Games **2013**, *4*, 648-669; doi:10.3390/g4040648

OPEN ACCESS

games

ISSN 2073-4336

www.mdpi.com/journal/games

Article

An Adaptive Learning Model in Coordination Games

Naoki Funai

Department of Economics, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK;
E-Mails: n.funai@bham.ac.uk or nfunaijp@yahoo.co.jp; Tel.: +44-744-780-5832

Received: 15 September 2013; in revised form: 4 November 2013 / Accepted: 7 November 2013 / Published: 15 November 2013

Abstract: In this paper, we provide a theoretical prediction of the way in which adaptive players behave in the long run in normal form games with strict Nash equilibria. In the model, each player assigns subjective payoff assessments to his own actions, where the assessment of each action is a weighted average of its past payoffs, and chooses the action which has the highest assessment. After receiving a payoff, each player updates the assessment of his chosen action in an adaptive manner. We show almost sure convergence to a Nash equilibrium under one of the following conditions: (i) that, at any non-Nash equilibrium action profile, there exists a player who receives a payoff, which is less than his maximin payoff; (ii) that all non-Nash equilibrium action profiles give the same payoff. In particular, the convergence is shown in the following games: the battle of the sexes game, the stag hunt game and the first order statistic game. In the game of chicken and market entry games, players may end up playing the action profile, which consists of each player's unique maximin action.

Keywords: payoff assessment; learning; coordination games

---

1. Introduction

Over the past few decades, learning models have received much attention in the theoretical and experimental literature of cognitive science. One such model is fictitious play, where players form beliefs about their opponents' play and best respond to these beliefs. In the fictitious play model, players know the payoff structure and their opponents' strategy sets.

Whereas there are other learning models in which players have only limited information about the game structure; players may not have information about payoff structure, opponents' strategy sets or they may not even know whether they are playing against other players. In this situation, they may not be

able to form beliefs about the way that the opponents play or all possible outcomes. What they do know is their own available actions and the results from the previous play, that is, the realized payoffs from chosen actions. Instead of forming beliefs about all possible outcomes, each player makes a subjective assessment on each of his actions based on the realized payoffs from the action and tends to pick the action that has achieved better results than the others.

One such model with limited information is the reinforcement learning model introduced by Erev and Roth [1] (ER, hereafter), where they model the observed behaviour of agents in the lab.[1] In their model, the agent chooses an action randomly, where the choice probability of the action is the fraction of the payoffs realized from the action over the total payoffs realized for all available actions.[2]

In another learning model, which is introduced by Sarin and Vahid [2] (SV, hereafter), each player makes a subjective payoff assessment on each of his actions, where the assessment is a weighted average of realized payoffs from the action, and chooses the action that has the highest assessment. After receiving a payoff, each player updates the assessment of the chosen action adaptively; the assessment of the chosen action is adjusted toward the received payoff.[3]

In this paper, we provide a theoretical prediction of the way in which adaptive players in the SV model behave in the long run in general games, mostly in coordination games, which are of interest to a wide range of researchers.[4] In this model, the initial assessment of each action is assumed to take a value between the maximum and the minimum payoff that the action can provide.[5] For instance, players may have experienced the game in advance, so they may use their knowledge of previous payoffs to form the initial assessment of each action. Given those assessments, each player chooses the action that has the highest assessment.

After playing a game and receiving a payoff, each player updates his assessment using the realized payoff; the new assessment of a chosen action is a convex combination of the current assessment and the realized payoff. In this paper, the sequence of weighting parameters is assumed to be stochastic,[6] meaning that the amount of new payoff information that each player incorporates into his new assessment in each period is uncertain. The randomness assumption is reasonable, because the weights are subjective and may depend on each player's capacity or emotion. For instance, a player may sometimes show inaccuracy in capturing new information because of his lack of concentration. The randomness also expresses the idea that each player's subjective weights on the new payoff information depend on his unexpected mood.

Since the initial assessment of each action is smaller than the best payoff that the action can give, each player increases his assessment of the action when he receives the best payoff. If there exists an

---

1 There also exists theoretical work on their model; for instance, see Beggs [3] and Laslier *et al.* [4].

2 Since the payoffs are assumed to be positive, each player increases the probability of choosing an action whenever the action is chosen.

3 It is worth noting that if the realized payoff of an action is lower than the assessment of the action, then the chance of the action being chosen in the next period becomes less likely.

4 As examples of experimental works on coordination games, Cooper, DeJong, Forsythe and Ross [5] and Van Huyck, Battalio and Beil [6] have investigated which among multiple Nash equilibria is the one played in the lab.

5 See, also, Sarin [7] for the justification of the assumption.

6 Later, we also consider cases where the weighting parameters are non-stochastic.

action profile at which each player receives the best payoff that his current action can give and they play the action profile in some period, then players will keep choosing the action profile in all subsequent periods. We call such an action profile the absorbing state; an action profile is absorbing if, once players play the action profile in some period, they play it in all subsequent periods.

Furthermore, there exist other cases where players stick to one action profile; if there exists a period in which, for each player, the assessment of his chosen action and the realized payoff is greater than the assessments of his other actions, then players keep choosing the action profile in all subsequent periods. It is shown that each pure Nash equilibrium is always a candidate of the convergence point, that is, for each strict Nash equilibrium, there exists a range of assessments for all players and actions, such that players stick to the Nash equilibrium forever. In addition, if: (i) at any non-Nash equilibrium action profile, at least one player receives the payoff, which is less than his maximin payoff; or (ii) all non-Nash equilibrium action profiles give the same payoff, then players end up playing a strict Nash equilibrium with probability one. To see this in detail, we focus on 2×2 coordination games and one non-2×2 coordination game.

First, we focus on 2×2 coordination games. For analytical purposes, we exclusively divide 2×2 coordination games into the following two categories: (i) at non-Nash equilibrium action profiles, there exists a player who receives the worst payoff; and (ii) there exists a non-Nash equilibrium action profile that each player's action corresponds to his unique maximin action. Then, it is shown that players end up playing a strict Nash equilibrium in coordination games in category (i), while players end up playing a strict Nash equilibrium or the action profile that consists of each player's unique maximin action in coordination games in category (ii). It is helpful for understanding the argument further to see well-known coordination games from each category. In particular, category (i) includes the battle of the sexes game, the pure coordination game and the stag hunt game, whereas category (ii) includes the game of chicken and market entry games.

Next, we focus on a non-2×2 coordination game introduced by Van Huyck, Battalio and Beil [6] (VHBB, hereafter) to compare the theoretical results from this model with their experimental results. In the game, each player is asked to pick a number from a finite set, and coordination is achieved when players pick the same number. If players fail to coordinate, the player who picks the smallest number among players' choices receives the highest payoff. In addition, each number gives a better payoff when the choice is closer to the smallest number among all the players' choices. We show that each Nash equilibrium is absorbing.[7] It is also shown that the smallest number of the players' choices weakly decreases over time. Next, we consider the case where the second best payoff from each action is lower than the payoff from the maximin action, which is the smallest number of their choice set. Hence, players are better off if they choose the smallest number of their choice set when they fail to pick the smallest number among the players' choices. In this case, we show that players end up playing a Nash equilibrium with probability one, which can be also observed in the experimental results by VHBB.

---

[7] It is absorbing if the minimum number gives different payoffs for opponents' choices. If it gives the same payoff for any opponents' choice, then we have to assume an inertia condition for the players' tie break rule for the corresponding Nash equilibrium to be absorbing. See the following argument.

It is also intriguing to consider cases where the weighting parameters are not stochastic. For example, we consider players who believe that the situation they are involved in is stationary, so that each action's assessment is the arithmetic mean of its past payoffs. We also consider the case where players believe that the environment is non-stationary and put the same weight on all new payoff information. Then, for each case, we show the necessary and sufficient condition for coordination failure, where players play non-Nash equilibrium action profiles alternately forever. In fact, in fictitious play, an example of the correlated play on off-diagonal action profiles is shown (Fudenberg and Levine [8]) and the empirical frequencies of the play should converge to the mixed Nash equilibrium in every 2×2 games with the diagonal property (Monderer and Shapley [9]). We, however, show a case in which the empirical frequencies of a correlated play on non-Nash equilibrium action profiles do not converge to the mixed Nash equilibrium.

2. Literature Review

In this paper, we investigate the convergence properties of the SV learning model in mainly coordination games. In another game, the prisoner's dilemma game, Sarin [7] shows that players end up playing for mutual cooperation or mutual defection. In this paper, players do not experience any stochastic emotional noise on their assessments; SV also investigate a case in which the decision maker experiences the stochastic emotional shocks on his assessments in each decision period, by which he may choose an action that does not have the highest assessment. Then, SV show that: (1) assessment of the action that is played infinitely often converges in distribution to a random variable whose expected value is the expected objective payoff; and (2) if one action first-order stochastically dominates the other, then the former action is played more often than the other on average. In the context of SV learning with the stochastic shocks in games, Leslie and Collins [10] investigate the model with slightly different updating rules and show convergence of strategies to a Nash distribution[8] in the partnership game and the zero-sum games. Using the SV updating rule, Cominetti, Melo and Sorin [11] show the general convergence result when each player's choice rule is the logistic choice rule. They show that players' choice probabilities converge to a unique Nash distribution if the noise term of the logistic choice rule for each player is big enough. By a property of the logistic choice rule, if its noise term becomes large, then the choice probability approaches a uniform distribution. Hence, players in their model are more likely to choose an action that does not have the highest assessment each time. However, players in the SV model without emotional shocks do not choose such actions; they always pick the action that they think is the best based on past payoff realizations. In this paper, even with the lack of such an exploration, it is shown that players end up playing a Nash equilibrium in several coordination games.

Lastly, some authors have provided empirical supports of this model. For instance, Sarin and Vahid [12] show that the SV model can explain the data by ER at least as well as the ER model does. Chen and Khoroshilov [13] show that among learning models comprising the ER model, the SV model

---

[8] Nash distribution is Nash equilibrium under stochastic perturbations on payoffs. If the expected values of the perturbations are zero, then Nash distribution coincides with the quantal response equilibrium proposed by McKelvey and Palfrey [15]

and the experience-weighted attraction learning model by Camerer and Ho [14], the SV model can best explain the data in coordination games and cost sharing games.

This paper is organized as follows. Section 3 provides the formal description of the model. Section 4 gives results in general games, while Sections 5 and 6 focus on 2×2 coordination games and a VHBB coordination game. Section 7 is devoted to the case in which weighting parameters are not stochastic. Section 8 concludes.

3. General Games

We consider a case in which M players play the same game repeatedly over periods. Let $N = \{1, ..., M\}$ be the set of players. In each period, $n \in \mathbb{N}$, each player chooses an action from his own action set simultaneously. Let S^i be the finite set of actions for player $i \in N$. After all the players choose actions, each player receives a payoff. If players play $(s^i)_{i \in N} \in \Pi_{i \in N} S^i$, then player i's realized payoff is denoted by $u^i(s^i, s^{-i})$, where $s^{-i} = (s^1, ..., s^{i-1}, s^{i+1}, ..., s^M)$. When choosing an action, each player does not know the payoff functions or the environment in which he is involved.

In each period, each player assigns subjective payoff assessments on his actions; let $Q_n^i(s^i) \in \mathbb{R}$ denote player i's assessment on action s^i in period n. Let $Q_n^i = (Q_n^i(s^i))_{s^i \in S^i}$ be the vector of assessments for all actions for player i. We assume that the initial assessment for each action and each player takes a value between the maximum and the minimum value that the action gives; thus:

$$Q_1^i(s^i) \in (\min_{s^{-i}} u^i(s^i, s^{-i}), \max_{s^{-i}} u^i(s^i, s^{-i}))$$

for all $i \in N$ and $s^i \in S^i$. If $\min_{s^{-i}} u^i(s^i, s^{-i}) = \max_{s^{-i}} u^i(s^i, s^{-i})$, then we assume that $Q_1^i(s^i) = \min_{s^{-i}} u^i(s^i, s^{-i}) = \max_{s^{-i}} u^i(s^i, s^{-i})$.

In each period, each player chooses the action that he believes will give the highest payoff; given his assessments, he chooses the action that has the highest assessment in the period. Therefore, if s_n^{i*} is the action that player i chooses in period n, then:

$$s_n^{i*} = \arg\max_{s^i} Q_n^i(s^i)$$

For a tie break situation, which arises when more than two actions have the highest assessment, we introduce two types of tie break rules. We say that a tie break rule satisfies the inertia condition if the rule chooses the action that was chosen in the last period; if actions that have the highest assessment were not chosen in the last period, then the rule picks one of the actions randomly. As a comparison, we also introduce another tie break condition, the uniform condition, where the rule picks each of the actions that have the highest assessment with equal probability. In the following argument, we specify a tie break rule if the result depends on the tie break rule; otherwise, the results do not depend on the tie break rule assumption.

After playing the game in each period, each player observes only his own payoff; players observe neither their opponents' actions nor their payoffs. Given his own realized payoff, each player updates his assessment of the action chosen in the previous period. Specifically, if player i receives a payoff $u_n^i(s^i, s^{-i})$ when players play (s^i, s^{-i}), then he updates Q_n^i as follows:

$$Q_{n+1}^i(s^i) = \begin{cases} (1 - \lambda_n^i(s^i))Q_n^i(s^i) + \lambda_n^i(s^i)u_n^i(s^i, s^{-i}) & \textit{if } s^i \textit{ is chosen in period } n \\ Q_n^i(s^i) & \textit{otherwise} \end{cases}$$

where $\lambda_n^i(s^i)$ is player i's weighting parameter for action s^i in period n. We assume that $\lambda_n^i(s^i)$ is a random variable that takes a value between zero and one; $\lambda_n^i(s^i) \in (0,1)$. It reflects the idea that players are uncertain how far to incorporate the new payoff information into their new assessments. The uncertainty can also be interpreted as players' emotional shocks. How far they incorporate the new payoff information depends on their random mood. We also assume that: (i) the sequence of weighting parameters, $\{\lambda_n^i(s^i)\}_{i,n,s^i}$, is independent among periods, players and actions and is identically distributed among periods; and (ii) each component, $\lambda_n^i(s^i)$, has a density function that is strictly positive on the domain $(0,1)$ for all i and s^i.

4. Results

In this section, we investigate the convergence results in general games. In later sections, we focus on more specific games, in particular, coordination games.

We first show a sufficient condition under which an action profile is absorbing. We say $(s^i)_{i\in I}$ is *absorbing* if once players play the action profile in a period, then they play it in all subsequent periods.

Proposition 1. *If $(s^i)_{i\in I}$ is such that (i) for all i:*

$$u^i(s^i, s^{-i}) = \max_{t^{-i}\in S^{-i}} u^i(s^i, t^{-i})$$

and (ii) for all i, there exists r^{-i}, such that:

$$\max_{t^{-i}\in S^{-i}} u^i(s^i, t^{-i}) > u^i(s^i, r^{-i})$$

then $(s^i)_{i\in I}$ is absorbing.

Proof. Consider the case where players pick the action profile, $(s^i)_{i\in I}$, in some period, n. In that case, player i receives the payoff, $u^i(s^i, s^{-i})$. Note that the value $u^i(s^i, s^{-i})$ is the maximum value that action s^i can give; therefore, by condition (ii), player i inflates the assessment of the action, s^i. Since the assessments of other actions do not change in the next period, player i plays action s^i in period $n+1$ again. Since this logic can be applied to other periods and we pick player i randomly, players play the same action profile in all the subsequent periods. □

Proposition 1 says that if at an action profile, each player receives a payoff that is strictly better than the other payoffs from his current action, then the action profile is absorbing.

If the inertia condition is always assumed for each player's tie break rule, then condition (ii) in Proposition 1 is not required. However, if the uniform condition is assumed, without condition (ii), players may not end up playing one action profile. As an extreme example, if two actions give the same payoff for any opponents' actions and the payoff is higher than any other payoffs that any other action can give, then he plays those two actions with equal probability forever.

From Proposition 1, it is easy to see that even action profiles that consist of dominated strategies for all players can be absorbing. To see this, assume that two players play the prisoner's dilemma game, which has the following payoff matrix:

	C	D
C	1,1	-1,2
D	2,-1	0,0

Note that the strategy "C" is strictly dominated by the strategy "D" for both players. Notice also that at (C,C), both players receive the highest payoffs from the action "C"; $u^i(C,C) = \max_{s^{-i} \in \{C,D\}} u^i(C, s^{-i})$, $\forall i \in N$. Hence, if players play (C,C) once, then they always play it afterwards.[9]

In the next statement, we show that player *i* stops playing an action if the assessment of the action becomes smaller than the minimum payoff that another action can give.

Proposition 2. *If $Q_n^i(s^i) < \min_{s^{-i}} u(t^i, s^{-i})$ in some period, n, for $t^i \neq s^i$, then player i does not choose s^i after period n.*

Proof. From the fact that $Q_n^i(t^i) \geq \min_{s^{-i}} u(t^i, s^{-i})$, we have the fact that $Q_n^i(t^i) > Q_n^i(s^i)$. Notice that s^i is not chosen in period n. Then, we have $Q_{n+1}^i(t^i) > \min_{s^{-i}} u(t^i, s^{-i}) > Q_{n+1}^i(s^i)$,[10] and player i will not choose s^i in period $n+1$. The same logic can be applied in later periods, and thus, player i will not choose s^i in any subsequent periods. □

Once the assessment of one action becomes lower than the worst payoff from another action, then the action will not be chosen forever. Therefore, if the worst payoff from one action is greater than the best payoff from another action, then the latter action is never chosen at any time. One natural question is whether players end up playing a strict Nash equilibrium. We say that players end up playing $(s^i)_{i \in I}$ if there exists n, such that for all periods after n, players play $(s^i)_{i \in N}$. If the condition, $Q_m^i(s^i) > Q_m^i(t^i)$, satisfies for all i, $m > n$ and $s^i \neq t^i$, then players end up playing $(s^i)_{i \in I}$.[11] In the following statement, we show that for any strict Nash equilibrium, there exist assessments for all players, such that they end up playing the strict Nash equilibrium:

Proposition 3. *For any strict Nash equilibrium and any period, there exist assessments for all players, such that they play the Nash equilibrium in the period and all subsequent periods.*

Proof. Let $(s^{i*})_{i \in N}$ be a strict Nash equilibrium and s^{i*} be player i's strategy at the strict Nash equilibrium. Then, we have the following condition; for all $i \in N$:

$$u^i(s^{i*}, s^{-i*}) > u^i(t^i, s^{-i*}) \tag{1}$$

for all $t^i \neq s^{i*}$. Now, we pick assessments of players, such that the following conditions are satisfied; for all i :

$$Q_n^i(s^{i*}) > Q_n^i(t^i) \tag{2}$$

---

[9] See, also, Sarin [7] for the result.

[10] Since the assessment of the chosen action is a convex combination of the realized payoff and the assessment of the previous period with $\lambda_n^i \in (0,1)$ for all i and n, we have $Q_{n+1}^i(t^i) > \min_{s^{-i}} u(t^i, s^{-i})$. Notice also that s^i is not chosen in period n, and thus, the assessment of the action is unchanged in the next period, $n+1$.

[11] This condition does not include some convergence case that happens when we assume the inertia condition for all players. In such a case, we can weaken the condition as follows: players end up playing $(s^i)_{i \in N}$ if there exist n and $(Q_n^i)_{i \in I}$, such that for all $m \geq n$, i and $t^i \neq s^i$, $Q_m^i(s^i) \geq Q_m^i(t^i)$, where player i chooses s^i in period n.

and:

$$u^i(s^{i*}, s^{-i*}) > Q^i_n(t^i) \tag{3}$$

for all $t^i \neq s^{i*}$. Note that Equation (3) can hold, since by Equation (1), the minimum value of the assessment of action t^i is less than or equal to $u^i(t^i, s^{-i*})$, which is strictly less than $u^i(s^{i*}, s^{-i*})$. Thus, by Equations (2) and (3), players play the strict Nash equilibrium in period n and:

$$Q^i_{n+1}(s^{i*}) \geq \min\{Q^i_n(s^{i*}), u^i(s^{i*}, s^{-i*})\} > Q^i_n(t^i) = Q^i_{n+1}(t^i)$$

for all $t^i \neq s^{i*}$. Therefore, players play the strict Nash equilibrium again in period $n+1$. Notice that we can apply the same argument for the following periods, and thus, players play the strict Nash equilibrium in all subsequent periods. □

Proposition 3 says that for any strict Nash equilibrium, there exists a case in which players end up playing the strict Nash equilibrium. However, it is possible that players end up playing a non-Nash equilibrium. Hence, it is natural to consider the case where if they converge to play one action profile, then it should be a strict Nash equilibrium.

In the following statements, we focus on the cases where all pure Nash equilibria are strict. We also assume that there do not exist any redundant actions that always give the same constant payoff; for any $i \in N$ and actions $s^i, t^i \in S^i$, $s^i \neq t^i$, the following condition does not hold:

$$u^i(s^i, s^{-i}) = u^i(t^i, t^{-i}) \text{ for all } s^{-i}, t^{-i} \in S^{-i}$$

Lemma 1. *For any initial assessments, players never end up playing $(s^i)_{i\in N}$ if $\exists i \in N$ s.t.:*

$$u^i(s^i, s^{-i}) \leq \max_{t^i \neq s^i} \min_{t^{-i} \in S^{-i}} u^i(t^i, t^{-i}) \tag{4}$$

where if the inequality holds with equality, then the following condition also holds; for $t^{i*} = \arg\max_{t^i \neq s^i} \min_{t^{-i} \in S^{-i}} u^i(t^i, t^{-i})$:

$$u^i(t^{i*}, t^{-i}) \neq u^i(t^{i*}, r^{-i}) \textit{ for } r^{-i} \neq t^{-i} \in S^{-i}$$

Proof. To prove the statement, we should consider the case in which Equation (4) holds strictly and the case in which Equation (4) holds with equality.In both cases, we prove by contradiction.

We first consider the case in which Equation (4) holds strictly. Then, we assume that there exist assessments, such that players end up playing $(s^i)_{i\in N}$. It implies that $\exists n, \forall m > n$:

$$\min\{Q_m(s^i), u^i(s^i, s^{-i})\} \geq Q_m(t^i) \geq \min_{t^{-i} \in S^{-i}} u^i(t^i, t^{-i})$$

$\forall t^i \in S^i$. Note that for $m > n$ and $t^i, \neq s^i$, $u^i(s^i, s^{-i})$ should be greater than $Q^i_m(t^i)$; it is the condition for the other actions not to be chosen in the next period. However, it contradicts the fact that Equation (4) holds strictly.

We next assume that Equation (4) holds with equality. We again assume that there exist assessments, such that players end up playing $(s^i)_{i\in N}$. Since we have the additional condition, it implies that $\forall m > n$:

$$\min\{Q_m(s^i), u^i(s^i, s^{-i})\} \geq Q_m(t^{i*}) > \min_{t^{-i} \in S^{-i}} u^i(t^{i*}, t^{-i})$$

However, it contradicts the fact that Equation (4) holds with equality. □

If Equation (4) is satisfied at non-Nash equilibrium action profiles, then players never end up playing one of them. It is also obvious that the condition is not satisfied at each strict Nash equilibrium. Equation (4) says that there exists a player who receives a payoff, which is less than his maximin payoff. Though the condition limits the class of games, still, there exist interesting games that satisfy the condition. For example, the stag hunt game satisfies Equation (4) at non-Nash equilibrium action profiles and has the following payoff matrix:

	Rabbit	Stag
Rabbit	1,1	2,0
Stag	0,2	5,5

At non-Nash equilibrium action profile, one player decides to hunt a stag, while the other player decides to hunt a rabbit. The player who decides to hunt a stag fails and receives nothing. Note that the payoff is less than the maximin payoff, which is obtained when both players decide to hunt a rabbit and share the rabbit.

Another coordination game that satisfies Equation (4) is the first order statistic game, where each player chooses a number from a finite set, and coordination occurs when all of them pick the same number. In addition, if players succeed in coordinating at a higher number, then they receive a better payoff. When they fail to coordinate on choosing the same number, the player who has chosen the smallest number receives the best payoff, the player who has chosen the second smallest number receives the second best payoff, and so on; the smaller number the player has chosen, the better payoff he receives. For example, we consider the case where each player picks a number from one to four and the payoff matrix of each player is expressed as follows:

	1	2	3	4
1	1	1.5	1.5	1.5
2	0	2	2.5	2.5
3	-1	0	3	3.5
4	-2	-1	0	4

The first column represents player i's choice, while the first row represents the minimum value of his opponents' choices. It is easy to see that at each Nash equilibrium, all players pick the same number. Since action 1 gives at least one, and players who fail to pick the smallest number receive at most zero, this game satisfies Equation (4).

In both games, Equation (4) holds strictly. In other games, such as the battle of the sexes games, Equation (4) holds weakly, in particular, $u^i(s^i, s^{-i}) = u^i(t^i, t^{-i})$ for all i and $(s^i), (t^i) \notin E^*$, where E^* is the set of pure Nash equilibria. For instance, a battle of the sexes game has the following payoff matrix:

	s_1^2	s_2^2
s_1^1	1,2	0,0
s_2^1	0,0	2,1

In the following theorem, we show that players end up playing a Nash equilibrium almost surely if: (i) Equation (4) is satisfied strictly at non-Nash equilibrium profiles; or (ii) if each player's payoffs at non-Nash equilibrium action profiles are equal.

Theorem 1. *Players end up playing a strict Nash equilibrium almost surely if (i)* $\forall (s^i)_{i\in N} \notin E^*$, $\exists i \in N$ *such that :*

$$u^i(s^i, s^{-i}) < \max_{t^i \neq s^i} \min_{t^{-i} \in S^{-i}} u^i(t^i, t^{-i}) \tag{5}$$

or (ii) $\forall i \in N$, $\forall (s^i)_{i\in N}, (t^i)_{i\in N} \notin E^*$,

$$u^i(s^i, s^{-i}) = u^i(t^i, t^{-i})$$

Proof. Case (i): (Only the intuition of the proof is provided here. For details, see the Appendix.) It is a direct consequence from Lemma 1 that if players end up playing one action profile, then it should be a strict Nash equilibrium. Therefore, it should be shown that they actually end up playing a strict Nash equilibrium. The intuition of the proof is as follows. Since off-diagonal action profiles cannot be played infinitely often, there exists a period after which players only play strict Nash equilibria. Since we consider games with strict Nash equilibrium, players should change their actions at the same time when they move from one Nash equilibrium to another Nash equilibrium. Note also that since the sequence of weighting parameters is independent and each component has a density function that is positive on its domain, perfect correlated play on strict Nash equilibria is impossible. Now, the detailed proofs are given in the following arguments.

Case (ii): Note that if condition (ii) satisfies, then the payoff from any Nash equilibrium should be greater than the payoff from non-Nash equilibrium; $u^i(s^{i*}, s^{-i*}) > u^i(s^i, s^{-i})$ for all $i \in N$, $(s^{i*}) \in E^*$ and $(s^i) \notin E^*$. Therefore, each strict Nash equilibrium is absorbing, and thus, once players play a strict Nash equilibrium, then they play it forever. By the same logic as the proof in (i), players cannot play only non-Nash equilibrium action profiles forever. That is, with probability one, players play a strict Nash equilibrium at some time, and then play it in all subsequent periods. □

Theorem 1 shows that players end up playing a strict Nash equilibrium almost surely in the games that satisfy condition (i) or (ii). To see this in detail, in the following sections, we investigate 2×2 and non-2×2 coordination games.

5. 2 ×2 Coordination Games

In this section, we focus on 2×2 coordination games, which have the following payoff matrix:

	s_1^2	s_2^2
s_1^1	a_{11}, b_{11}	a_{12}, b_{12}
s_2^1	a_{21}, b_{21}	a_{22}, b_{22}

where $a_{11} > a_{21}$, $a_{22} > a_{12}$, $b_{11} > b_{12}$ and $b_{22} > b_{21}$. Note that given the conditions, the pure Nash equilibria are (s_1^1, s_1^2) and (s_2^1, s_2^2). For the purposes of the analysis, we exclusively divide 2×2

coordination games into the following two categories. In the first category, at non-Nash equilibrium action profiles, there exists at least one player who receives his worst payoff; $a_{kl} \geq a_{lk}$ and $b_{lk} \geq b_{kl}$ for $k \neq l$. In the second category, there exists an action profile at which each player's action corresponds to his unique maximin action: $a_{kl} > a_{lk}$ and $b_{lk} > b_{kl}$ for $k \neq l$. Then, we have the following result:

Proposition 4. *With probability one, players end up playing: (i) a strict Nash equilibrium in 2 × 2 coordination games in the first category; and (ii) a Nash equilibrium or the action profile at which each player's action corresponds to his unique maximin action in 2 × 2 coordination games in the second category.*

Proof. (i) Note that by Theorem 1, we only have to consider the case in which only one of the inequalities holds with equality: without loss of generality, we assume that $a_{12} = a_{21}$ and $b_{12} > b_{21}$. Notice that (s_1^1, s_1^2), which is a strict Nash equilibrium, is absorbing. Since at (s_2^1, s_1^2), player 2 receives the payoff, b_{21}, which is strictly lower than the worst payoff of another action, b_{12}; players never play the action profile infinitely many times. Since player 1 receives the worst payoff at (s_1^1, s_2^2), players never end up playing the action profile. Therefore, the last case to be considered is that players play only (s_1^1, s_2^2) and (s_2^1, s_2^2) alternately forever without ending up playing one of the action profiles. However, it cannot happen, since $a_{12} < a_{22}$. Therefore, players end up playing one of strict Nash equilibria.

(ii) Without loss of generality, we assume that $a_{21} > a_{12}$ and $b_{21} > b_{12}$. Note that (s_1^1, s_2^2) is not played infinitely many times, since, at the action profile, each player receives his worst payoff. It is easy to show that there exists a case in which players end up playing (s_2^1, s_1^2). What we have to show is that players actually end up playing a strict Nash equilibrium or (s_2^1, s_1^2) with probability one. To show that, we consider the following four possible cases according to the number of absorbing states under the uniform condition for each player's tie break rule.[12]

(1) Consider the case where there exist two absorbing states that correspond to strict Nash equilibria. Since strict Nash equilibria are absorbing, it is obvious that they end up playing a strict Nash equilibrium or (s_2^1, s_1^2).

(2) Consider the case where there exists one absorbing state that corresponds to a strict Nash equilibrium. Let such a strict Nash equilibrium be (s_1^1, s_1^2). The only case to be considered is that players play only (s_2^1, s_1^2) and (s_2^1, s_2^2) alternately without converging one of them. Since s_1^2 and s_2^2 are played infinitely many times, it should be that $b_{21} = b_{22}$, which contradicts the condition for the coordination game. Thus, players end up playing a strict Nash equilibrium or (s_2^1, s_1^2).

(3) Consider the case where (s_2^1, s_1^2) is absorbing. It can be shown by the argument in Theorem 1 that perfect correlation on strict Nash equilibria is impossible, and thus, players end up playing a strict Nash equilibrium or (s_2^1, s_1^2).

(4) Lastly, consider the case where there exists no absorbing state. Then, the following condition on players' payoffs should hold: $a_{21} \geq a_{22}$ and $b_{21} \geq b_{11}$, where at least one of them should hold with equality. Without loss of generality, we assume that $a_{21} = a_{22}$ and $b_{21} \geq b_{11}$. Note that once s_2^1 is

[12] The uniform condition for each players' tie break rule is only used for the categorization; the result does not depend on the condition. Note that if an action profile is absorbing under the uniform condition, then it is absorbing under any other condition for each player's tie break rule.

played, it is played in all subsequent periods.[13] Since, by the same logic above, (s_2^1, s_1^2) and (s_2^1, s_2^2) are not played alternately forever, they end up playing a strict Nash equilibrium or (s_2^1, s_1^2). □

In fact, this categorization helps us to understand the long run outcomes of the adaptive learning process in 2×2 coordination games. In the following argument, we focus on specific coordination games: the battle of the sexes game, the stag hunt game, the game of chicken and market entry games.

Consider first the battle of the sexes game, which has the following payoff matrix:

	Opera	Football
Opera	$1, 2$	$0, 0$
Football	$0, 0$	$2, 1$

In this game, the row player prefers going to a football game together to going to an opera together, while the column player enjoys going to the opera together rather than going to the football game together. However, players are worse off when they fail to coordinate to go to one of them. By Proposition 4, we know that players end up playing a strict Nash equilibrium almost surely.

There exists another form of the battle of the sexes game, which has the following payoff matrix:

	Opera	Football
Opera	1,2	0,0
Football	0.5,0.5	2,1

Notice that the row player enjoys going to a football game alone rather than going to an opera alone. The column player is in the opposite situation; she enjoys going to the opera alone rather than going to the football game alone. In this case, it is a possible outcome that players fail to coordinate, and they end up playing their favored actions (football, opera).

We secondly consider the stag hunt game, which has the following payoff matrix:

	Stag	Rabbit
Stag	10,10	0,8
Rabbit	8,0	4,4

In the stag hunt game, at each off-diagonal action profile, one player receives the worst payoff. Therefore, by Theorem 1 or Proposition 4, players end up playing a strict Nash equilibrium almost surely.

We thirdly consider the game of chicken:

	Swerve	Stay
Stay	1,-1	-10,-10
Swerve	0,0	-1,1

The game describes the following situation. There are two drivers who are facing each other to show their braveness. When one driver swerves while his opponent stays, he shows his cowardice to the

[13] Tie break situation can be ignored, since it happens with zero probability.

audience. If both drivers swerve, then both of them are safe and receive nothing. However, the best outcome for each driver is that he stays, while the opponent swerves, so that he can show his braveness. Whereas the worst scenario is that both drivers stay and have a severe accident. Note that each player receives the worst payoff at (stay, stay). Therefore, by Proposition 4, players end up playing a strict Nash equilibrium or (swerve, swerve).

Lastly, consider a market entry game, which has the following payoff matrix:

	Stay Out	Enter
Enter	100,0	-50,-50
Stay Out	0,0	0,100

In this game, players have to decide to enter a market or stay out from the market. If a player decides to stay out, regardless of his opponent's action, he receives nothing. If one player decides to enter while his opponent decides to stay out, he enjoys the profit from the market. However, if both players decide to enter, players face severe competition and earn negative profit. In this case, by Proposition 4, we know that players end up playing a strict Nash equilibrium or (stay out, stay out) almost surely.

6. VHBB Coordination Games

In this section, we consider the coordination game proposed by Van Huyck, Battalio and Beil [6], where there exist M players with $S^i = S = \{1, 2, ..., J\}$ for all $i \in N = \{1, ..., M\}$, and players have the following payoff function:

$$u^i(s^i, s^{-i}) = a(\min\{s^1, ..., s^M\}) - bs^i$$

where $a > b > 0$ for all $i \in N$. If J=4, then player i's payoffs are shown by the following matrix:

	1	2	3	4
1	a-b	a-b	a-b	a-b
2	a-2b	2a-2b	2a-2b	2a-2b
3	a-3b	2a-3b	3a-3b	3a-3b
4	a-4b	2a-4b	3a-4b	4a-4b

where the numbers in the first column correspond to player i's actions and the numbers in the first row correspond to the minimum values of the opponents' actions. It is easy to check that $(j, j, j, ..., j)$, $j \in S$, is a pure Nash equilibrium and is absorbing[14] under the inertia condition.

Proposition 5. *Assume the inertial condition for each player's tie break rule. Then, for $j \in S$, the pure Nash equilibrium $(j, j, ..., j)$ is absorbing.*

[14] It is easy to check that Nash equilibria, except $(1, 1, ..., 1)$, are absorbing. However, under the inertia condition for each player's tie break rule, $(1, 1, ..., 1)$ is also absorbing.

When a player is choosing the smallest action among players' actions, he is receiving the best payoff that the action can give. Therefore, the player does not change his action when he is choosing the smallest action, except when he chooses one and is facing a tie break situation. If the inertia condition is satisfied, then he chooses one forever, and the minimum value of actions does not increase over time. Moreover, since the minimum value is bounded below, it converges.

Proposition 6. *Assume the inertia condition for each player's tie break rule. Then, the minimum value of actions among players is non-increasing over periods and converges almost surely.*

We now assume that each action's second best payoff, $a(j-1)-bj$ for $j \in S/\{1\}$, is less than the secure payoff, $a-b$. That is:

$$a(j-1)-bj < a-b$$

for all $j \in S/\{1\}$. This means that each player receives a payoff better than his maximin payoff only when his choice is the smallest among all players' choices. Since there exists at least one player who receives a payoff, which is less than his maximin payoff at non-Nash equilibrium action profiles, by Theorem 1, players end up playing a Nash equilibrium.

Corollary 1. *If $a(j-1)-bj < a-b$ for all $j \in S\backslash\{1\}$, then players end up playing a pure Nash equilibrium almost surely.*

7. Non-Random Weighting Parameters

7.1. Coordination Failure

In this section, we assume that players' weighting parameters are not random variables. For example, players may believe that all past experiences equally represent the corresponding action's value, that is, players believe that the environments in which they are involved are stationary. Therefore, in each period, players put the same weight on all past experiences and players' assessments become the arithmetic mean of past payoffs. Note that the weighting parameters for each player are as follows: $\lambda_n^i(s_j^i) = \frac{1}{\tau(n)+1}$ for all $i \in N$ and $s_j^i \in S^i$, where $\tau(n)$ is the number of times that the action, s_j^i, is played until period n.

We also consider the players who have the following weighting parameters: $\lambda_n^i(s_j^i) = \lambda$ for all i, s_j^i and n as in Sarin and Vahid [12]; all players have constant weighting parameters in all periods, that is, both players always put the same weight on the received payoff in each period. It is reasonable to assume this condition if players believe that the situation they are facing is non-stationary. If λ is close to one, then players believe that only the most recent payoffs give information about the values of corresponding actions. If λ is close to zero, then players believe that initial assessments of actions mostly represent the actions' value.

In this section, we consider the battle of the sexes game, in which players may play off-diagonal action profiles alternately without ending up at a Nash equilibrium. In detail, we first consider the case where $\lambda_n^i(s_j^i) = \frac{1}{\tau(n)+1}$ for all i, s_j^i and n and off-diagonal payoffs for each player are all equivalent; $a_{12} = a_{21}, b_{12} = b_{21}$. In particular, we assume that $a_{12} = 0$ and $b_{12} = 0$.

As an example, consider the case where players' initial assessments are as follows: $Q_1^1(s_1^1) = 0.2$, $Q_1^1(s_2^1) = 0.2+\epsilon, Q_1^2(s_1^2) = 0.2+\epsilon, Q_1^2(s_2^2) = 0.2$, where $\epsilon \in (0, 0.2)$ is an irrational number. In

this case, in the first period, they play (s_2^1, s_1^2), and both players receive a payoff of zero. In period 2, players' assessments are as follows: $Q_2^1(s_1^1) = 0.2, Q_2^1(s_2^1) = \frac{1}{2}(0.2+\epsilon), Q_2^2(s_1^2) = \frac{1}{2}(0.2+\epsilon), Q_2^2(s_2^2) = 0.2$. Notice that the assessments of s_1^1 and s_2^2 are greater than the assessments of s_2^1 and s_1^2. Hence, players play (s_1^1, s_2^2), and both players receive a payoff of zero. Using the payoff information in period 2, they update their assessments, and they have the following assessments in period 3: $Q_3^1(s_1^1) = \frac{1}{2}(0.2)$, $Q_3^1(s_2^1) = \frac{1}{2}(0.2+\epsilon)$, $Q_3^2(s_1^2) = \frac{1}{2}(0.2+\epsilon)$, $Q_3^2(s_2^2) = \frac{1}{2}(0.2)$. Then, players play (s_2^1, s_1^2) in period 3. Notice that their assessments of action s_1^1 and s_2^2 never coincide with the assessments of action s_2^1 and s_1^2 at any period because of ϵ. After period 3, players play (s_2^1, s_1^2) until the corresponding assessments become lower than the assessments of (s_1^1, s_2^2). After the event, players again switch back to play (s_1^1, s_2^2), and so on.

When $\lambda_n^i(s_j^i) = \frac{1}{\tau(n)+1}$ for all i, s_j^i and n, the following statement shows the condition of initial assessments for coordination failures, which is the play on off-diagonal action profiles alternately. In this section, we assume that players' tie break rules satisfy the inertia condition.

Proposition 7. *In 2 × 2 coordination games with $a_{12} = a_{21} = b_{12} = b_{21} = 0$, under the inertia condition, if $\lambda_n^i(s_j^i) = \frac{1}{\tau(n)+1}$ for all i, s_j^i and n, then the necessary and sufficient condition for the coordination failure is as follows:*

$$\frac{Q_1^1(s_2^1)}{Q_1^1(s_1^1)} = \frac{Q_1^2(s_1^2)}{Q_1^2(s_2^2)}$$

Proof. See the Appendix. □

This result says that players will play non-Nash equilibria alternately forever, if and only if players' ratios of initial assessments "coordinate".

Next, we consider the players who have the following weighting parameters: $\lambda_n^i(s_j^i) = \lambda$ for all i, s_j^i and n. Then, the necessary and sufficient condition of initial assessments for the coordination failure is as follows:

Proposition 8. *In 2 × 2 coordination games with $a_{12} = a_{21} = b_{12} = b_{21} = 0$, under the inertia condition, if $\lambda_n^i(s_j^i) = \lambda$ for all i, s_j^i and n, then the necessary and sufficient condition for the coordination failure is as follows; for some $z \in \mathbb{Z}$:*

$$(1-\lambda)^{z-1} > \frac{Q_1^1(s_2^1)}{Q_1^1(s_1^1)} \geq (1-\lambda)^z \text{ and } (1-\lambda)^{z-1} > \frac{Q_1^2(s_1^2)}{Q_1^2(s_2^2)} \geq (1-\lambda)^z$$

or

$$(1-\lambda)^{z-1} \geq \frac{Q_1^1(s_2^1)}{Q_1^1(s_1^1)} > (1-\lambda)^z \text{ and } (1-\lambda)^{z-1} \geq \frac{Q_1^2(s_1^2)}{Q_1^2(s_2^2)} > (1-\lambda)^z$$

Proof. See the Appendix. □

Since players play a Nash equilibrium forever if they coordinate once on the Nash equilibrium, for each case, the negation of the condition is the one for the success of coordination. For instance, if off-diagonal payoffs are all zero and players are frequentists, then they coordinate in some period and in all subsequent periods, if and only if the initial assessments for both players and actions should satisfy the following condition: $\frac{Q_1^1(s_2^1)}{Q_1^1(s_1^1)} \neq \frac{Q_1^2(s_1^2)}{Q_1^2(s_2^2)}$.

7.2. Non-Convergence to a Mixed Nash Equilibrium

It is an interesting question whether the empirical frequencies of play on the off-diagonal action profiles converge to the mixed Nash equilibrium. In fictitious play, Monderer and Shapley [9] show that every 2×2 game with the diagonal property[15] has the fictitious play property; the empirical frequencies of past play, which is a belief of players about an opponent player's behaviour, converges to a Nash equilibrium.

First note that 2×2 coordination games with $a_{21} = a_{12} = b_{12} = b_{21} = 0$ also have the diagonal property. In this case, under the condition of coordination failure, players forever play off-diagonal action profiles alternately. However, the frequency of the play need not converge to the mixed Nash equilibrium. We show this by an example. Consider a pure coordination game, which has the following payoff matrix:

	s_1^2	s_2^2
s_1^1	1,1	0,0
s_2^1	0,0	1,1

We assume that weighting parameters and initial assessments for players are as follows: $\lambda_n^1(s_1^1) = \lambda_n^2(s_2^2) = \frac{1}{2}$, $\lambda_n^1(s_2^1) = \lambda_n^2(s_1^2) = \frac{1}{4}$, $Q_1^1(s_1^1) = Q_1^2(s_2^2) = \frac{1}{2}$, $Q_1^1(s_2^1) = Q_1^2(s_1^2) = \frac{1}{4}$. Under the inertia condition for both players, it is easy to see that players play action profiles in the following order: $(s_1^1, s_2^2) \to (s_1^1, s_2^2) \to (s_2^1, s_1^2) \to (s_1^1, s_2^2) \to (s_1^1, s_2^2) \to (s_2^1, s_1^2) \to \ldots$. In period 1, they play (s_1^1, s_2^2), and the assessments of s_1^1 and s_2^2 become $\frac{1}{4}$. Because of the inertia condition, they choose (s_1^1, s_2^2) again in period 2, and their assessments become $\frac{1}{8}$. Now, players change to play (s_2^1, s_1^2) in period 3, and the assessments of s_2^1 and s_1^2 become $\frac{1}{16}$. In period 4, players return to play (s_1^1, s_2^2), and so on. Therefore, the empirical frequencies of play for both players converge to $((\frac{2}{3}, \frac{1}{3}), (\frac{1}{3}, \frac{2}{3}))$, while the mixed Nash equilibrium in this game is $((\frac{1}{2}, \frac{1}{2}), (\frac{1}{2}, \frac{1}{2}))$.

8. Conclusions and Discussion

In this paper, we provide a theoretical foundation of adaptive learning in games with strict Nash equilibria. Different from the belief-based learning model, players in this model have limited information about their decision-making environment. We provide conditions under which players end up playing a strict Nash equilibrium; it happens when: (i) at the non-Nash equilibrium action profile, one player receives a payoff that is less than his maximin payoff; and (ii) all the payoffs from non-Nash equilibrium action profiles are the same for each player. We also investigate specific cases, such as the VHBB

---

[15] The game has the diagonal property if $\alpha \neq 0$ and $\beta \neq 0$, where:

$$\alpha = a_{11} + a_{22} - a_{12} - a_{21}$$

and:

$$\beta = b_{11} + b_{22} - b_{12} - b_{21}$$

coordination game, the stag hunt game, the battle of the sexes game, the game of chicken and market entry games.

This model can be also interpreted as a population model. Consider the situation in which there exist two large populations of naive players. In each period, one player is picked from each population randomly and plays a 2×2 coordination game, but he can play the game only once.[16] After each player plays the game, he reports the payoff that he has received to each population. We assume that each population does not share information with the other population. Each population accumulates information as a public assessment, which consists of realized payoffs and the initial assessment. In each period, the public assessment of the action that is played is updated, using realized payoffs as defined above; the convex combination of the realized payoff and the public assessment in the previous period. Each player may not know whether he is playing a game, but he knows the public assessment. Using the public assessment, each player chooses an action that has the highest public assessment.

For example, consider the battle of the sexes game. After the result of going to the opera or to football, both players report the realized payoff to the population that they belong to, so that people in the population can make an assessment before they play the game themselves. The result above says that players from two different populations never coordinate when initial assessments satisfy the condition in Proposition 7 when they are frequentists. Otherwise, players coordinate to play one of the pure Nash equilibria.

Appendix

A. Proof of Theorem 1

A.1. Detailed Proof for Case (i) in Theorem 1

At any non-Nash equilibrium action profile, there exists a player who is receiving a payoff that is less than his maximin payoff. Assume that a non-Nash equilibrium action profile, denoted by $(s^i)_i$, is played in period $n-1$. Then, a player whose payoff is less than his maximin payoff never plays his current action again if the assessment of his current action becomes lower than his maximin payoff. Note that the probability of such an event is bounded below by the following probability:

$$\Pr(Q_n^i(s^i) \in (u^i(s^i, s^{-i}), \min_{t^{-i} \in S^{-i}} u^i(t^i, t^{-i}))) \mid A)$$

where $A = \{Q_{n-1}^i(s^i) = \max_{s^{-i}} u^i(s^i, s^{-i})\}$. Since the sets of players and actions are finite, if players play a non-Nash equilibrium action profile infinitely often, then the player who receives a worse payoff stops playing the action with probability one. Therefore, players do not play a non-Nash equilibrium action profile infinitely often. Hence, we assume that players only play some Nash equilibrium action profiles. The remaining cases to be considered are that players play some strict Nash equilibria alternately without converging one of them. Since the game that we consider here has only strict Nash equilibria, all players should change their strategies at the same time when they change from one Nash

[16] Alternatively, each population is so large that the probability that a player plays a game again is almost zero.

equilibrium to another. Let $(s^{i*})_{i\in N}$ and $(s^{i**})_{i\in N}$ be two different strict Nash equilibrium action profiles that are played infinitely often. In this argument, we assume that players play only those two strict Nash equilibria alternately. The argument can be extended easily to the case where players play more than two Nash equilibria. Note that since players change one strict Nash equilibrium action profile to another strict Nash equilibrium action profile at the same time, all players should receive the payoffs that are strictly less than their current assessments. It also should be true that $u^i(s^{i*}, s^{-i*}) = u^i(s^{i**}, s^{-i**})$ for all i, and each player i's assessment never reaches the level $u^i(s^{i*}, s^{-i*})$ in a finite period. In the following argument, we show that players fail to play strict Nash equilibria alternately with probability one; to show that, we restrict our attention to the periods in which players change from $(s^{i**})_{i\in N}$ to $(s^{i*})_{i\in N}$.

By the assumption on weighting parameters, we can ignore the case where $Q^i(s^{i*}) = Q^i(s^{i**})$. Now, we consider period n, such that the condition $Q^i_{n-1}(s^{i*}) > Q^i_{n-1}(s^{i**}) > Q^i_n(s^{i*})$ holds. Then, for any small $\varepsilon \in (0,1)$, there exist $0 < c^i, d^i < 1$, such that:

$$\Pr(Q^i_n(s^{i*}) \in (\varepsilon u^i(s^{i**}, s^{-i**}) + (1-\varepsilon)Q^i_{n-1}(s^{i**}), Q^i_{n-1}(s^{i,**})) \mid B) \leq c^i$$

and:

$$\Pr(Q^i_n(s^{i*}) \in (u^i(s^{i**}, s^{-i**}), (1-\varepsilon)u^i(s^{i**}, s^{-i**}) + \varepsilon Q^i_{n-1}(s^{i**})) \mid B) \leq d^i$$

where $B := \{Q^i_{n-1}(s^{i*}) > Q^i_{n-1}(s^{i**}) > Q^i_n(s^{i*})\}$. Note that:

$$\begin{aligned}
& \Pr(Q^i_n(s^{i*}) \in (\varepsilon u^i(s^{i**}, s^{-i**}) + (1-\varepsilon)Q^i_{n-1}(s^{i**}), Q^i_{n-1}(s^{i**})) \mid B) \\
= \; & \frac{\Pr(\varepsilon u^{i**} + (1-\varepsilon)Q^{i**}_{n-1} < \lambda u^{i**} + (1-\lambda)Q^{i*}_{n-1} < Q^{i**}_{n-1})}{\Pr(\lambda u^{i**} + (1-\lambda)Q^{i*}_{n-1} < Q^{i**}_{n-1})} \\
= \; & \frac{F(K) - F((1-\varepsilon)K)}{F(K)} \\
= \; & 1 - \frac{F((1-\varepsilon)K)}{F(K)}
\end{aligned}$$

and:

$$\begin{aligned}
& \Pr(Q^i_n(s^{i*}) \in (u^i(s^{i**}, s^{-i**}), (1-\varepsilon)u^i(s^{i**}, s^{-i**}) + \varepsilon Q^i_{n-1}(s^{i**})) \mid B) \\
= \; & \frac{\Pr(u^{i**} < \lambda u^{i**} + (1-\lambda)Q^{i*}_{n-1} < (1-\varepsilon)u^{i**} + \varepsilon Q^{i**}_{n-1})}{\Pr(\lambda u^{i**} + (1-\lambda)Q^{i*}_{n-1} < Q^{i**}_{n-1})} \\
= \; & \frac{F(\varepsilon K)}{F(K)}
\end{aligned}$$

where $u^{i**} = u^i(s^{i**}, s^{-i**})$, $\lambda = \lambda^i_n(s^{i**})$, $Q^{i*}_{n-1} = Q^i_{n-1}(s^{i*})$, $Q^{i**}_{n-1} = Q^i_{n-1}(s^{i**})$, $F(x) = \Pr((1-\lambda) \leq x)$ for $x \in (0,1)$ and $K = \frac{Q^{i**}_{n-1} - u^{i**}}{Q^{i*}_{n-1} - u^{i**}}$. Notice that for any $K \in (0,1]$, $\frac{F(cK)}{F(K)} \in (0,1)$ and $\lim_{K\to 0} \frac{F(cK)}{F(K)} = \lim_{K\to 0} \frac{cf(cK)}{f(K)} = c$, where $c \in \{\varepsilon, 1-\varepsilon\}$, f is the density function for the weighting parameter and $f(0) < \infty$.

Therefore, for player i, with probability one, there exist infinitely many periods, n, such that $Q^i_n(s^{i*})$ is not too close to or not too far from $Q^i_n(s^{i**})$:

$$Q^i_n(s^{i*}) < \varepsilon u^i(s^{i**}, s^{-i**}) + (1-\varepsilon)Q^i_{n-1}(s^{i**})$$

and:

$$Q_n^i(s^{i*}) > (1-\varepsilon)u^i(s^{i**}, s^{-i**}) + \varepsilon Q_{n-1}^i(s^{i**})$$

In the following argument, we focus on the cases where both conditions hold when player i changes his action from s^{i*} to s^{i**}.

Now, we consider period n in which players have just switched to $(s^{i**})_{i\in N}$, so that both conditions above hold for player i. In the following argument, we consider the cases in which a coordination failure happens in period $n+1$. Notice that for the case:

$$Q_n^j(s^{j*}) \in [\varepsilon u^j(s^{j**}, s^{-j**}) + (1-\varepsilon)Q_n^j(s^{j**}), Q_n^j(s^{j**}))$$

for $j \neq i$, we have:

$$\Pr(Q_{n+1}^i(s^{i**}) \geq Q_n^i(s^{i*})) \times \ \Pr(Q_{n+1}^j(s^{j**}) < Q_n^j(s^{j*})) \geq e_{1,ij}$$

while for the case:

$$Q_n^j(s^{j*}) < \varepsilon u^j(s^{j**}, s^{-j**}) + (1-\varepsilon)Q_n^j(s^{j**})$$

for $j \neq i$, we have:

$$\Pr(Q_{n+1}^i(s^{i**}) < Q_n^i(s^{i*})) \times \ \Pr(Q_{n+1}^j(s^{j**}) \geq Q_n^j(s^{j*})) \geq e_{2,ij}$$

for some $e_{1ij} > 0$ and $e_{2,ij} > 0$. In any case, the probability that players fail to play the same strict Nash equilibrium in period $n+1$ has a positive probability, which has the lower bound, $\min_{h\in\{1,2\}} \min_{ij,i\neq j}\{e_{h,ij}\} > 0$. Since players change from $(s^{i**})_{i\in N}$ to $(s^{i*})_{i\in N}$ infinitely many times, players fail to play a strict Nash equilibrium with probability one, which contradicts the hypothesis. Therefore, the only possibility is that players play only one strict Nash equilibrium after some period. □

B. Proof of Proposition 7

We assume that each player's initial assessments of both actions are different. Then, the condition of coordination failure under the inertia condition for each player's tie break rule is as follows: for $j \neq k$ and (1) for the initial assessment, $Q_1^i(s_j^i) > Q_1^i(s_k^i)$ and $Q_1^{-i}(s_k^{-i}) > Q_1^{-i}(s_j^{-i})$; and (2) for any n, $Q_n^i(s_j^i) \geq Q_n^i(s_k^i)$ and $Q_n^{-i}(s_k^{-i}) \geq Q_n^{-i}(s_j^{-i})$, where, if one of the inequalities holds, then: (i) $Q_{n-1}^i(s_j^i) > Q_{n-1}^i(s_k^i)$ and $Q_{n-1}^{-i}(s_k^{-i}) > Q_{n-1}^{-i}(s_j^{-i})$; and (ii) $Q_{n+1}^i(s_j^i) < Q_{n+1}^i(s_k^i)$ and $Q_{n+1}^{-i}(s_k^{-i}) < Q_{n+1}^{-i}(s_j^{-i})$. Let $\hat{Q}_t^i(s_j^i)$ be the assessment of action s_j^i when only (s_j^i, s_k^{-i}) is played t times, where $j \neq k$. Then, it can be easily verified that the condition for coordination failure is equivalent to the following condition; for any $u, t \in \mathbb{N}$, $\hat{Q}_u^i(s_j^i) \geq \hat{Q}_t^i(s_k^i)$ and $\hat{Q}_u^{-i}(s_k^{-i}) \geq \hat{Q}_t^{-i}(s_j^{-i})$, where if one of inequalities holds, then $\hat{Q}_{u-1}^i(s_j^i) > \hat{Q}_t^i(s_k^i)$, $\hat{Q}_{u-1}^{-i}(s_k^{-i}) > \hat{Q}_t^{-i}(s_j^{-i})$, $\hat{Q}_{u+1}^i(s_j^i) < \hat{Q}_t^i(s_k^i)$ and $\hat{Q}_{u+1}^{-i}(s_k^{-i}) < \hat{Q}_t^{-i}(s_j^{-i})$ for $j \neq k$. Therefore, in the following proofs, we use the latter condition.

The important factor of this argument is that the players change actions at the same time and they "coordinate" at coordination failure. If the players coordinate on diagonal action profiles once, then they succeed in coordinating. Therefore, if the following conditions are satisfied, players never coordinate; for any m and $n \in \mathbb{N}$:

$$\frac{1}{n}Q_1^1(s_j^1) \geq \frac{1}{m}Q_1^1(s_k^1) \textit{ and } \frac{1}{n}Q_1^2(s_k^2) \geq \frac{1}{m}Q_1^2(s_j^2)$$

holds, where equalities among them do not hold consecutively; if one of the equalities holds at m, n, then both inequalities hold strictly at m, $n-1$ and m, $n+1$. In the following argument, we show that this condition is equivalent to the following condition:

$$\frac{Q_1^1(s_2^1)}{Q_1^1(s_1^1)} = \frac{Q_1^2(s_1^2)}{Q_1^2(s_2^2)}$$

To make this clear, we assume first that if one of the inequalities holds with equality, then both inequalities should hold with equality. Then, the original condition above can be expressed as follows:

$$Q_1^1(s_2^1) < \frac{m}{n}Q_1^1(s_1^1) \text{ and } Q_1^2(s_2^2) > \frac{n}{m}Q_1^2(s_1^2)$$

or:

$$Q_1^1(s_2^1) > \frac{m}{n}Q_1^1(s_1^1) \text{ and } Q_1^2(s_2^2) < \frac{n}{m}Q_1^2(s_1^2)$$

or:

$$Q_1^1(s_2^1) = \frac{m}{n}Q_1^1(s_1^1) \text{ and } Q_1^2(s_2^2) = \frac{n}{m}Q_1^2(s_1^2).$$

Note that if $\frac{Q_1^1(s_2^1)}{Q_1^1(s_1^1)}$ is a rational number, then there exist m and n, such that $\frac{Q_1^1(s_2^1)}{Q_1^1(s_1^1)} = \frac{m}{n}$. By the last condition, we should have $\frac{Q_1^2(s_2^2)}{Q_1^2(s_1^2)} = \frac{n}{m}$, that is, $\frac{Q_1^2(s_2^2)}{Q_1^2(s_1^2)}$ should be a rational number, too. If $\frac{Q_1^1(s_2^1)}{Q_1^1(s_1^1)}$ is an irrational number, then $\frac{Q_1^2(s_2^2)}{Q_1^2(s_1^2)}$ should be also an irrational number. If $\frac{Q_1^1(s_2^1)}{Q_1^1(s_1^1)} \neq \frac{Q_1^2(s_1^2)}{Q_1^2(s_2^2)}$, say if $\frac{Q_1^1(s_2^1)}{Q_1^1(s_1^1)} > \frac{Q_1^2(s_1^2)}{Q_1^2(s_2^2)}$, then there exists a rational number, $\frac{m}{n}$, such that $\frac{Q_1^1(s_2^1)}{Q_1^1(s_1^1)} > \frac{m}{n} > \frac{Q_1^2(s_1^2)}{Q_1^2(s_2^2)}$. This means that $Q_1^1(s_2^1) > \frac{m}{n}Q_1^1(s_1^1)$ and $Q_1^2(s_2^2) > \frac{n}{m}Q_1^2(s_1^2)$, and it contradicts the conditions above. Hence, the following relation, $\frac{Q_1^1(s_2^1)}{Q_1^1(s_1^1)} = \frac{Q_1^2(s_1^2)}{Q_1^2(s_2^2)}$, is the only case that satisfies the condition above.

Now consider the other cases. There exist m and $n \in \mathbb{N}$, such that $\frac{1}{n}Q_1^i(s_1^i) = \frac{1}{m}Q_1^i(s_2^i)$ and $\frac{1}{n}Q_1^j(s_2^j) \neq \frac{1}{m}Q_1^j(s_1^j)$, say $\frac{1}{n}Q_1^j(s_2^j) > \frac{1}{m}Q_1^j(s_1^j)$.[17] Then:

$$\frac{1}{n-1}Q_1^i(s_1^i) > \frac{1}{m}Q_1^i(s_2^i) \text{ and } \frac{1}{n-1}Q_1^j(s_2^j) > \frac{1}{m}Q_1^j(s_1^j)$$

and:

$$\frac{1}{n+1}Q_1^i(s_1^i) < \frac{1}{m}Q_1^i(s_2^i) \text{ and } \frac{1}{n+1}Q_1^j(s_2^j) < \frac{1}{m}Q_1^j(s_1^j)$$

should hold. Notice that $\frac{Q_1^i(s_1^i)}{Q_1^i(s_2^i)}$ should be a rational number, $\frac{n}{m}$. Moreover, by the conditions above, we have $\frac{n+1}{m} > \frac{Q_1^j(s_2^j)}{Q_1^j(s_1^j)} > \frac{n}{m}$. It is easy to see that at $2n$ and $2m$, the following conditions also satisfy: $\frac{1}{2n}Q_1^i(s_1^i) = \frac{1}{2m}Q_1^i(s_2^i)$ and $\frac{1}{2n}Q_1^j(s_2^j) > \frac{1}{2m}Q_1^j(s_1^j)$. Thus, we have the following condition: $\frac{2n+1}{2m} > \frac{Q_1^j(s_2^j)}{Q_1^j(s_1^j)} > \frac{2n}{2m}$. Using the same logic, the condition should be satisfied for any kn and km, where $k \in \mathbb{N}$.

---

[17] If $\frac{1}{n}Q_1^j(s_2^j) < \frac{1}{m}Q_1^j(s_1^j)$, then:

$$\frac{1}{n}Q_1^i(s_1^i) < \frac{1}{m-1}Q_1^i(s_2^i) \text{ and } \frac{1}{n}Q_1^j(s_2^j) < \frac{1}{m-1}Q_1^j(s_1^j)$$

If $k \to \infty$, then the condition becomes as follows; $\frac{n}{m} \geq \frac{Q_1^j(s_2^j)}{Q_1^j(s_1^j)} > \frac{n}{m}$. However, there do not exist initial assessments that satisfy this condition.[18] Therefore, the necessary and sufficient condition for initial assessments for the coordination failure in this case is equivalent to the following condition:

$$\frac{Q_1^1(s_2^1)}{Q_1^1(s_1^1)} = \frac{Q_1^2(s_1^2)}{Q_1^2(s_2^2)}.$$

□

C. Proof of Proposition 8

It can be shown that the following condition is equivalent to the condition for the coordination failure in the coordination game; for any t, there exists u such that:

$$\hat{Q}_t^i(s_1^i) \in (\hat{Q}_{u+1}^i(s_2^i), \hat{Q}_u^i(s_2^i)] \; and \; \hat{Q}_t^{-i}(s_2^{-i}) \in (\hat{Q}_{u+1}^{-i}(s_1^{-i}), \hat{Q}_u^{-i}(s_1^{-i})]$$

or:

$$\hat{Q}_t^i(s_1^i) \in [\hat{Q}_{u+1}^i(s_2^i), \hat{Q}_u^i(s_2^i)) \; and \; \hat{Q}_t^i(s_2^{-i}) \in [\hat{Q}_{u+1}^{-i}(s_1^{-i}), \hat{Q}_u^{-i}(s_1^{-i}))$$

for all i.[19] Since $\hat{Q}_t^i(s_j^i) = (1-\lambda)^t \hat{Q}_0^i(s_j^i)$, the condition in Proposition 8 can be easily derived. □

Acknowledgments

The author would like to thank Rajiv Sarin for his support and guidance over many years. I am also grateful to Antonella Iannni, Indrajit Ray, Rodrigo Velez and seminar audiences at Texas A& M University, the University of Birmingham and UECE Lisbon meetings, 2012, for helpful comments and suggestions.

Conflicts of Interest

The author declares no conflict of interest.

References

1. Erev, I.; Roth, A. Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *Am. Econ. Rev.* **1998**, *88*, 848–881.
2. Sarin, R.; Vahid, F. Payoff assessments without probabilities: A simple dynamic model of choice. *Games Econ. Behav.* **1999**, *28*, 294–309.

---

[18] If $\frac{1}{n}Q_1^j(s_2^j) < \frac{1}{m}Q_1^j(s_1^j)$, then it satisfies that $\frac{n+1}{m} < \frac{Q_1^j(s_2^j)}{Q_1^j(s_1^j)} < \frac{n}{m}$. By the same argument, There exist no initial assessments that satisfy the conditions.

[19] For example, if $\hat{Q}_m^i(s_1^i) > \hat{Q}_0^i(s_2^i)$, then we assume that $\hat{Q}_{-1}^i(s_2^i)$ is the maximum payoff, which both actions give, so that $\hat{Q}_m^i(s_1^i) \in (\hat{Q}_0^i(s_2^i), \hat{Q}_{-1}^i(s_2^i)]$. In addition, let $\hat{Q}_\infty^i(s_2^i)$ be the minimum payoff that both actions give and $\hat{Q}_{\infty+1}^i(s_2^i)$ be the minimum payoff of those which both actions give. Then, if $\hat{Q}_m^i(s_1^i) \leq \hat{Q}_\infty^i(s_2^i)$, $\hat{Q}_m^i(s_1^i) \in (\hat{Q}_{\infty+1}^i(s_2^i), \hat{Q}_\infty^i(s_2^i)]$.

3. Beggs, A.W. On the convergence of reinforcement learning. *J. Econ. Theory* **2005**, *122*, 1–36.
4. Laslier, J.F.; Topol, R.; Walliser, B. A behavioral learning process in games. *Games Econ. Behav.* **2001**, *37*, 340–366.
5. Cooper, R.W.; DeJong, D.V.; Forsythe, R.; Ross, T.W. Selection criteria in coordination games: Some experimental results. *Am. Econ. Rev.* **1990**, *80*, 218–233.
6. Van Huyck, J.; Battalio, R.C.; Beil, R.O. Tacit coordination games, strategic uncertainty, and coordination failure. *Am. Econ. Rev.* **1990**, *80*, 234–248.
7. Sarin, R. Simple play in the prisoner's dilemma. *J. Econ. Behav. Organ.* **1999**, *40*, 105–113.
8. Fudenberg, D.; Levine, D.K. *The Theory of Learning in Games*; MIT Press: Cambridge, MA, USA, 1998.
9. Monderer, D.; Shapley, L.S. Fictitious play for games with identical interests. *J. Econ. Theory* **1996**, *68*, 258–265.
10. Leslie, D.S.; Collins, E.J. Individual q-Learning in normal form games. *SIAM J. Control Optim.* **2005**, *44*, 495–514.
11. Cominetti, R.; Melo, E.; Sorin, S. A payoff-based learning and its application to traffic games. *Games Econ. Behav.* **2010**, *70*, 71–83.
12. Sarin, R.; Vahid, F. Predicting how people play games: A simple dynamic model of choice. *Games Econ. Behav.* **2001**, *34*, 104–122.
13. Chen, Y.; Khoroshilov, Y. Learning under limited information. *Games Econ. Behav.* **2003**, *44*, 1–25.
14. Camerer, C.; Ho, T.H. Experience-weighted attraction learning in normal form games. *Econometrica* **1999**, *67*, 827–874.
15. McKelvey, R.; Palfrey, T.R. Quantal response equilibria for normal form games. *Games Econ. Behav.* **1995** *10*, 6–38.