j \to \infty$, the denominator converges to zero and the numerator converges to $-\overline{DQ}_{j,s_j} > k > 0$. Therefore, the fraction converges to infinity. Thus, by L'Hospital's rule, $\sigma_j x_{j,s_j}(\overline{DQ})(1 - x_{j,s_j}(\overline{DQ}))$ converges to zero as $\sigma_j \to \infty$.

Next, we consider the remaining terms in inequality (9), For $k \in \{1, ..., |S_j| - 1\}$, there exist $\lambda \in (0, 1)$ and $\overline{DQ} = \lambda DQ_k + (1 - \lambda)D_{k+1}$ such that

$$|x_{j,s_j}(DQ_k) - x_{j,s_j}(DQ_{k+1})| = |\frac{\partial x_{j,s_j}}{\partial DQ_{j,t_j}}(\overline{DQ})(DQ_{j,t_j} - D\pi^*_{j,t_j})|.$$

Note that

$$\begin{aligned}\frac{\partial x_{j,s_j}}{\partial DQ_{j,t_j}}(\overline{DQ}) =& \frac{-\sigma_j \exp(\sigma_j DQ_{j,s_j}) \exp(\sigma_j DQ_{j,t_j})}{(1 + \sum_{t_j \neq s^*_j} \exp(\sigma_j DQ_{j,t_j}))^2} \\ =& -\sigma_j x_{j,s_j}(\overline{DQ}) x_{j,t_j}(\overline{DQ}).\end{aligned}$$

By the same logic, $-\sigma_j x_{i,s_i}(\overline{DQ}) x_{i,t_i}(\overline{DQ})$ converges to zero as $\sigma_j \to \infty$.

Therefore, for any ε, there exists $\overline{\sigma}$ such that for any $\sigma_j > \overline{\sigma}$,

$$\begin{aligned}|x_{j,s_j} - x^*_{j,s_j}| \leq& \sum_{t_j \neq s^*_j} |\frac{\partial x_{j,s_j}}{\partial DQ_{j,t_j}}(\overline{DQ})(DQ_{j,t_j} - D\pi^*_{j,t_j})| \\ \leq& \sum_{t_j \neq s^*_j} |\frac{\partial x_{j,s_j}}{\partial DQ_{j,t_j}}(\overline{DQ})| \cdot |DQ_{j,t_j} - D\pi^*_{j,t_j}| \\ \leq& \varepsilon ||DQ - D\pi^*||_\infty.\end{aligned}$$

Thus, we have

$$\begin{aligned}|x_{-i,s_{-i}} - x^*_{-i,s_{-i}}| \leq& \sum_{n=0}^{I-2} |x_n - x_{n+1}| \\ \leq& \sum_{j \neq i} |x_{j,s_j} - x^*_{j,s_j}| \\ \leq& \varepsilon(I-1)||DQ - D\pi^*||_\infty.\end{aligned}$$

Lastly,

$$\begin{aligned}|F_{i,s_i}(DQ) - D\pi^*_{i,s_i}| =& |\sum_{s_{-i}} (\pi_i(s_i, s_{-i}) - \pi_i(s^*_i, s_{-i}))(x_{-i,s_{-i}} - x^*_{-i,s_{-i}})| \\ \leq& \sum_{s_{-i}} |\pi_i(s_i, s_{-i}) - \pi_i(s^*_i, s_{-i})| \cdot |(x_{-i,s_{-i}} - x^*_{-i,s_{-i}})| \\ \leq& \sum_{s_{-i}} |\pi_i(s_i, s_{-i}) - \pi_i(s^*_i, s_{-i})| \varepsilon(I-1)||DQ - D\pi^*||_\infty \\ \leq& K|S|\varepsilon(I-1)||DQ - D\pi^*||_\infty,\end{aligned}$$

where $K := \max_{i,s_i} \sum_{s_{-i}} |\pi_i(s_i, s_{-i}) - \pi_i(s_i^*, s_{-i})|$. Since we randomly pick i and s_i and the right-hand side of the inequality above does not depend on i and s_i, we have

$$||F(DQ) - D\pi^*||_\infty \leq K|S|\varepsilon(I-1)||DQ - D\pi^*||_\infty.$$

Thus, when we take $\sigma = (\sigma_i)$ large enough so that $\varepsilon < \frac{1}{K|S|(I-1)}$, $\beta := K|S|\varepsilon(I-1)$ becomes less than 1. □

Lastly, we show that the stochastic adaptive learning process almost surely converges to the LQRE corresponding to the surviving strategy profile. Note that since $D\pi^*_{i,s_i} = \pi_i(s_i, x^*_{-i}) - \pi_i(s_i^*, x^*_{-i}) < -k$ for any i and $s_i \neq s_i^*$, we have $x^*_{i,s_i} < \varepsilon$ for sufficiently large σ_i.[27] That is, $||x^* - s^*||_\infty < \varepsilon$ for sufficiently large $\sigma = (\sigma_i)$.

Proposition. For any sufficiently small $\varepsilon > 0$, there exists $\overline{\sigma}$ such that for any $\sigma = (\sigma_i)$ with $\sigma_i > \overline{\sigma}$ for each i, the stochastic adaptive learning process almost surely converges to the LQRE which is ε-close to the strategy profile which uniquely survives the iterated elimination of strategies strictly dominated by pure strategies.

Proof. We have shown that with probability one, there is N such that for any $n > N$, inequality (8) holds. Then by following the argument of Tsitsiklis (1994), the proof of Theorem 3 in particular, we obtain the result. Below, we replicate the proof of Tsitsiklis (1994) along with the argument in this paper.

First, as in the proof, we focus on the process $\{DQ'_n = (DQ'_{n,i,s_i})\}$, where $DQ'_{n,i,s_i} := DQ_{n,i,s_i} - D\pi^*_{i,s_i}$. We know that there exists B_0 such that $||DQ'_n||_\infty \leq B_0$ for any n, as the initial assessments and payoffs are assumed to be bounded and the new assessment of each strategy is a convex combination of the observed payoff and the assessment of the previous period. Note that

$$\begin{aligned}
DQ'_{n+1,i,s_i} =& DQ_{n+1,i,s_i} - D\pi^*_{i,s_i} \\
=& (1-\lambda_{n,i})(DQ_{n,i,s_i} - D\pi^*_{i,s_i}) + \lambda_{n,i}(\pi_{n,i,s_i} - D\pi^*_{i,s_i}) \\
=& (DQ_{n,i,s_i} - D\pi^*_{i,s_i}) + \lambda_{n,i}\big((\pi_{n,i,s_i} - D\pi^*_{i,s_i}) - (DQ_{n,i,s_i} - D\pi^*_{i,s_i})\big) \\
=& DQ'_{n,i,s_i} + \lambda_{n,i}\big((\overline{\pi}_{n,i,s_i} - D\pi^*_{i,s_i}) - DQ'_{n,i,s_i} + w_{n,i,s_i}\big),
\end{aligned}$$

where $\overline{\pi}_{n,i,s_i} := \mathbb{E}[\pi_{n,i,s_i} \mid \mathcal{F}_n]$ and $w_{n,i,s_i} := \pi_{n,i,s_i} - \overline{\pi}_{n,i,s_i}$, which is a martingale difference noise.

[27] Note that

$$\begin{aligned}
x^*_{i,s_i} =& \frac{\exp(\sigma_i D\pi^*_{i,s_i})}{1 + \sum_{t_i \neq s_i, s_i^*} \exp(\sigma_i D\pi^*_{i,t_i})} \\
\leq& \exp(\sigma_i(-k)) \to 0
\end{aligned}$$

as $\sigma_i \to \infty$.

Fix $\varepsilon > 0$ such that $\beta(1+2\varepsilon) < 1$ and for each k, let B_k be such that

$$B_{k+1} = \beta(1+2\varepsilon)B_k,\ k \geq 0.$$

Note that $\{B_k\}$ converges to zero as $k \to \infty$.

Suppose that there exists n_k such that $||DQ'_n||_\infty \leq B_k$ for all $n \geq n_k$. Here, we utilise the result we have obtained, so that we let $n_k > N$, where N is the one such that inequality (8) holds. Then we show that there exists $n_{k+1} \geq n_k$ such that $||DQ'_n||_\infty \leq B_{k+1}$ for all $n \geq n_{k+1}$, which completes the proof of $DQ'_n = DQ_n - D\pi^*$ converging to zero.

Let W_{n,i,s_i} be such that

$$W_{n+1,i,s_i} = (1-\lambda_{n,i})W_{n,i,s_i} + \lambda_{n,i}w_{n,i,s_i}.$$

Then we have $W_{n,i,s_i} \to 0$ as $n \to \infty$.

Now, for any period n_0, we define $W_{n_0|n_0,i,s_i} = 0$ and for $n \geq n_0$,

$$W_{n+1|n_0,i,s_i} = (1-\lambda_{n,i})W_{n|n_0,i,s_i} + \lambda_{n,i}w_{n,i,s_i}.$$

Then, for any $\delta > 0$, there exists some M such that $|W_{n+1|m,i,s_i}| \leq \delta$ for all $n+1 > m \geq M$.[28]

[28] See Lemma 2 of Tsitsiklis (1994). Note that

$$\begin{aligned}
W_{n+1|n_0,i,s_i} =& (1-\lambda_{n,i})W_{n|n_0,i,s_i} + \lambda_{n,i}w_{n,i,s_i} \\
=& (1-\lambda_{n,i})W_{n|n_0,i,s_i} + W_{n+1|n,i,s_i} \\
=& (1-\lambda_{n,i})(1-\lambda_{n-1,i})W_{n-1|n_0,i,s_i} + (1-\lambda_{n,i})\lambda_{n-1,i}w_{n-1,i,s_i} + \lambda_{n,i}w_{n,i,s_i} \\
=& \prod_{l=n-1}^{n}(1-\lambda_{l,i})W_{n-1|n_0,i,s_i} + (1-\lambda_{n,i})W_{n|n-1,i,s_i} + W_{n+1|n,i,s_i} \\
=& \prod_{l=n-1}^{n}(1-\lambda_{l,i})W_{n-1|n_0,i,s_i} + W_{n+1|n-1,i,s_i} \\
=& \prod_{l=n-2}^{n}(1-\lambda_{l,i})W_{n-2|n_0,i,s_i} + (1-\lambda_{n-1,i})W_{n-1|n-2,i,s_i} + W_{n+1|n-1,i,s_i} \\
=& \prod_{l=n-2}^{n}(1-\lambda_{l,i})W_{n-2|n_0,i,s_i} + W_{n+1|n-2,i,s_i} \\
=& \cdots \\
=& \prod_{l=m}^{n}(1-\lambda_{l,i})W_{m|n_0,i,s_i} + W_{n+1|m,i,s_i},
\end{aligned}$$

and for $n_0 = 0$, we have

$$|W_{n+1|m,i,s_i}| \leq |W_{n+1|0,i,s_i}| + |W_{m|0,i,s_i}|.$$

Notice that the two terms on the right-hand side of the inequality above converge to zero as m and $n+1$ converge to infinity, as $W_{n+1|0,i,s_i} = W_{n+1,i,s_i}$ and $W_{m|0,i,s_i} = W_{m,i,s_i}$.

Let $\nu_k \geq n_k$ be such that for $n \geq \nu_k \geq n_k$, $|W_{n|\nu_k,i,s_i}| \leq \beta\varepsilon B_k$ and $||DQ'_n||_\infty \leq B_k$. Let $Y_{\nu_k,i,s_i} = B_k$ and

$$Y_{n+1,i,s_i} = (1-\lambda_{n,i})Y_{n,i,s_i} + \lambda_{n,i}\beta B_k$$

for $n \geq \nu_k$.

Lemma 9.

$$-Y_{n,i,s_i} + W_{n|\nu_k,i,s_i} \leq DQ'_{n,i,s_i} \leq Y_{n,i,s_i} + W_{n|\nu_k,i,s_i}$$

Proof. Suppose that the inequality above holds for some n. Then

$$\begin{aligned} DQ'_{n+1,i,s_i} =& (1-\lambda_{n,i})DQ'_{n,i,s_i} + \lambda_{n,i}(\overline{\pi}_{n,i,s_i} - D\pi^*_{i,s_i} + w_{n,i,s_i}) \\ \leq& (1-\lambda_{n,i})(Y_{n,i,s_i} + W_{n|\nu_k,i,s_i}) + \lambda_{n,i}\beta B_k + \lambda_{n,i}w_{n,i,s_i} \\ =& Y_{n+1,i,s_i} + W_{n+1|\nu_k,i,s_i} \end{aligned}$$

A symmetrical argument yields $-Y_{n+1,i,s_i} + W_{n+1|\nu_k,i,s_i} \leq DQ'_{n+1,i,s_i}$:

$$\begin{aligned} DQ'_{n+1,i,s_i} =& (1-\lambda_{n,i})DQ'_{n,i,s_i} + \lambda_{n,i}(\overline{\pi}_{n,i,s_i} - D\pi^*_{i,s_i} + w_{n,i,s_i}) \\ \geq& (1-\lambda_{n,i})(-Y_{n,i,s_i} + W_{n|\nu_k,i,s_i}) - \lambda_{n,i}\beta B_k + \lambda_{n,i}w_{n,i,s_i} \\ =& -Y_{n+1,i,s_i} + W_{n+1|\nu_k,i,s_i} \end{aligned}$$

□

Now, we know that Y_{n,i,s_i} converges to βB_k as $n \to \infty$ and thus,

$$\limsup_{n\to\infty} ||DQ'_{n,i,s_i}||_\infty \leq \beta(1+\varepsilon)B_k < B_{k+1},$$

which completes the proof. □

Appendix B Proof for Proposition 3

We first show the almost sure convergence of player 3's assessment profile and, consequently, the convergence of her choice probability profile.

Lemma 10. The assessment profile of player 3, $\{Q_{n,3} = (Q_{n,3,s_3}, Q_{n,3,t_3})\}$, almost surely converges to $(0,-1)$. In addition, the choice probability profile of player 3, $\{(x_{n,3,s_3}, x_{n,3,t_3}) = (\frac{\exp(\sigma_3 Q_{n,3,s_3})}{\exp(\sigma_3 Q_{n,3,s_3})+\exp(\sigma_3 Q_{n,3,t_3})}, \frac{\exp(\sigma_3 Q_{n,3,t_3})}{\exp(\sigma_3 Q_{n,3,s_3})+\exp(\sigma_3 Q_{n,3,t_3})})\}$, almost surely converges to $(x^*_{3,s_3}, x^*_{3,t_3}) = (\frac{\exp(\sigma_3(0))}{\exp(\sigma_3(0))+\exp(\sigma_3(-1))}, \frac{\exp(\sigma_3(-1))}{\exp(\sigma_3(0))+\exp(\sigma_3(-1))})$.

Proof. Note that player 3 receives a payoff of 0 from strategy s_3 and a payoff of -1 from t_3, regardless of which strategy profile players 1 and 2 choose. Therefore, the assessments of s_3 and t_3 converge to 0 and -1, respectively. Since the logit choice rule is continuous with respect to the assessment profile, the corresponding choice probabilities also converge. □

Since the event that player 3's assessment profile converges occurs with probability one, we proceed with the argument conditional on this event.

Next, we focus on the convergence of the assessment profiles of player 1 and player 2. Note that for each $i \in \{1, 2\}$,

$$\begin{aligned}
Q_{n+1,i,s_i} =& Q_{n,i,s_i} + \lambda_{n,i}\Big(\sum_{s_{-i},s_3} \pi_i(s_i, s_{-i}, s_3)\mathbb{1}_{n,-i,s_{-i}}\mathbb{1}_{n,3,s_3} - Q_{n,i,s_i}\Big) \\
=& Q_{n,i,s_i} + \lambda_{n,i}\Big(\sum_{s_{-i},s_3} \pi_i(s_i, s_{-i}, s_3)x_{n,-i,s_{-i}}x_{n,3,s_3} - Q_{n,i,s_i} + M_{n,i,s_i}\Big) \\
=& Q_{n,i,s_i} + \lambda_{n,i}\Big(\sum_{s_{-i},s_3} \pi_i(s_i, s_{-i}, s_3)x_{n,-i,s_{-i}}x^*_{3,s_3} - Q_{n,i,s_i} + M_{n,i,s_i} + \eta_{n,i,s_i}\Big),
\end{aligned}$$

where

$$\begin{aligned}
M_{n,i,s_i} :=& \sum_{s_{-i},s_3} \pi_i(s_i, s_{-i}, s_3)\mathbb{1}_{n,-i,s_{-i}}\mathbb{1}_{n,3,s_3} - \sum_{s_{-i}} \pi_i(s_i, s_{-i}, s_3)x_{n,-i,s_{-i}}x_{n,3,s_3} \text{ and} \\
\eta_{n,i,s_i} :=& \sum_{s_{-i}} \pi_i(s_i, s_{-i}, s_3)x_{n,-i,s_{-i}}x_{n,3,s_3} - \sum_{s_{-i}} \pi_i(s_i, s_{-i}, s_3)x_{n,-i,s_{-i}}x^*_{3,s_3}.
\end{aligned}$$

Note that M_{n,i,s_i} is a martingale difference noise and η_{n,i,s_i} converges to zero with probability one, as $x_{n,3,s_3}$ almost surely converges to x^*_{3,s_3}. Note that

$$\begin{aligned}
Q_{n+1,i,s_i} =& Q_{n,i,s_i} + \lambda_{n,i}\Big(\sum_{s_{-i},s_3} \pi_i(s_i, s_{-i})x_{n,-i,s_{-i}}x^*_{3,s_3} - Q_{n,i,s_i} + M_{n,i,s_i} + \eta_{n,i,s_i}\Big) \\
=& Q_{n,i,s_i} + \lambda_{n,i}\Big(\sum_{s_{-i}} \pi_i(s_i, s_{-i}, x^*_3)x_{n,-i,s_{-i}} - Q_{n,i,s_i} + M_{n,i,s_i} + \eta_{n,i,s_i}\Big),
\end{aligned}$$

where

$$\pi_i(s_i, s_{-i}, x^*_3) := \sum_{s_3} \pi_i(s_i, s_{-i}, s_3)x^*_{3,s_3}.$$

Then, we can utilise the analysis of the two-player game between players 1 and 2 with payoff function $\pi_i(s_i, s_{-i}, x^*_3)$, where the corresponding payoff matrix is shown in Figure 3.

Notice that if

$$\begin{aligned}
&10 > 12x^*_{3,s_3} + 6(1 - x^*_{3,s_3}) = 6x^*_{3,s_3} + 6 \text{ and} \\
&6(1 - x^*_{3,s_3}) > 2,
\end{aligned}$$

Figure 3: s_3

	s_2	t_2
s_1	$10, 10$	$6x^*_{3,t_3}, 12x^*_{3,s_3} + 6x^*_{3,t_3}$
t_1	$12x^*_{3,s_3} + 6x^*_{3,t_3}, 6x^*_{3,t_3}$	$2, 2$

that is, if $x^*_{3,s_3} < \frac{2}{3}$, s_i strictly dominates t_i for each $i \in \{1, 2\}$. Therefore, the result follows from Appendix B of Funai (2025) and Proposition 2 of this paper, even with an additional noise term that converges to zero almost surely.

Appendix C Proof for Proposition 4

We first show the following lemma.

Lemma 11. If s_i strictly dominates t_i, then $x_{n,i,t_i} \to 0$ almost surely.

Proof. Note that for $DQ_{n,i} := Q_{n+1,i,s_i} - Q_{n+1,i,t_i}$,

$$\begin{aligned} DQ_{n+1,i} &= DQ_{n,i} + \lambda_{n,i}\big((\pi_{n,i,s_i} - \pi_{n,i,t_i}) - DQ_{n,i}\big) \\ &= \prod_{m=0}^{n}(1-\lambda_{m,i})DQ_{0,i} + \sum_{m=1}^{n}\lambda_{m,i}\Big(\prod_{l=m+1}^{n}(1-\lambda_{l,i})\Big)\big((\pi_{l,i,s_i} - \pi_{l,i,t_i})\big) \end{aligned}$$

and

$$\begin{aligned} (\pi_{n,i,s_i} - \pi_{n,i,t_i}) &= \sum_{s_{-i}}(\pi_i(s_i, s_{-i}) - \pi_i(t_i, s_{-i}))\mathbb{1}_{n,-i,s_{-i}} \\ &\geq \min_{s_{-i}\in S_{-i}}(\pi_i(s_i, s_{-i}) - \pi_i(t_i, s_{-i})) =: m_{s_i,t_i}, \end{aligned}$$

where $m_{s_i,t_i} > 0$ as s_i strictly dominates t_i. Therefore, for any $\varepsilon > 0$ small enough so that $m_{s_i,t_i} - \varepsilon > 0$, there exists N such that for any $n > N$,

$$DQ_{n,i} \geq m_{s_i,t_i} - \varepsilon > 0$$

almost surely. This implies that

$$\begin{aligned} x_{n,i,t_i} &= \frac{\exp(\sigma_{n,i}Q_{n,i,t_i})}{\sum_{u_i\in S_i}\exp(\sigma_{n,i}Q_{u_i})} \\ &= \frac{1}{1 + \exp(\sigma_{n,i}(Q_{n,i,s_i} - Q_{n,i,t_i})) + \sum_{u_i\neq s_i,t_i}\exp(\sigma_{n,i}(Q_{u_i} - Q_{t_i}))} \\ &\to 0 \end{aligned}$$

as $n \to \infty$, since for large n,

$$\exp(\sigma_{n,i}(Q_{n,i,s_i} - Q_{n,i,t_i})) \geq \exp(\sigma_{n,i}(m_{s_i,t_i} - \varepsilon)),$$

and $\exp(\sigma_{n,i}(m_{s_i,t_i} - \varepsilon)) \to \infty$ as n and $\sigma_{n,i}$ diverge. □

Using Lemma 11, we prove the claim by mathematical induction. Let s_i be a strategy which strictly dominates t_i at the k-th step of the iterated elimination of strictly dominated strategies. Recall that

$$DQ_{n+1,i} = DQ_{n,i} + \lambda_{n,i}\big((\pi_{n,i,s_i} - \pi_{n,i,t_i}) - DQ_{n,i}\big).$$

Here, the difference $\pi_{n,i,s_i} - \pi_{n,i,t_i}$ can be expressed as follows.

$$\begin{aligned}
&\pi_{n,i,s_i} - \pi_{n,i,t_i} = \sum_{s_{-i} \in S_{-i}} \pi_i(s_i, s_{-i}) - \pi_i(t_i, s_{-i}) \mathbb{1}_{n,-i,s_{-i}} \\
&= \sum_{s_{-i} \in S_{-i} \setminus S_{k,-i}} \pi_i(s_i, s_{-i}) - \pi_i(t_i, s_{-i}) \mathbb{1}_{n,-i,s_{-i}} + \sum_{s_{-i} \in S_{k,-i}} \pi_i(s_i, s_{-i}) - \pi_i(t_i, s_{-i}) \mathbb{1}_{n,-i,s_{-i}},
\end{aligned}$$

where $S_{k,-i}$ is the set of surviving strategy profiles of players except i at k-th step of iterated elimination of strictly dominated strategies. We now show that for any small $\varepsilon > 0$ such that $m_{k,s_i,t_i} - \varepsilon > 0$, there exists N such that for any $n > N$, we have

$$DQ_{n,i} > m_{k,s_i,t_i} - \varepsilon > 0$$

almost surely, where $m_{k,s_i,t_i} := \min_{s_{-i} \in S_{k,-i}} (\pi_i(s_i, s_{-i}) - \pi_i(t_i, s_{-i})) > 0$. Note that

$$\begin{aligned}
&DQ_{n+1,i} - DQ_{n,i} = \lambda_{n,i}\big((\pi_{n,i,s_i} - \pi_{n,i,t_i}) - DQ_{n,i}\big) \\
&= \lambda_{n,i}\Big(\sum_{s_{-i} \in S_{k,-i}} (\pi_i(s_i, s_{-i}) - \pi_i(t_i, s_{-i})) \mathbb{1}_{n,-i,s_{-i}} + \sum_{s_{-i} \in S_{-i} \setminus S_{k,-i}} (\pi_i(s_i, s_{-i}) - \pi_i(t_i, s_{-i})) \mathbb{1}_{n,-i,s_{-i}} - DQ_{n,i} \Big) \\
&= \lambda_{n,i}\Big(\sum_{s_{-i} \in S_{k,-i}} (\pi_i(s_i, s_{-i}) - \pi_i(t_i, s_{-i})) x_{n,-i,s_{-i}} - DQ_{n,i} + DM_{n,i,s_i,t_i} + \eta_{n,i,s_i,t_i} \Big),
\end{aligned}$$

where

$$DM_{n,i,s_i,t_i} := \sum_{s_{-i} \in S_{-i}} (\pi_i(s_i, s_{-i}) - \pi_i(t_i, s_{-i})) \mathbb{1}_{n,-i,s_{-i}} - \sum_{s_{-i} \in S_{-i}} (\pi_i(s_i, s_{-i}) - \pi_i(t_i, s_{-i})) x_{n,-i,s_{-i}}$$

and

$$\eta_{n,i,s_i,t_i} := \sum_{s_{-i} \in S_{-i} \setminus S_{k,-i}} (\pi_i(s_i, s_{-i}) - \pi_i(t_i, s_{-i})) x_{n,-i,s_{-i}},$$

which converges to 0 almost surely by the mathematical induction hypothesis. Therefore,

$$DQ_{n+1,i} = \prod_{m=0}^{n}(1-\lambda_{m,i})DQ_{0,i}$$
$$+\sum_{m=1}^{n}\lambda_{m,i}(\prod_{l=m+1}^{n}(1-\lambda_{l,i}))(\sum_{s_{-i}\in S_k}\big(\pi_i(s_i,s_{-i})-\pi_i(t_i,s_{-i})\big)x_{m,-i,s_{-i}}+DM_{m,i,s_i,t_i}+\eta_{m,i,s_i,t_i}).$$

By a similar argument in Lemma 1, we have

$$DQ_{n,i} > m_{i,s_i,t_i} - \varepsilon > 0$$

for large enough n.[29] Again, by the same argument as in Lemma 11, we have $x_{n,i,t_i} \to 0$ almost surely as $n \to \infty$ and $\sigma_{n,i} \to \infty$.

Appendix D Proof for Proposition 5

Note that if $\pi_i(s_i,s_{-i}) - \pi_i(t_i,s_{-i}) = k$ and $\pi_i(s_i,t_{-i}) - \pi_i(t_i,t_{-i}) = l$ for some i, $s_{-i} \neq t_{-i}$ and $0 < k < l$, then for any $n > 0$ and $DQ_{n,i} := Q_{n,i,s_i} - Q_{n,i,t_i}$, the probabilities of $DQ_{n,i}$ approaching k and l are positive. This means that $DQ_{n,i}$ does not almost surely converge. Since x_{n,i,s_i} is strictly increasing with respect to $DQ_{n,i}$, x_{n,i,s_i} does not also converge almost surely.

In the following, we provide a more rigorous argument, we prove by contradiction in particular. The intuition is as follows. Suppose that $DQ_{n,i}$ almost surely converges to a random variable DQ. Then, for any $\varepsilon > 0$, there exists N such that for each $n > N$, $|DQ_{n,i} - DQ| < \varepsilon$. (i) If $DQ > \frac{k+l}{2}$ and $DQ_{n,i}$ is in the ε-neighbourhood of DQ in period n, then the probability that $DQ_{n+d,i} < k+\varepsilon$ is $\eta > 0$, where (a) d is the maximum number of consecutive periods in which s_{-i} is chosen so that $DQ_{n,i}$ becomes ε-close to k, and (b) η is the probability that s_{-i} is chosen consecutively for d periods. This contradicts almost sure convergence. (ii) If $DQ \leq \frac{k+l}{2}$ and $DQ_{n,i}$ is in the ε-neighbourhood of DQ in period n, then the probability that $DQ_{n+d',i} > l - \varepsilon$ is $\eta' > 0$, where (c) d' is the maximum number of consecutive periods in which t_{-i} is chosen so that $DQ_{n,i}$ becomes ε-close to l, and (d)

[29] Note that

$$\prod_{m=0}^{n}(1-\lambda_{m,i})D_{0,i},\ \sum_{m=1}^{n}\lambda_{m,i}(\prod_{l=m+1}^{n}(1-\lambda_{l,i}))DM_{m,i,s_i,t_i}\ \text{ and }\ \sum_{m=1}^{n}\lambda_{m,i}(\prod_{l=m+1}^{n}(1-\lambda_{l,i}))\eta_{m,i,s_i,t_i}$$

converge to zero almost surely as $n \to \infty$. Also,

$$\sum_{m=1}^{n}\lambda_{m,i}(\prod_{l=m+1}^{n}(1-\lambda_{l,i}))(\sum_{s_{-i}\in S_{k,-i}}\big(\pi_i(s_i,s_{-i})-\pi_i(t_i,s_{-i})\big)x_{m,-i,s_{-i}}$$

becomes greater than $m_{i,s_i,t_i} - \varepsilon$ as $n \to \infty$.

η' is the probability that t_{-i} is chosen consecutively for d periods. This also contradicts almost sure convergence.

Let $DQ_{n,i} := Q_{n,i,s_i} - Q_{n,i,t_i}$ and DQ be a random variable to which $DQ_{n,i}$ almost surely converges. Then we assume that

$$\mathbb{P}(\{DQ_{n,i} \to DQ\}) = \mathbb{P}(\cap_{c=1}^{\infty} \cup_{N=1}^{\infty} \cap_{n=N}^{\infty} \{\omega \in \Omega : |DQ_{n,i}(\omega) - DQ(\omega)| < \frac{1}{c}\}) = 1.$$

Since for each c,

$$\cap_{c=1}^{\infty} \cup_{N=1}^{\infty} \cap_{n=N}^{\infty} \{\omega \in \Omega : |DQ_{n,i}(\omega) - DQ(\omega)| < \frac{1}{c}\} \subset \cup_N \cap_{n=N}^{\infty} \{\omega \in \Omega : |DQ_{n,i}(\omega) - DQ(\omega)| < \frac{1}{c}\},$$

we have

$$\mathbb{P}(\cup_{N=1}^{\infty} \cap_{n=N}^{\infty} \{\omega \in \Omega : |DQ_{n,i}(\omega) - DQ(\omega)| < \frac{1}{c}\}) = 1. \tag{10}$$

In the following argument, we provide a contradictory argument for equation (10) for small enough $\frac{1}{c}$, which means that the assumption that $DQ_{n,i}$ almost surely converges to DQ is false.

To do so, we consider the partition $\{\{DQ(\omega) > \frac{k+l}{2}\}, \{DQ(\omega) \leq \frac{k+l}{2}\}\}$ and express equation (10) as follows:

$$\begin{aligned}
&\mathbb{P}\big(\cup_N \cap_{n=N}^{\infty} (\{\omega \in \Omega : |DQ_{n,i}(\omega) - DQ(\omega)| < \frac{1}{c}\} \cap (\{DQ(\omega) > \frac{k+l}{2}\})\big) \\
&+ \mathbb{P}\big(\cup_N \cap_{n=N}^{\infty} (\{\omega \in \Omega : |DQ_{n,i}(\omega) - DQ(\omega)| < \frac{1}{c}\} \cap (\{DQ(\omega) \leq \frac{k+l}{2}\})\big) \\
&= 1.
\end{aligned}$$

Then for small enough $\frac{1}{c}$, we derive a contradiction.

Before doing so, we note two points. First, note that for each n and ω,

$$\begin{aligned}
DQ_{n,i}(\omega) \in &(\min\{\min_{s_{-i}}(\pi_i(s_i, s_{-i}) - \pi_i(t_i, s_{-i})), Q_{0,i,s_i} - Q_{0,i,t_i}\}, \\
&\max\{\max_{s_{-i}}(\pi_i(s_i, s_{-i}) - \pi_i(t_i, s_{-i})), Q_{0,i,s_i} - Q_{0,i,t_i}\}) \\
&=: (m, M).
\end{aligned}$$

Since $DQ_{n,i}$ is a convex combination of $DQ_{n-1,i}$ and the payoff difference

$$(\pi_{n-1,i,s_i} - \pi_{n-1,i,t_i}) \in \big(\min_{s_{-i}}(\pi_i(s_i, s_{-i}) - \pi_i(t_i, s_{-i})), \max_{s_{-i}}(\pi_i(s_i, s_{-i}) - \pi_i(t_i, s_{-i}))\big),$$

we have $\mathbb{P}(\{DQ_{n,i}(\omega) \in \mathbb{R}\}) = \mathbb{P}(\{DQ_{n,i}(\omega) \in (m, M)\}) = 1$.

Second, note that (i) $DQ_{n,i}$ approaches k if only s_{-i} is chosen from period n onwards, and (ii) for any ε, the number of periods that $DQ_{n,i}$ reaches ε-neighbourhood of k depends

on the value of $DQ_{n,i}$ but not on n itself, as the weighting parameter is fixed and constant over periods.[30] Then, for any n and $DQ_{n,i}$, the number of periods that $DQ_{n,i}$ reaches within ε-neighbourhood of k is the largest when $DQ_{n,i} = M$ if $|M-k| > |m-k|$ and $DQ_{n,i} = m$ if $|M-k| \leq |m-k|$.

Let $DQ_{n,i} = M$ for some n and $n' > n$ be such that $DQ_{n',i} < k+\varepsilon$ when only s_{-i} is chosen after period n.[31] Note that (i) the number of periods reaching the neighbourhood, $d := (n'-n)$, does not depend on n itself, as λ is constant over periods, and (ii) for any $DQ_{n,i} \in (m, M)$, $DQ_{n',i} < k+\varepsilon$ when only s_{-i} is chosen from period n to period n'. Then let $\eta := (\overline{x}_{s_{-i}})^d$, where $\overline{x}_{s_{-i}}$ is the smallest probability of s_{-i} being chosen in each period.[32] Utilising the same argument, we let d' be the number of periods during which only t_{-i} is chosen so that $DQ_{n,i}$ reaches ε-neighbourhood of l, and $\eta' := (\overline{x}_{t_{-i}})^{d'}$.

We are now ready to go back to show a contradiction. We first consider the case in which

$$\mathbb{P}\big(\cup_{N=1}^{\infty} \cap_{n=N}^{\infty} \{\omega \in \Omega : |DQ_{n,i}(\omega) - DQ(\omega)| < \frac{1}{c}\} \cap \{DQ(\omega) > \frac{k+l}{2}\}\big) > 0.$$

[30]Note that for any $n' > n$,

$$\begin{aligned} DQ_{n',i} =& (1-\lambda)(DQ_{n'-1,i}) + \lambda(\pi_{n'-1,i,s_i} - \pi_{n'-1,i,t_i}) \\ =& (1-\lambda)^2(DQ_{n'-2,i}) + \lambda(1-\lambda)(\pi_{n'-2,i,s_i} - \pi_{n'-2,i,t_i}) + \lambda(\pi_{n'-1,i,s_i} - \pi_{n'-1,i,t_i}) \\ =& \cdots \\ =& (1-\lambda)^{(n'-n)}(DQ_{n,i}) + \sum_{k=0}^{(n'-n)-1} \lambda(1-\lambda)^k(\pi_{n'-k-1,i,s_i} - \pi_{n'-k-1,i,t_i}) \end{aligned}$$

which approaches k when only s_{-i} is chosen after period n. Note that when only s_{-i} is chosen after period n and approaches k, the number of periods that $DQ_{n',i}$ reaches the ε-neighbourhood of k depends on (i) the difference between n' and n, $n'-n$, and (ii) the value of $DQ_{n,i}$. However, note that it does not depend on n itself, as λ does not depend on n. In other words, if $DQ_{n,i} = DQ_{m,i}$ for some $n \neq m$, the number of periods that $DQ_{n,i}$ and $DQ_{m,i}$ reach the ε-neighbourhood of k when only s_{-i} is chosen after period n and m, $n'-n$ and $m'-m$, are the same.

[31]When $DQ_{n,i} = m$, as $m < k+\varepsilon$, $DQ_{n,i} < k+\varepsilon$.

[32]Note that

$$\mathbb{P}(\{s_{-i} \text{ is chosen in period } n\} \mid \mathcal{F}_n) = \prod_{j \neq i} x_{n,j,s_j}.$$

Note also that for each j,

$$\begin{aligned} x_{n,j,s_j} =& \frac{\exp(\sigma_j Q_{n,j,s_j})}{\sum_{u_j} \exp(\sigma_j Q_{n,j,u_j})} \\ \geq& \overline{x}_{j,s_j} := \frac{1}{1 + \sum_{s_j \neq t_j} \exp(\sigma_j \overline{D}_{n,j,s_j,t_j})} > 0, \end{aligned}$$

where $\overline{D}_{n,j,s_j,t_j} := (M-n)$. Then $\overline{x}_{s_{-i}} := \prod_{j \neq i} \overline{x}_{j,s_j}$.

Then since

$$\begin{aligned}&\cup_{N=1}^{\infty}\cap_{n=N}^{\infty}\{\omega\in\Omega:|DQ_{n,i}(\omega)-DQ(\omega)|<\frac{1}{c}\}\cap\{DQ(\omega)>\frac{k+l}{2}\}\\&\subset\cup_{N=1}^{\infty}\cap_{n=N}^{\infty}\{\omega\in\Omega:DQ_{n,i}(\omega)>\frac{k+l}{2}-\frac{1}{c}\},\end{aligned}$$

we have

$$\mathbb{P}\big(\cup_{N=1}^{\infty}\cap_{n=N}^{\infty}\{\omega\in\Omega:DQ_{n,i}(\omega)>\frac{k+l}{2}-\frac{1}{c}\}\big)>0.$$

Note that there exists N such that[33]

$$p_N:=\mathbb{P}\big(\cap_{n=N}^{\infty}(\{\omega\in\Omega:DQ_{n,i}(\omega)>\frac{k+l}{2}-\frac{1}{c}\}\big)>0,$$

[33]Such N exists, otherwise,

$$\begin{aligned}&\mathbb{P}\big(\cup_{N=1}^{\infty}\cap_{n=N}^{\infty}(\{\omega\in\Omega:DQ_{n,i}(\omega)>\frac{k+l}{2}-\frac{1}{c}\})\\&\leq\sum_{N=1}^{\infty}\mathbb{P}\big(\cap_{n=N}^{\infty}(\{\omega\in\Omega:DQ_{n,i}(\omega)>\frac{k+l}{2}-\frac{1}{c}\})\\&=0,\end{aligned}$$

which contradicts with the hypothesis.

and for each $n > N' \geq N$, we have[34]

$$\begin{aligned}
&\mathbb{P}(\{DQ_{n+d,i}(\omega) < k + \frac{1}{c}\} \cap \cap_{m=N'}^{n}\{DQ_{m,i}(\omega) > \frac{k+l}{2} - \frac{1}{c}\}) \\
&= \mathbb{P}(\{DQ_{n+d,i}(\omega) < k + \frac{1}{c}\} \mid \cap_{m=N'}^{n}\{DQ_{m,i}(\omega) > \frac{k+l}{2} - \frac{1}{c}\}) \\
&\times \mathbb{P}(\cap_{m=N'}^{n}\{DQ_{m,i}(\omega) > \frac{k+l}{2} - \frac{1}{c}\}) \\
&\geq \eta p_N \\
&> 0.
\end{aligned}$$

[34]Note that for each n,

$$\begin{aligned}
&\mathbb{P}(\{DQ_{n+d,i} < k + \frac{1}{c}\} \mid \mathcal{F}_n) \geq \mathbb{P}(\{\text{Only } s_{-i} \text{ is chosen from } n \text{ to } n+d\} \mid \mathcal{F}_n) \geq \eta, \\
&\mathbb{E}[\mathbb{P}(DQ_{n+d,i} < k + \frac{1}{c} \mid \mathcal{F}_n)] = \mathbb{P}(DQ_{n+d,i} < k + \frac{1}{c}) \geq \eta, \\
&\mathbb{P}(DQ_{n+d',i} > l - \frac{1}{c} \mid \mathcal{F}_n) \geq \mathbb{P}(\{\text{Only } t_{-i} \text{ is chosen from} n \text{ to } n+d'\} \mid \mathcal{F}_n) \geq \eta', \text{ and} \\
&\mathbb{E}[\mathbb{P}(DQ_{n+d',i} > l - \frac{1}{c} \mid \mathcal{F}_n)] = \mathbb{P}(DQ_{n+d',i} > l - \frac{1}{c}) \geq \eta'.
\end{aligned}$$

Note also that for each $A_n \in \mathcal{F}_n$ such that $\mathbb{P}(A_n) > 0$,

$$\mathbb{E}[\mathbb{1}_{A_n}\mathbb{P}(\{DQ_{n+d,i} < k + \frac{1}{c}\} \mid \mathcal{F}_n)] = \mathbb{P}(A_n \cap \{DQ_{n+d,i} < k + \frac{1}{c}\}) \geq \mathbb{P}(A_n)\eta,$$

where the last inequality holds due to the monotonicity of the expectation. Therefore,

$$\mathbb{P}(\{DQ_{n+d,i} < k + \frac{1}{c})\} \mid A_n) = \frac{\mathbb{P}(A_n \cap \{DQ_{n+d,i} < k + \frac{1}{c}\})}{\mathbb{P}(A_n)} \geq \eta.$$

This implies that for N,[35]

$$\mathbb{P}\big(\cap_{n=N}^{\infty}(\{\omega\in\Omega: DQ_{n,i}(\omega)>\frac{k+l}{2}-\frac{1}{c}\})\big)\leq\lim_{M\to\infty}(1-\eta)^M=0,$$

which contradicts the hypothesis.

Next, we consider the remaining case, in which

$$\mathbb{P}\big(\cup_{N=1}^{\infty}\cap_{n=N}^{\infty}\{\omega\in\Omega:|DQ_{n,i}(\omega)-DQ(\omega)|<\frac{1}{c}\}\cap\{DQ(\omega)>\frac{k+l}{2}\}\big)=0,$$

and thus

$$\mathbb{P}\big(\cup_{N=1}^{\infty}\cap_{n=N}^{\infty}\{\omega\in\Omega:|DQ_{n,i}(\omega)-D(\omega)|<\frac{1}{c}\}\cap\{DQ(\omega)\leq\frac{k+l}{2}\}\big)=1.$$

Then, since

$$\begin{aligned}&\cup_{N=1}^{\infty}\cap_{n=N}^{\infty}\{\omega\in\Omega:|DQ_{n,i}(\omega)-D(\omega)|<\frac{1}{c}\}\cap\{DQ(\omega)\leq\frac{k+l}{2}\}\\&\subset\cup_{N=1}^{\infty}\cap_{n=N}^{\infty}\{\omega\in\Omega:DQ_{n,i}(\omega)\leq\frac{k+l}{2}+\frac{1}{c}\},\end{aligned}$$

[35]In detail, note that

$$\begin{aligned}&\mathbb{P}\big(\cap_{n=N}^{\infty}(\{\omega\in\Omega:DQ_{n,i}(\omega)>\frac{k+l}{2}-\frac{1}{c}\})\\=&\lim_{N'\to\infty}\mathbb{P}\big(\cap_{n=N}^{N'}(\{\omega\in\Omega:DQ_{n,i}(\omega)>\frac{k+l}{2}-\frac{1}{c}\})\\\leq&\lim_{M\to\infty}\mathbb{P}\big(\cap_{m=0}^{M}(\{\omega\in\Omega:DQ_{N+md,i}(\omega)>\frac{k+l}{2}-\frac{1}{c}\})\\\leq&\lim_{M\to\infty}\mathbb{P}\big((\{\omega\in\Omega:DQ_{N+Md,i}(\omega)>\frac{k+l}{2}-\frac{1}{c}\}\mid\cap_{m=0}^{M-1}\{\omega\in\Omega:DQ_{N+md,i}(\omega)>\frac{k+l}{2}-\frac{1}{c}\})\\&\times\mathbb{P}\big((\{\omega\in\Omega:DQ_{N+(M-1)d,i}(\omega)>\frac{k+l}{2}-\frac{1}{c}\}\mid\cap_{m=0}^{M-2}\{\omega\in\Omega:DQ_{N+md,i}(\omega)>\frac{k+l}{2}-\frac{1}{c}\})\\&\cdots\\&\times\mathbb{P}\big((\{\omega\in\Omega:DQ_{N+d,i}(\omega)>\frac{k+l}{2}-\frac{1}{c}\}\mid\{\omega\in\Omega:DQ_{N,i}(\omega)>\frac{k+l}{2}-\frac{1}{c}\})\\&\times\mathbb{P}(\{\omega\in\Omega:DQ_{N,i}(\omega)>\frac{k+l}{2}-\frac{1}{c}\})\\\leq&\lim_{M\to\infty}(1-\eta)^M\\=&0,\end{aligned}$$

Note also that for small enough $\frac{1}{c}$ and $A_n\in\mathcal{F}_n$,

$$\begin{aligned}&\mathbb{P}(\{DQ_{n+d,i}(\omega)>\frac{k+l}{2}-\frac{1}{c}\}\mid A_n\})\\\leq&\mathbb{P}(\{DQ_{n+d,i}(\omega)\geq k+\frac{1}{c}\}\mid A_n\})\\\leq&(1-\eta).\end{aligned}$$

we have

$$\mathbb{P}\big(\cup_{N=1}^{\infty}\cap_{n=N}^{\infty}(\{\omega\in\Omega: DQ_{n,i}(\omega)\leq\frac{k+l}{2}+\frac{1}{c}\}\big)>0.$$

Note that for some N,

$$p'_N:=\mathbb{P}\big(\cap_{n=N}^{\infty}(\{\omega\in\Omega: D_{n,i}(\omega)\leq\frac{k+l}{2}+\frac{1}{c}\}\big)>0,$$

and thus we have[36]

$$\mathbb{P}\big(\cap_{n=N}^{\infty}(\{\omega\in\Omega: DQ_{n,i}(\omega)\leq\frac{k+l}{2}+\frac{1}{c}\}\big)\leq\lim_{M\to\infty}(1-\eta')^M=0,$$

Therefore, for each case, we have a contradiction.

Appendix E Proof for Proposition 6

We first show that the choice probability of a strictly dominated strategy converges to zero almost surely. Then we follow the same procedure as before to show that the stochastic

[36]In detail,

$$\begin{aligned}
&\mathbb{P}\big(\cap_{n=N}^{\infty}(\{\omega\in\Omega: DQ_{n,i}(\omega)\leq\frac{k+l}{2}+\frac{1}{c}\}\big)\\
=&\lim_{N'\to\infty}\mathbb{P}\big(\cap_{n=N}^{N'}(\{\omega\in\Omega: DQ_{n,i}(\omega)\leq\frac{k+l}{2}+\frac{1}{c}\}\big)\\
\leq&\lim_{M\to\infty}\mathbb{P}\big(\cap_{m=0}^{M}(\{\omega\in\Omega: DQ_{N+md',i}(\omega)\leq\frac{k+l}{2}+\frac{1}{c}\}\big)\\
\leq&\lim_{M\to\infty}\mathbb{P}\big((\{\omega\in\Omega: DQ_{N+Md',i}(\omega)\leq\frac{k+l}{2}+\frac{1}{c}\}\mid\cap_{m=0}^{M-1}\{\omega\in\Omega: DQ_{N+md',i}(\omega)\leq\frac{k+l}{2}+\frac{1}{c}\}\big)\\
&\times\mathbb{P}\big((\{\omega\in\Omega: DQ_{N+(M-1)d',i}(\omega)\leq\frac{k+l}{2}+\frac{1}{c}\}\mid\cap_{m=0}^{M-2}\{\omega\in\Omega: DQ_{N+md',i}(\omega)\leq\frac{k+l}{2}+\frac{1}{c}\}\big)\\
&\cdots\\
&\times\mathbb{P}\big((\{\omega\in\Omega: DQ_{N+d',i}(\omega)\leq\frac{k+l}{2}+\frac{1}{c}\}\mid\{\omega\in\Omega: DQ_{N,i}(\omega)\leq\frac{k+l}{2}+\frac{1}{c}\}\big)\\
&\times\mathbb{P}(\{\omega\in\Omega: DQ_{N,i}(\omega)\leq\frac{k+l}{2}+\frac{1}{c}\})\\
\leq&\lim_{M\to\infty}(1-\eta')^M\\
=&0.
\end{aligned}$$

Note that for small enough $\frac{1}{c}$,

$$\begin{aligned}
&\mathbb{P}(\{DQ_{n+d',i}(\omega)\leq\frac{k+l}{2}+\frac{1}{c}\}\mid\cap_{m=N'}^{n}\{DQ_{m,i}(\omega)\leq\frac{k+l}{2}+\frac{1}{c}\}\})\\
\leq&\mathbb{P}(\{DQ_{n+d',i}(\omega)\leq l-\frac{1}{c}\}\mid\cap_{m=N'}^{n}\{DQ_{m,i}(\omega)\leq\frac{k+l}{2}-\frac{1}{c}\}\})\\
\leq&(1-\eta').
\end{aligned}$$

adaptive learning process converges to the surviving strategy profile.

Lemma 12. For each strictly dominated strategy t_i, $x_{n,i,t_i} \to 0$ almost surely.

Proof. Let t_i be strictly dominated by s_i. That is,

$$\pi_i(t_i, s_{-i}) < \pi_i(s_i, s_{-i})$$

for each $s_{-i} \in S_{-i}$. Note that

$$\begin{aligned} Q_{n+1,i,s_i} =& (1-\lambda)Q_{n,i,s_i} + \lambda\pi_{n,i,s_i} \\ =& (1-\lambda)^2 Q_{n-1,i,s_i} + (1-\lambda)\lambda\pi_{n-1,i,s_i} + \lambda\pi_{n,i,s_i} \\ =& \cdots \\ =& (1-\lambda)^{n+1} Q_{0,i,s_i} + \sum_{m=1}^{n} \lambda(1-\lambda)^{n-m}\pi_{m,i,s_i}. \end{aligned}$$

Note that $(1-\lambda)^{n+1}Q_{0,i,s_i} \to 0$ as $n \to \infty$. Note also that

$$\begin{aligned} \pi_{n,i,s_i} - \pi_{n,i,t_i} =& \sum_{s_{-i}\in S_{-i}} \pi_i(s_i, s_{-i})\mathbb{1}_{n,-i,s_{-i}} - \sum_{s_{-i}\in S_{-i}} \pi_i(t_i, s_{-i})\mathbb{1}_{n,-i,s_{-i}} \\ =& \sum_{s_{-i}\in S_{-i}} \big(\pi_i(s_i, s_{-i}) - \pi_i(t_i, s_{-i})\big)\mathbb{1}_{n,-i,s_{-i}} \\ \geq& D_{s_i,t_i} \end{aligned}$$

where $D_{s_i,t_i} := \min_{s_{-i}\in S_{-i}}(\pi_i(s_i, s_{-i}) - \pi_i(t_i, s_{-i})) > 0$. Therefore, for any small $\varepsilon > 0$ such that $D_{s_i,t_i} - \varepsilon > 0$, there exists N such that for any $n > N$,

$$\begin{aligned} Q_{n+1,i,s_i} - Q_{n+1,i,t_i} =& (1-\lambda)^{n+1}(Q_{0,i,s_i} - Q_{0,i,t_i}) + \sum_{m=1}^{n} \lambda(1-\lambda)^{n-m}(\pi_{m,i,s_i} - \pi_{m,i,t_i}) \\ \geq& D_{s_i,t_i} - \varepsilon > 0. \end{aligned}$$

Therefore,

$$\begin{aligned} x_{n,i,t_i} =& \frac{\exp(\sigma_{n,i}Q_{n,i,t_i})}{\sum_{u_i}\exp(\sigma_{n,i}Q_{n,i,u_i})} \\ =& \frac{1}{1 + \exp(\sigma_{n,i}(Q_{n,i,s_i} - Q_{n,i,t_i})) + \sum_{u_i\neq s_i,t_i}\exp(\sigma_{n,i}(Q_{n,i,u_i} - Q_{n,i,t_i}))} \\ \leq& \frac{1}{\exp(\sigma_{n,i}(D_{s_i,t_i} - \varepsilon))} \\ \to& 0 \end{aligned}$$

as $n \to \infty$.[37] □

Then, by following the argument in Proposition 4, we obtain the result.

References

[1] Beggs, A. W., 2005. On the convergence of reinforcement learning. J. Econ. Theory 122, 1–36.

[2] 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.

[3] Börgers, T., Sarin, R., 1997. Learning Through Reinforcement and Replicator Dynamics. J. Econ. Theory 77, 1–14. https://doi.org/10.1006/jeth.1997.2319.

[4] 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

[5] 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.

[6] 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.

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

[8] 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.

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

[10] 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.

[37] Note that as $\exp(x) > 0$ for $x \in \mathbb{R}$, $1 + \exp(\sigma_{n,i}(Q_{n,i,s_i} - Q_{n,i,t_i})) + \sum_{u_i \neq s_i,t_i} \exp(\sigma_{n,i}(Q_{n,i,u_i} - Q_{n,i,t_i})) > \exp(\sigma_{n,i}(D_{s_i,t_i} - \varepsilon))$, that is, $x_{n,i,t_i} < \frac{1}{\exp(\sigma_{n,i}(D_{s_i,t_i} - \varepsilon))}$. Since $D_{s_i,t_i} - \varepsilon > 0$, the right-hand side converges to 0 as $\sigma_{n,i} \to \infty$.

[11] Funai, N., 2025. Stochastic adaptive learning with committed players in games with strict Nash equilibria, Games Econ. Behav. 154, 351–376. https://doi.org/10.1016/j.geb.2025.09.011

[12] Hofbauer, J., Hopkins, E., 2005. Learning in perturbed asymmetric games. Games Econ. Behav. 52, 133–152.

[13] Hofbauer, J., Sandholm, W.H., 2002. On the global convergence of stochastic fictitious play. Econometrica 70, 2265–2294. https://doi.org/10.1111/j.1468-0262.2002.00440.x

[14] Hofbauer, J., Sandholm, W. H., 2011. Survival of dominated strategies under evolutionary dynamics. Theor. Econ. 6, 341–377. https://doi.org/10.3982/TE771

[15] Hofbauer, J., Weibull, J.W., 1996. Evolutionary selection against dominated strategies. J. Econ. Theory 71, 558–573. https://doi.org/10.1006/jeth.1996.0133.

[16] 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.

[17] Laraki R., Mertikopoulos, P., 2013. Higher order game dynamics. J. Econ. Theory 148, 2666–2695. https://doi.org/10.1016/j.jet.2013.08.002.

[18] Leslie, D. S., Collins, E. J., 2006. Generalised weakened fictitious play. Games Econ. Behav. 56, 285–298. https://doi.org/10.1016/j.geb.2005.08.005.

[19] 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.

[20] Milgrom, P., Roberts, J., 1991. Adaptive and sophisticated learning in normal form games. Games Econ. Behav. 3, 82–100. https://doi.org/10.1016/0899-8256(91)90006-Z.

[21] Monderer, D., Shapley, L. S., 1996. Fictitious play property for games with identical interests. J. Econ. Theory 68, 258–265. https://doi.org/10.1006/jeth.1996.0014.

[22] Nachbar, J. I., 1990. "Evolutionary" selection dynamics in games: convergence and limit properties. Int. J. Game Theory 19, 59–89. https://doi.org/10.1007/BF01753708.

[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] Samuelson, L., Zhang, J., 1992. Evolutionary stability in asymmetric games. J. Econ. Theory 57, 363–391. https://doi.org/10.1016/0022-0531(92)90041-F.

[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] Sutton, R. S., Barto, A. G., 2018. Reinforcement learning: an introduction. 2nd edition. Cambridge, MA: MIT Press.

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

[28] Viossat, Y., 2015. Evolutionary dynamics and dominated strategies. Econ. Theory Bull. 3, 91–113. https://doi.org/10.1007/s40505-014-0062-4.

[29] Viossat, Y., Zapechelnyuk, A., 2013. No-regret dynamics and fictitious play, J. Econ. Theory 148, 825–842. https://doi.org/10.1016/j.jet.2012.07.003.

[30] 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.

[31] 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.

[32] Young, H., P., 1998. Individual Strategy and Social Structure. Princeton, N.J.: Princeton University Press