Control and Cybernetics

vol. **35** (2006) No. 2

Learning in games with bounded memory

by

Jaideep Roy

Department of Economics, LUMS
Lancaster University
Lancaster LA1 4YX, United Kingdom
e-mail: jroy@lancaster.ac.uk

Abstract: The paper studies infinite repetition of finite strategic form games. Players use a backward looking learning behavior and face bounds on their cognitive capacities. We show that for any given belief-probability over the set of possible outcomes where players have no experience, games can be payoff classified and there always exists a stationary state in the space of action profiles. In particular, if the belief-probability assumes all possible outcomes without experience to be equally likely, in one class of Prisoners' Dilemmas where the uniformly weighted average defecting payoff is higher than the cooperative payoff and the uniformly weighted average cooperative payoff is lower than the defecting payoff, play converges in the long run to the static Nash equilibrium while in the other class of Prisoners' Dilemmas where the reverse holds, play converges to cooperation. Results are applied to a large class of 2×2 games.

Keywords: cognitive complexity, bounded logistic quantal response learning, long run outcomes.

1. Introduction

The Prisoners' Dilemma represents social situations characterized by the existence of what is commonly known as the free rider problem. Let us consider two examples from the author's hometown. The city of Calcutta is renowned in India for *Durga Puja*, a religious festival during the month of October. There is hardly any corner of the city without a *pandel* (a Bengali term for a temporary tent-like construction) with idols of *Durga* the mother goddess, and her entire family. The festival is financed by collecting private donations from households of respective neighborhoods. Over the years, more and more Calcuttans have voluntarily contributed successively higher donations. On the other hand, the city is also renowned (and this time all over the world) for her neglected streets and localities. Some local private organizations have made efforts to collect private donations from member families of respective neighborhoods for the maintenance of their localities. Such efforts have repeatedly failed in most of the neighborhoods where such organizational efforts had been undertaken. Some cases showed complete failure from the very beginning while others eventually ceased to exist as the number of private contributors became negligible over time. When asked individually however, every citizen of Calcutta showed great concern and enthusiasm regarding the cleanliness of their localities.

These two social situations, the likes of which are surely widespread, can be represented by a model of infinitely repeated Prisoners' Dilemma where at each period players decide whether or not to contribute to a public good. What is surprising is that while in the former we see that agents cooperate and achieve the Pareto dominant outcome (and the festival of *Durga Puja* is celebrated with great enthusiasm), in the latter the same agents behave as predicted by the static Nash equilibrium (and Calcutta remains neglected forever). Do we have a unified theory that is able to explain this difference in social behavior? To have one we need to be able to distinguish one Prisoners' Dilemma from another. Is the Prisoners' Dilemma representing the former example different from the one representing the latter?

As we will argue, a large part of the existing literature on repeated games does not answer this question. Nevertheless, since cooperative behavior is observed quite often, many authors have tried to explain convergence to such Pareto dominant outcomes in similar settings. The general intuition has been that when players are rational and sufficiently patient, almost any feasible payoff allocation that gives each player at least his minimum security level can be realized in an equilibrium of any repeated game. In the game theory literature, these feasibility theorems have been referred to as *folk theorems*. Naturally, the predictive power of these results is rather weak. Neyman (1985) shows that bounded complexity on part of the players justifies cooperation in finitely repeated Prisoners' Dilemma. Recently, the literature has taken a turn towards modeling ways by which players may actually learn to play games and see if such learning behaviors indeed lead to some equilibrium or stationary state.

Bendor et al. (1995) show that if both players are aspiration driven and aspirations are static, players may exhibit long run cooperative behavior under certain initial conditions. Karandikar et al. (1998) show that with evolving aspirations, players will cooperate most of the time in a large class of games which includes the Prisoners' Dilemma. Kim (1999) studies a satisficing model of learning where a 2×2 game is repeated by case-based players a la Gilboa and Schmeidler (1995) and shows that infrequent and simultaneous experiments with cooperative strategies are essential for convergence to cooperative behavior in the Prisoners' Dilemma.

However, as mentioned before, most of these models are not sufficient in answering why the same group of players may in some Prisoners' Dilemma converge to cooperative behavior while in some other to the Nash equilibrium. One may build up a model with two stationary long run outcomes, one where players cooperate and the other where players defect and conclude that such a diversity is a matter of chance and the evolution of play. We do not think this approach is satisfactory in answering our question. Given a specified behavior rule, we need a theory that is able to classify Prisoners' Dilemmas into groups in terms of observable variables and show that players converge to cooperation with probability one in all Prisoners' Dilemmas falling in one such group while they converge with probability one to the static Nash equilibrium in all Prisoners' Dilemmas falling in the other.

In order to resolve this apparent puzzle, we study infinite repetition of a class of games (our initial and fundamental analysis is, however, kept general) which all have the basic essence of the Prisoners' Dilemma that ensures existence of the free rider problem as cited in the examples above. Players are assumed to learn to play the game over time from past experience. The learning model used by players is a variant of *logistic quantal response learning* (LQRL) a la McKelvey and Palfrey (1995). We assume that players have cognitive bounds in analyzing histories of past plays in order to decide upon their current choice probabilities. We call this learning mechanism *bounded logistic quantal response learning* (BLQRL). Our model captures the fact that players cannot always look very long in the past if the past is also very heterogenous and therefore (we call it) cognitively complex.

Our notion of complexity in decision making differs significantly from that of implementational complexity and finite automata as in Kalai and Stanford (1988) or Binmore and Samuelson (1992), among others, which is typically forward looking and can be roughly thought of as equal to the number of states that a repeated game strategy can induce. The notion of complexity used in this paper is backward looking and is close to the basic idea of finite memory. However, we justify it for two reasons. Firstly, players forget distant past. In addition, the rate at which they forget the past increases with the heterogeneity of realized history. This implies that players not only have finite memory but the length of their memory evolves probabilistically, conditional on future realization of actions which carve the path of play, thus making the length of their memory path dependent. Furthermore, players are parameterized by their degree of rationality in deciding upon the choice probabilities. McKelvey and Palfrey (1995) show that in any finite game where players use the quantal response learning, play converges to a Nash equilibrium when this rationality parameter approaches infinity. Thus, in our model we expect play to converge to the Nash equilibrium when complexity bounds and the rationality parameter go to infinity, as essentially then the two models are equivalent. We show that (i) if players face cognitive bounds, for any finite stage game, the play converges to stationary probabilities on the space of action profiles for all values of the rationality parameter; (ii) in addition, for the Prisoners' Dilemma, when the rationality parameter approaches infinity, the play converges to the Nash defect with probability one if the weighted average defecting payoff is higher than the cooperative one and the weighted average cooperative payoff is lower than the defecting one, where the weights are determined by players' beliefs regarding outcomes without experience. (The weighted average defecting payoff is simply the expected value of individual payoff from a given pure strategy of a normal form game under the assumption that other players in the game play all pure strategies with equal probabilities). However, if the reverse is true, that is if the weighted average defecting payoff is lower than that of cooperation and the weighted average cooperative payoff is higher than that of defection (which is possible in some class of Prisoners' Dilemma games), the play converges to cooperative outcome with probability one. In comparison with the result obtained in McKelvey and Palfrey (1995) where they show that with infinite memory, as rationality increases unboundedly, quantal response learning behavior almost always select a Nash equilibrium, our analysis suggests that the assumption of infinite memory is crucial in their results as we show that if memory is finite, then even if rationality is increased unboundedly, players may indeed converge to other (non-Nash) outcomes, and in particular to the cooperative outcome in some Prisoners' Dilemma games. This is the main contribution of this paper and perhaps throws some light behind why Durga Puja goes on while at the same time some parts of Calcutta remain untidy. Whether the conditions asserted in this paper hold in reality in these two examples is still an empirical question.

We then apply the general results to study long run behavior in a large class of 2 × 2 games which includes Pure Coordination, Common Interest, and Chicken. Our results call for experiments where subjects play 2×2 games studied in this paper and satisfying restrictions as demanded by the theory provided. One easy way of capturing and controlling the notion of cognitive bounds in such an experimental setup could be to impose *time restrictions* within which subjects need to decide upon their current actions.

The rest of the paper is organized as follows. In Section 2 we formally state the model. In Section 3 we analyze exclusively the issues regarding memory and information sets, establishing some properties of an appropriate stochastic process which help us prove the two convergence results in Section 4. Section 5 uses these convergence results in selecting out unique long run outcomes in Prisoners' Dilemmas under appropriate payoff restrictions and suggests a possible resolution to the paradox described above. Section 6 studies other 2 × 2 games and applies the general results. An informal discussion on beliefs without experience is provided in Section 7. Finally the paper draws its conclusions in Section 8.

2. The general environment

Consider the following 2×2 game Γ :

$$\begin{array}{ccc|c|c|}
 & & \multicolumn{2}{c}{\text{player } 2} \\
 & & \multicolumn{1}{c}{C} & \multicolumn{1}{c}{D} \\
\cline{3-4}
\text{player } 1 & C & \beta, \beta & \varepsilon, \alpha \\
\cline{3-4}
 & D & \alpha, \varepsilon & \theta, \theta \\
\cline{3-4}
\end{array}$$

where $\varepsilon, \theta, \beta$ and α are all assumed to be positive. With $\infty > \alpha > \beta > \theta > \varepsilon > 0$, Γ is the celebrated Prisoners' Dilemma. (The assumption that all entries in the payoff matrix are positive is discussed in Subsection 2.2). Γ is repeated infinitely between two players, player 1 and player 2. The action sets are $A_1 = A_2 = \{C, D\}$ with $A = \{C, D\}^2, a \in A$. Let $u_i(t) : A \to \{\varepsilon, \theta, \beta, \alpha\}$ be the period t payoff function for player i. By $\Gamma^t(A, u)$ we will mean the t-th repetition of $\Gamma(A, u), t = 1, 2, ...$. Let $H^t = A^{t-1}$ (the $(t-1)-$fold Cartesian product of A) be the set of all possible sequences of realized action pairs till the beginning of period t. Let $h^t \in H^t$ and denote $a(t) \in A$ as the action pair realized at period t. With the first round of play being $t = 1$, it is clear that $H^1 = \emptyset$, $H^2 = A$, $H^3 = A^2$, etc. We may also think of h^t as a vector with $(t-1)$ components, each of which are elements of A. Furthermore, we may also think of each element a of A as a vector with two components. Interpreting histories and action profiles as vectors in appropriate spaces will be helpful in what follows. Typically, a repeated game strategy for player i at date t is a function $f_i^t : H^t \rightrightarrows A_i$. This implicitly assumes that players use entire histories in order to decide upon their current actions and therefore are able to use very long as well as possibly complicated histories as information sets. We would like to rule this out. So, suppose instead that players have cognitive bounds and therefore can only analyze histories of a certain cognitive complexity. For any $1 \leq \tau_b \leq \tau_a \leq t$, denote by $h^t(\tau_b, \tau_a) \subset h^t$ a segment of h^t whose first period is τ_b and last period is τ_a. Under some abuse of notation, let $H^t(\tau_b, \tau_a)$ be the set of all possible segments of all possible elements of H^t.

2.1. Cognitive complexity

Assume that players are unable to recognize patterns of histories. We agree that this is a relatively strong assumption and hope to relax it in future research. An interested reader may see Sonsino (1997) where a model of learning with the possibility of pattern recognition is studied. The cognitive complexity of a given history, or segment of history, is determined by its length and by its variability. The length of a history is straightforward to define, as it is simply the number of elements appearing in the history. Formally, length is an integer valued function $\ell : H^t(\tau_b, \tau_a) \to \mathbb{Z}_+$ defined as $\ell(h^t(\tau_b, \tau_a)) = \tau_a - \tau_b + 1$.

To obtain a formal definition of variability, we start defining, for a given segment of history $h^t(\tau_b, \tau_a)$ the *finest partition* $\mathcal{P}(h^t(\tau_b, \tau_a))$ as the partition putting in the same set identical elements. For example, if

$$\begin{aligned} h^t(\tau_b,\tau_a) &= \{a,a,a',a,a'',a,a'\}\,, \text{ then} \\ \mathcal{P}(h^t(\tau_b,\tau_a)) &= \{\{a,a,a,a\}\,,\{a',a'\}\,,\{a''\}\} \end{aligned}$$

and if

$$\begin{aligned} h^t(\tau_b,\tau_a) &= \{a,a,a,a,a,a\}\,, \text{ then} \\ \mathcal{P}(h^t(\tau_b,\tau_a)) &= \{\{a,a,a,a,a,a\}\}\,. \end{aligned}$$

The measure of variability we use is simply the cardinality of this finest partition minus 1. Formally, variability is an integer valued function $v : H^t(\tau_b,\tau_a) \to \mathbb{Z}_+$ defined as $v(h^t(\tau_b,\tau_a)) = |\mathcal{P}(h^t(\tau_b,\tau_a))| - 1$. Thus, variability is directly related to the number of different elements appearing in a given history. From the above examples,

$$\begin{aligned} v\left(\{a,a,a',a,a'',a,a'\}\right) &= 2 \text{ and} \\ v\left(\{a,a,a,a,a,a\}\right) &= 0. \end{aligned}$$

Given any history segment $h^t(\tau_b,\tau_a)$, we define the cognitive complexity (*ccomp*) of $h^t(\tau_b,\tau_a)$ as a function

$$ccomp\left(h^t(\tau_b,\tau_a)\right) = G\left(\ell(h^t(\tau_b,\tau_a)), v(h^t(\tau_b,\tau_a))\right) \qquad (1)$$

with $G(\cdot)$ strictly increasing in $\ell(h^t(\tau_b,\tau_a))$ and $v(h^t(\tau_b,\tau_a))$. In this paper we adopt the following linear functional form for $G(\cdot)$:

$$G\left(\ell(h^t(\tau_b,\tau_a)), v(h^t(\tau_b,\tau_a))\right) = \ell(h^t(\tau_b,\tau_a)) + v(h^t(\tau_b,\tau_a)), \qquad (2)$$

this form not affecting the basic spirit of the results obtained.

The cognitive complexity of a history segment simply equals the length of the segment plus its degree of heterogeneity. More generally, our definition of cognitive complexity implies that cognitive complexity of an information set increases with the cardinality of the data set and the variation or heterogeneity of its composition. Note, however, that sequences of realized action profiles with very simple patterns (like Tit-for-Tat for example) may turn out to be significantly heterogenous.

2.2. Behavior rules for players

Players face cognitive bounds which equal a positive integer ζ and is the same across players. They are assumed to look at the most recent history segments. For any repetition period t, let $\hat{h}^t \subset h^t$ be a segment of h^t satisfying the following two conditions:

$$(i)\ ccomp(\hat{h}^t) \leq \zeta \text{ and } (ii)\ \hat{h}^t \in \underset{h^t(\tau_b,t-1)}{\arg\max} \left\{ccomp(h^t(\tau_b,t-1))\right\}. \qquad (3)$$

Thus, $\hat{h}^t$ is the most recent history segment (since the most recent period in this segment is $t-1$) with the highest cognitive complexity that players can analyze at time t given their cognitive bound ζ. We assume that players always look at history segments $\hat{h}^t$. To avoid abuse of notation, let $\vec{\in}$ denote the relation 'is a component of the vector' and $\vec{\notin}$ the negation of $\vec{\in}$. Let $|A_i| = N$ denote the number of actions available to player $i = 1, 2$ (in the case of Prisoners' Dilemma, $N = 2$ for both players). For any period t and for all $k = 1, 2, ..., N$ such that $\exists a \vec{\in} \hat{h}^t$ with $a_i^k \vec{\in} a$, define $\pi_i^k(t)$as

$$\pi_i^k(t) = \frac{\sum\limits_{\left\{a \vec{\in} \hat{h}^t | a_i^k \vec{\in} a\right\}} u_i(a)}{\left|\left\{a \vec{\in} \hat{h}^t \mid a_i^k \vec{\in} a\right\}\right|},\ i = 1, 2 \tag{4}$$

and for all $k = 1, 2, ..., N$ such that $\nexists a \vec{\in} \hat{h}^t$ with $a_i^k \vec{\in} a$, define $\pi_i^k(t)$ as

$$\pi_i^k(t) = \sum_{r=1}^{N} w_i\left(k, r\right) u_i(a_i^k, a_j^r),\ i, j = 1, 2, \tag{5}$$

where $w_i\left(k, r\right) \geq 0$ is the belief probability held by player i regarding the outcome (a_i^k, a_j^r) when she herself plays a_i^k, with $\sum\limits_{r=1}^{N} w_i\left(k, r\right) = 1$ for all i. Eqs. (4) and (5) define what we will refer to as perceived payoffs of each action. In this paper we will work with the special case where $w_i\left(k, r\right) = w_i\left(k, r'\right)$ for every $r, r' \in \{1, 2, ..., N\}$. This is the case when players believe that all outcomes which are possible when they use strategies with no experience are equally likely. The results thus obtained with this simplification can in spirit be generalized. An extensive discussion on this issue is provided in Section 7.

Eq. (4) says that if there is some element a in the history segment $\hat{h}^t$ such that player i has taken action a_i^k, then player i evaluates the perceived payoff of playing a_i^k in the subsequent period as the arithmetic mean of payoffs obtained in all cases in which he played a_i^k in $\hat{h}^t$. For example, consider Γ and let

$$\hat{h}^t = \{(C, D), (C, C), (C, C), (D, C), (D, D)\}.$$

Then, $\pi_1^C(t) = \left(2\beta + \varepsilon\right)/3$, $\pi_1^D(t) = \left(\alpha + \theta\right)/2$, $\pi_2^C(t) = \left(2\beta + \varepsilon\right)/3$ and $\pi_2^D(t) = \left(\alpha + \theta\right)/2$. On the other hand, if $\hat{h}^t$ does not contain any action profile in which player i takes action a_i^k, the perceived payoff of action a_i^k is simply the arithmetic mean of all possible payoffs to player i. For example, let

$$\hat{h}^t = \{(C, C), (C, D), (C, C), (C, C), (C, D)\}.$$

Then, $\pi_1^D(t) = \left(\alpha + \theta\right)/2$.

To distinguish our way of computing perceived payoffs from that in the existing literature on quantal response learning, let $t = 5$, $\zeta = 2$ and consider the history

$$h^t = \{(C, C), (D, D), (C, D), (D, C)\} .$$

From Eq. (3), $\hat{h}^t = \{(D, C)\}$. Then, in our formulation,

$$\pi_1^C(t) = (\beta + \varepsilon)/2, \pi_1^D(t) = \alpha, \pi_2^C(t) = \varepsilon, \text{ and } \pi_2^D(t) = (\alpha + \theta)/2$$

while in the standard formulation, as used in McKelvey and Palfrey,

$$\pi_1^C(t) = (\beta+\varepsilon)/2, \pi_1^D(t) = (\theta+\alpha)/2, \pi_2^C(t) = (\beta+\varepsilon)/2 \text{ and } \pi_2^D(t) = (\theta+\alpha)/2,$$

since in the history h^t, player 1 has played C two times, once receiving β (when player 2 also played C) and the other time receiving ε (since there player 2 had played D) and hence the perceived payoff held by player 1 from the action C is simply $(\beta + \varepsilon)/2$. The rest of the perceived payoffs in the standard set up is computed analogously. Notice that with cognitive bounds, player 2 thinks that C is a relatively unfruitful action. This is because he forgets (or, in a time-constrained experimental setup, does not have enough time to realize) that C had actually yielded a high payoff equal to β when his opponent also played C. This is not the case with unbounded cognition as then players will be able to keep account of all past experiences.

After computing $\pi_i^k(t)$ for every k, players assign probabilities with which each action is chosen for play in period t. Let $\sigma_i^k(t)$ be the choice probability assigned by player i to action a_i^k at period t. Using the Logit framework (this is in fact a "Luce" model of choice, see Luce, 1959):

$$\sigma_i^k(t) = \frac{\left[\pi_i^k(t)\right]^\lambda}{\sum\limits_{k'=1}^{N} \left[\pi_i^{k'}(t)\right]^\lambda} \tag{6}$$

with $\lambda \in [0, \infty]$ being a parameter measuring the degree of rationality of the players. By the use of the term 'rationality' we mean closeness of players' choice probabilities to those of the myopic best response choices. The functional form of $\sigma_i^k(t)$ as in Eq. (6) is also used in Chen, Thisse and Friedman (1997). This formulation is invariant to linear transformations of the perceived payoffs. However, it is undefined if for a history segment, the sum of payoff experiences is zero for at least one player. To avoid this problem, we assume that the payoff matrix of Γ has only positive entries. An alternative formulation, adopted by McKelvey and Palfrey (1995), is

$$\sigma_i^k(t) = \frac{e^{\lambda \pi_i^k(t)}}{\sum\limits_{k'=1}^{N} e^{\lambda \pi_i^{k'}(t)}} \, . \tag{7}$$

This formulation has the advantage of taking care of negative payoffs but is not invariant to linear transformations. In appendix A.3 we show that our results hold with this alternative formulation as well.

Notice that $\hat{h}^t$ is the memory of the players and $\ell\left(\hat{h}^t\right)$ is a stochastic process determined by the evolution of play. In the next section, we will study the nature of this stochastic process. As far as the rationality parameter is concerned, it is straightforward to see that if $\lambda = 0$, players always play each action with probability equal to $1/N$. However, as λ increases, actions with better past experiences receive higher choice probabilities ($\lambda = \infty$ coinciding with myopic expected payoff maximization). From a statistical point of view, players may be thought of as committing errors in deciding upon the choice probabilities. Then, λ varies inversely with these errors.

Mention must also be made of the fact that players always have the option of using some belief updating mechanism (like the moving averages technique) which may only require them to keep track of few basic statistics. In such settings, the model studied in this paper can be approximated to cases where players make errors in such updating with such errors decreasing in ζ and λ. In our case, with $\lambda = 0$, there is a unique equilibrium at the centroid of the $(N^2 - 1)$-dimensional simplex of probability measures over action profiles for any value of ζ. On the other extreme, McKelvey and Palfrey (1995) show that with $\zeta = \infty$, as $\lambda \to \infty$, every Logit equilibrium approaches the Nash equilibrium of the static game. The above kind of probability choice function is commonly known as logistic quantal response function ($LQRF$) as in McKelvey and Palfrey (1995). The learning process with which we deal here is therefore a variant of the logistic quantal response learning ($LQRL$) of Mookherjee and Sopher (1997) in the sense that information sets are not any more the entire history that players confront; rather they are only segments of these histories up to what cognitive bounds enable players to look at. We therefore call our learning process bounded logistic quantal response learning ($BLQRL$). Mookherjee and Sopher bring such a learning model to experimental test (incidentally, I was one of the subjects of that experiment held at the Delhi School of Economics) and their data revealed some support to this process of behavior. Note that in our model, players have two bounds. The first one is a bound on their cognitive capability by which they are unable to analyze history segments of complexity higher than this bound. The second one is a bound on their rationality in computing choice probabilities while looking at these history segments. Two players may have enough cognitive capacity to analyze long history segments but one who is endowed with a higher value of λ will play more frequently those actions which yielded higher payoffs in the past.

3. Analysis of memory and stochastic information sets

In this section we show that the stochastic process generated by the behavior rules of players in $\Gamma^{\infty}(A, u)$ can be represented by a Markov chain for an appropriate state space defined below when the cognitive bound faced by players is finite.

In order to precisely state this Markov chain, we begin with the following two lemmas.

LEMMA 3.1 *For any repetition period t, for any $h^t \in H^t$, and for any $\zeta < \infty$, $\zeta - 1 \leq ccomp(\hat{h}^t) \leq \zeta$.*

Proof. The right hand inequality is straightforward. Suppose the left hand inequality is violated and that (without loss of generality) $ccomp(\hat{h}^t) = \zeta - 2$. Increase $\ell(\hat{h}^t)$ to $\ell(\hat{h}^t) + 1$ by including the last element of h^t that did not enter $\hat{h}^t$. Call this extension $\hat{h}^{t\prime}$. Then, $v(\hat{h}^{t\prime})$ either equals $v(\hat{h}^t)$ or is equal to $v(\hat{h}^t) + 1$. It suffices to consider the case $v(\hat{h}^{t\prime}) = v(\hat{h}^t) + 1$. Then,

$$ccomp(\hat{h}^{t\prime}) = ccomp(\hat{h}^t) + 2 = \zeta$$

and therefore

$$\hat{h}^t \notin \left\{ \underset{h^t(\tau_b, t-1)}{\arg\max} \left\{ ccomp(h^t(\tau_b, t-1)) \right\} \mid ccomp(h^t(\tau_b, t-1)) \leq \zeta \right\},$$

a contradiction. ■

Although players can analyze histories with cognitive complexity equal to ζ, the evolution of play may be such that players are not able to use their full cognitive capacity at each repetition period because there may exist moments of play when, if they try to look at one more element of the history they confront, they get confused (or they run out of time in time-constrained experiments). In other words, there may be periods where there is no immediate history segment that matches exactly their cognitive bound and therefore players have no other choice but to analyze that segment which has the highest cognitive complexity given their cognitive bound. However, the above lemma also shows that there will never come a period where they will under utilize their capacity beyond one unit of their cognition.

For any given positive integer x, define the set $H(x) := \big\{ h \in \bigcup_{t=1}^{x+1} H^t \mid \ell(h) + v(h) = x \big\}$. To understand the set $H(x)$, fix such an integer x (say $x = 2$). Consider the sets of histories H^1, H^2 and H^3 and take their union. Then,

$$\bigcup_{t=1}^{(x=2)+1} H^t \equiv H^1 \cup H^2 \cup H^3 = \left\{ \begin{array}{c} \underbrace{\varnothing}_{H^1}, \underbrace{\{a\}, \{a'\} ....., \{a''\} ...,}_{H^2\text{: all elements from } A} \\ \underbrace{\{a, a\}, \{a', a'\}, \{a, a'\}, ..., \{a'', a'''\} ...}_{H^3\text{: all elements from } A^2} \end{array} \right\}.$$

From the above collection, now pick all elements such that the cognitive complexity of each element picked is exactly equal to 2. The collection you end up with is $H\left(2\right)$ which in this case will constitute all elements in A^2 which have the same element of A twice, that is $\{a,a\},\{a',a'\},\{a'',a''\}$ and so on. The process is identical for any $x \geq 1$. Let $\mathcal{B}(\zeta) := \bigcup\limits_{\zeta-1\leq x\leq \zeta} H(x)$. Given Lemma (3.1), $\mathcal{B}(\zeta)$ is the set of all possible sequences of elements of A with cognitive complexities between $\zeta-1$ and ζ and therefore can be thought of as the set of all possible $\hat{h}^t$ for all periods t. In other words, at any period the information set used by players must belong to $\mathcal{B}(\zeta)$.

As A is finite and $\zeta < \infty$, $\mathcal{B}(\zeta)$ is a finite set (since $\ell(\hat{h}) \leq \zeta\ \forall \hat{h} \in \mathcal{B}(\zeta)$) and we call $\mathcal{B}(\zeta)$ the state space.

Definition 3.1 *Given any $\hat{h} \in \mathcal{B}(\zeta)$, $\hat{h}' \in \mathcal{B}(\zeta)$ is an immediate successor of $\hat{h}$, denoted by ${}^{(1)}(\hat{h}' \mid \hat{h})$, iff (i) the $\ell(\hat{h}')-$th component of $\hat{h}'$ is a new far right addition to $\hat{h}$ and (ii) for each integer $0 \leq n \leq \ell(\hat{h})-1$, the $n-$th last component of $\hat{h}$ is the $(n+1)-$th last component of $\hat{h}'$.*

Not all pairs of elements of $\mathcal{B}(\zeta)$ can be immediate successors of each other. Given any history segment $\hat{h}$, a new realized action profile arrives in the subsequent period, which necessarily becomes the last element of the immediate next history segment players utilize. However, since this may actually alter the variability of realized history, in order to satisfy the cognitive bound, some initial elements of the previous history segment may need to be deleted in the construction of its immediate successor. For example consider Γ and let $\zeta = 2$. Then,

$$\mathcal{B}(\zeta) = \left\{ \begin{array}{c} \{(C,C)\},\{(D,D)\},\{(C,D)\},\{(D,C)\},\{(C,C),(C,C)\}, \\ \{(D,D),(D,D)\},\{(C,D),(C,D)\},\{(D,C),(D,C)\} \end{array} \right\}.$$

Notice that with $\zeta = 2$ and given Lemma 3.1, vectors consisting of two components can be present only if they are identical. Let $\hat{h} = \{(D,D),(D,D)\}$ and let the current action profile be (C,C). Then the immediate successor of $\hat{h}$ is (C,C). In fact, history segments $\{(D,D),(D,D)\},\{(C,C)\},\{(C,D)\}$ and $\{(D,C)\}$ are the only possible immediate successors of $\hat{h}$.

Let $\hat{h}$ be the current state of the stochastic process. In our model, knowledge of $\hat{h}$ is all that is needed to determine the probability of the next period history segment. This observation implies that our process is a Markov chain. We therefore drop the use of time as a subscript for the period history segments. The process moves from this current state $\hat{h}$ to an immediate successor state $\hat{h}'$ according to the following transition rule. Denote by $\Pr(a(t) \in A \mid \hat{h})$ the conditional probability of realizing action profile $a = (a_1, a_2)$ at period t given information set $\hat{h}$. We know that $\Pr(a(t) \in A \mid \hat{h}) = \sigma_1^{a_1}(t).\sigma_2^{a_2}(t)$, where $\sigma_i^{a_i}(t)$ is given by Eq. (6). $\Pr(a \in A \mid \hat{h})$ then induces a probability distribution for $\hat{h}'$ given the complexity bound of the players. We denote this induced probability distribution by $f\,{}_{\lambda}^{(1)}(\hat{h}' \mid \hat{h})$ and call it the transitional probability distribution (the use of λ as a subscript for f is to highlight the dependence of f on λ).

Thus the Markov chain is denoted by $M \equiv \left\langle \mathcal{B}(\zeta), f_{\lambda}^{(1)}(\hat{h}' \mid \hat{h}) \right\rangle$ *(*$\langle \mathcal{B}(\zeta), f_{\lambda}\rangle$ *in short) where* $B(\zeta)$ *is the state space and* f_{λ} *is the one-step transitional probability.*

Denote by $\mathcal{B}(\zeta)_{f_{\lambda}}^{(1)}(\hat{h})$ the support of $f_{\lambda}^{(1)}(\hat{h}' \mid \hat{h})$.

Lemma 3.2 *If* $\hat{h}' \in \mathcal{B}(\zeta)_{f_{\lambda}}^{(1)}(\hat{h})$,*then* $\ell(\hat{h}') \leq \ell(\hat{h}) + 1$.

Proof. To see this observe that if $\ell(\hat{h}') > \ell(\hat{h}) + 1$, implying, without loss of generality, that $\ell(\hat{h}') = \ell(\hat{h}) + 2$, it must be that the first component (call it a^*) of $\hat{h}'$ is the last component in $h \in H^t$, for some $t \geq \left[\frac{\zeta}{2}\right]$, after which h is truncated to form $\hat{h}$. Since $a^* \vec{\notin} \hat{h}$, it must be that for the extended vector $(a^*, \hat{h})$, $ccomp(a^*, \hat{h}) > \zeta$. But by construction, $(a^*, \hat{h}) \subset \hat{h}'$ if $\ell(\hat{h}') = \ell(\hat{h}) + 2$ implying that $ccomp(\hat{h}') > ccomp(a^*, \hat{h})$ and thus $\hat{h}' \notin \mathcal{B}(\zeta)$ and therefore cannot belong to a subset of it. ■

Between two consecutive periods, players cannot gain more than one unit length of memory. However, they may lose their current memory by more than one period if realization of play increases the heterogeneity of the current path of play significantly. As examples, let $\zeta = 4$ and consider the evolution of play given by the following end tail of a period t history $h^t = \{......., a', a, a, a\}$. Then, $\hat{h}^t = \{a, a, a\}$. Suppose now that at period t, the realized action profile is a. Then, $\hat{h}^{t+1} = \{a, a, a, a\}$ and therefore $\ell(\hat{h}^{t+1}) = 4 > \ell(\hat{h}^t) = 3$. Now consider the end tail $h^t = \{.......a'', a', a, , a'\}$. Here, $\hat{h}^t = \{a', a, a'\}$. Suppose now that at period t, the realized action profile is a''. In that case, $\hat{h}^{t+1} = \{a', a''\}$ and therefore $\ell(\hat{h}^{t+1}) = 2 < \ell(\hat{h}^t) = 3$.

In general, let ${}^{(n)}(\hat{h}' \mid \hat{h})$ denote the relation that $\hat{h}'$ is an n-period ahead successor of $\hat{h}$ and consider the set $\mathcal{B}(\zeta)_{f_{\lambda}}^{(n)}(\hat{h})$. Before ending this section, we prove the following three lemmas on reducibility and periodicity of $M \equiv \langle \mathcal{B}(\zeta), f_{\lambda}\rangle$.

Lemma 3.3 *If* $\zeta < \infty$, *then* $\forall \hat{h}, \hat{h}' \in \mathcal{B}(\zeta)$, $\exists n(\hat{h}, \hat{h}') < \infty$ *such that* $\hat{h}' \in \mathcal{B}(\zeta)_{f_{\lambda}}^{(n(\hat{h}, \hat{h}'))}(\hat{h})$, *i.e., every state is reachable from any state in some finite time.*

Proof. Since A is a finite set, $\zeta < \infty$, and $\forall a \in A, \forall \hat{h} \in \mathcal{B}(\zeta)$ we have $\Pr(a \in A \mid \hat{h}) > 0$, we can begin to construct any $\hat{h}'$ by adding from the right the first element of $\hat{h}'$ to $\hat{h}$, then adding to the far right the second element of $\hat{h}'$ to the previous extension of $\hat{h}$,....., then adding to the far right the last element of $\hat{h}'$ to the previous extension of $\hat{h}$. This only needs finitely many such addition steps for any $\hat{h}, \hat{h}' \in \mathcal{B}(\zeta)$. ■

To move to the next lemma, we require the notion of periodicity of Markov chains. Given the Markov chain $M \equiv \langle \mathcal{B}(\zeta), f_\lambda \rangle$, the state $\hat{h}$ has period d if

$$d = GCD \left\{ \tau \in \mathbb{Z}_+ \mid f_\lambda^{(\tau)}(\hat{h} \mid \hat{h}) > 0 \right\}$$

where GCD stands for greatest common divisor. If $d = 1$ for all $\hat{h} \in \mathcal{B}(\zeta)$, the Markov chain $M \equiv \langle \mathcal{B}(\zeta), f_\lambda \rangle$ is called aperiodic.

Lemma 3.4 *(i) If $\zeta = 1$, then $M = \langle \mathcal{B}(\zeta), f_\lambda \rangle$ is aperiodic.*
(ii) If $\zeta \geq 2$, $M = \langle \mathcal{B}(\zeta), f_\lambda \rangle$ has period greater than one.

Proof. (i) When $\zeta = 1$, $\hat{h}$ is a singleton set, implying that $\mathcal{B}(1) = A$. Since $\Pr(a' \mid a) > 0 \ \forall a, a' \in A$, the result follows.

(ii) Take any $\zeta \geq 2$ and any $\hat{h} \in \mathcal{B}(\zeta)$. The set $\mathcal{B}(\zeta)_{f_\lambda}^{(1)}(\hat{h})$ is such that $\forall \hat{h}' \in \mathcal{B}(\zeta)_{f_\lambda}^{(1)}(\hat{h}), \exists m(\hat{h}')$with $m(\hat{h}')$-th last element of $\hat{h}$ being the first element of $\hat{h}', m(\hat{h}') + 1$-th last element of $\hat{h}$ being the second element of $\hat{h}', ...,$ last element of $\hat{h}$ being the second last element of $\hat{h}'$, and the last element of $\hat{h}'$ being a new far right addition to $\hat{h}$. Now, $\hat{h}' = \hat{h}$ only if $\hat{h}$ is a finite constant sequence of elements of A. Since such sequences form only a proper subset of $\mathcal{B}(\zeta), M \equiv \langle \mathcal{B}(\zeta), f_\lambda \rangle$ loses aperiodicity. ■

Since $M \equiv \langle \mathcal{B}(\zeta), f_\lambda \rangle$ is irreducible by Lemma 3.3, any two states $\hat{h}$ and $\hat{h}'$ will have the same period and therefore we can define the period of the Markov chain itself. For example, with $\zeta = 2$, $\hat{h}$ can be either $\{a, a\}$ or $\{a'\}$. If $\hat{h} = \{a, a\}$ and $\hat{h}' = \{a\}$, $f_\lambda^{(1)}(\hat{h}' \mid \hat{h}) = 0$ since realization of action profile a given a history segment $\{a, a\}$ implies that $\hat{h}' = \{a, a\}$.

4. Convergence results

Having established the above properties of the Markov chain $M \equiv \langle \mathcal{B}(\zeta), f_\lambda \rangle$, in this section we prove two convergence results, one when $\zeta = 1$, the case when the Markov chain is aperiodic, and the other when $\zeta \geq 2$, and the Markov chain loses aperiodicity. These results will help us characterize long run outcomes of Γ^∞. As the result for $\zeta = 1$ will turn out to be the fundamental one, we will state and provide a self contained proof of it by closely following Theorem 8.9 in Billingsley (1986). The theorem in Billingsley (1986) is as follows: *If the state space is finite and the Markov chain is irreducible and aperiodic, then there is a stationary distribution $\{\pi_i\}$, and*

$$\left| p_{ij}^{(n)} - \pi_j \right| \leq A\rho^n,$$

where $A \geq 0$ and $0 \leq \rho < 1$.

We will then see that our result with $\zeta \geq 2$ is a direct consequence of the result with $\zeta = 1$ once we further show that our original Markov chain defined above satisfies an additional requirement of *primitivity*, a term to be defined later in this section. We will deal with these issues separately in the following two subsections.

4.1. $\zeta = 1$

In the following theorem we prove that if $\zeta = 1$, there is a long run stationary probability distribution for the evolution of the states of M. Let $|\mathcal{B}(\zeta)|$ denote the cardinality of $\mathcal{B}(\zeta)$. Note that $\mathcal{B}(1) = A$. Define

$$\gamma_\lambda = \min_{\hat{h}, \hat{h}' \in \mathcal{B}(1)} f_\lambda^{(1)}(\hat{h}' \mid \hat{h}). \tag{8}$$

γ_λ is the minimum probability, over all possible pairs of history segments, of moving from a given current state to its immediate successor state. It therefore gives a lower bound to the probability of evolution of the path of play in the sense that given a realized path of play up to time t, γ_λ is the lowest probability with which the most unlikely future path of play begins to evolve.

THEOREM 4.1 *If* $\zeta = 1, M = \langle \mathcal{B}(1), f_\lambda \rangle$ *has a stationary distribution* $f_\lambda^*(\cdot)$ *such that* $\forall \hat{h}, \hat{h}' \in \mathcal{B}(1)$,

$$\left| f_\lambda^{(n)}(\hat{h} \mid \hat{h}') - f_\lambda^*(\hat{h}) \right| \leq (1 - \gamma_\lambda \, |\mathcal{B}(1)|)^n$$

with $(1 - \gamma_\lambda \, |\mathcal{B}(\zeta)|) \in [0, 1)$.

Proof. Let $m^{(n)}(\hat{h}) = \underset{\hat{h}'}{\min} f_\lambda^{(n)}(\hat{h} \mid \hat{h}')$, and $M^{(n)}(\hat{h}) = \underset{\hat{h}'}{\max} f_\lambda^{(n)}(\hat{h} \mid \hat{h}')$. It follows that $m^{(n+1)}(\hat{h}) = \underset{\hat{h}'}{\min} \sum_{\hat{h}''} f_\lambda^{(1)}(\hat{h}'' \mid \hat{h}').f_\lambda^{(n)}(\hat{h} \mid \hat{h}'') \geq \underset{\hat{h}'}{\min} \sum_{\hat{h}''} f_\lambda^{(1)}(\hat{h}'' \mid \hat{h}')m^{(n)}(\hat{h}) = m^{(n)}(\hat{h})$ and $M^{(n+1)}(\hat{h}) = \underset{\hat{h}'}{\max} \sum_{\hat{h}''} f_\lambda^{(1)}(\hat{h}'' \mid \hat{h}').f_\lambda^{(n)}(\hat{h} \mid \hat{h}'') \leq \max_{\hat{h}'} \sum_{\hat{h}''} f_\lambda^{(1)}(\hat{h}'' \mid \hat{h}').M^{(n)}(\hat{h}) = M^{(n)}(\hat{h})$. Since $m^{(n)}(\hat{h}) \leq M^{(n)}(\hat{h}) \; \forall \hat{h} \in \mathcal{B}(1)$, we have

$$0 \leq m^{(1)}(\hat{h}) \leq m^{(2)}(\hat{h}) \leq ... \leq M^{(2)}(\hat{h}) \leq M^{(1)}(\hat{h}) \leq 1. \tag{9}$$

From the aperiodicity of $M = \langle \mathcal{B}(1), f_\lambda \rangle$, we know that $f_\lambda^{(1)}(\hat{h} \mid \hat{h}') > 0 \forall \hat{h}, \hat{h}' \in \mathcal{B}(1)$ so that $0 < \gamma_\lambda \leq \frac{1}{|\mathcal{B}(1)|}$. Fix any two states $\hat{h}^*$ and $\hat{h}^{**}$. For any arbitrary function $g \circ f_\lambda^{(1)}$ let $\sum_{\geq} g(f_\lambda^{(1)}(\hat{h}))$ be the summation over $\hat{h} \in \mathcal{B}(1)$ such that $f_\lambda^{(1)}(\hat{h} \mid \hat{h}^*) \geq f_\lambda^{(1)}(\hat{h} \mid \hat{h}^{**})$ and $\sum_{<} g(f_\lambda^{(1)}(\hat{h}))$ be the summation over $\hat{h} \in \mathcal{B}(1)$ such that $f_\lambda^{(1)}(\hat{h} \mid \hat{h}^*) < f_\lambda^{(1)}(\hat{h} \mid \hat{h}^{**})$. Then,

$$\begin{aligned} &\textstyle\sum_{\geq} \left[f_\lambda^{(1)}(\hat{h} \mid \hat{h}^*) - f_\lambda^{(1)}(\hat{h} \mid \hat{h}^{**}) \right] \\ &+ \textstyle\sum_{<} \left[f_\lambda^{(1)}(\hat{h} \mid \hat{h}^*) - f_\lambda^{(1)}(\hat{h} \mid \hat{h}^{**}) \right] = 0. \end{aligned} \tag{10}$$

Since

$$\sum\nolimits_{\geq} f_{\lambda}^{(1)}(\hat{h} \mid \hat{h}^{**}) + \sum\nolimits_{<} f_{\lambda}^{(1)}(\hat{h} \mid \hat{h}^{*}) \geq \gamma_{\lambda} \left|\mathcal{B}(1)\right|,$$

we get

$$\begin{aligned}\sum\nolimits_{\geq} \left[f_{\lambda}^{(1)}(\hat{h} \mid \hat{h}^{*}) - f_{\lambda}^{(1)}(\hat{h} \mid \hat{h}^{**})\right] &= 1 - \sum\nolimits_{<} f_{\lambda}^{(1)}(\hat{h} \mid \hat{h}^{*}) \\ -\sum\nolimits_{\geq} f_{\lambda}^{(1)}(\hat{h} \quad \mid \quad \hat{h}^{**}) & \\ &\leq 1 - \gamma_{\lambda} \left|\mathcal{B}(1)\right|. \end{aligned} \tag{11}$$

From Eqs. (10) and (11), we have

$$\begin{aligned} f_{\lambda}^{(n+1)}(\hat{h}' \quad \mid \quad & \hat{h}^{*}) - f_{\lambda}^{(n+1)}(\hat{h}' \mid \hat{h}^{**}) \\ &= \sum_{\hat{h}} \left(f_{\lambda}^{(1)}(\hat{h} \mid \hat{h}^{*}) - f_{\lambda}^{(1)}(\hat{h} \mid \hat{h}^{**})\right) f_{\lambda}^{(n)}(\hat{h}' \mid \hat{h}) \\ &\leq \sum\nolimits_{\geq} \left(f_{\lambda}^{(1)}(\hat{h} \mid \hat{h}^{*}) - f_{\lambda}^{(1)}(\hat{h} \mid \hat{h}^{**})\right) M^{(n)}(\hat{h}') \\ &\quad + \sum\nolimits_{<} \left(f_{\lambda}^{(1)}(\hat{h} \mid \hat{h}^{*}) - f_{\lambda}^{(1)}(\hat{h} \mid \hat{h}^{**})\right) m^{(n)}(\hat{h}') \\ &= \sum\nolimits_{\geq} \left(f_{\lambda}^{(1)}(\hat{h} \mid \hat{h}^{*}) - f_{\lambda}^{(1)}(\hat{h} \mid \hat{h}^{**})\right) \left(M^{(n)}(\hat{h}') - m^{(n)}(\hat{h}')\right) \\ &\leq (1 - \gamma_{\lambda} \left|\mathcal{B}(1)\right|) \left(M^{(n)}(\hat{h}') - m^{(n)}(\hat{h}')\right). \end{aligned}$$

Since $\hat{h}^{*}$ and $\hat{h}^{**}$ are arbitrary, we can write

$$M^{(n+1)}(\hat{h}) - m^{(n+1)}(\hat{h}) \leq (1 - \gamma_{\lambda} \left|\mathcal{B}(1)\right|) \left(M^{(n)}(\hat{h}) - m^{(n)}(\hat{h})\right)$$

implying that

$$M^{(n)}(\hat{h}) - m^{(n)}(\hat{h}) \leq (1 - \gamma_{\lambda} \left|\mathcal{B}(1)\right|)^{n} \tag{12}$$

From Eqs. (9) and (12), we know that $M^{(n)}(\hat{h})$ and $m^{(n)}(\hat{h})$ have a common limit. Call this limit $f_{\lambda}^{*}(\hat{h})$. From Eq. (12), we get

$$\left|f_{\lambda}^{(n)}(\hat{h} \mid \hat{h}') - f_{\lambda}^{*}(\hat{h})\right| \leq (1 - \gamma_{\lambda} \left|\mathcal{B}(1)\right|)^{n}. \tag{13}$$

Since $0 < \gamma_{\lambda} \leq \frac{1}{|\mathcal{B}(\zeta)|}$, $1 - \gamma_{\lambda} \left|\mathcal{B}(1)\right| \in [0, 1)$. Therefore, $\lim_{n\to\infty} (1 - \gamma_{\lambda} \left|\mathcal{B}(1)\right|)^{n} = 0$ implying that $f_{\lambda}^{(n)}(\hat{h} \mid \hat{h}') \to f_{\lambda}^{*}(\hat{h})$ as $n \to \infty$. ■

We are now in a position to deal with the case of $\zeta \geq 2$.

4.2. $\zeta \geq 2$

Lemma 3.4 shows that the Markov chain $M \equiv \langle \mathcal{B}(\zeta), f_\lambda \rangle$ is not aperiodic for $\zeta \geq 2$ and this prevents the existence of limiting probabilities. However, an interesting property of the Markov chain, proved in Proposition 4.1 of this section, is that for each value of ζ there exists a positive integer $n(\zeta)$ such that $f_\lambda^{n(\zeta)}(\hat{h} \mid \hat{h}') > 0$ for all $\hat{h}, \hat{h}' \in \mathcal{B}(\zeta)$. This makes the stochastic matrix primitive (a nonnegative matrix A is *primitive* if there exists some finite $k > 0$ such that $A^k >> 0$) and enables us to convert the original Markov chain $M \equiv \langle \mathcal{B}(\zeta), f_\lambda \rangle$ defined in Section 3 into its $n(\zeta)$- step Markov chain $M^{n(\zeta)} \equiv \left\langle \mathcal{B}(\zeta), f_{\lambda, n(\zeta)} \right\rangle$ such that a unit transition period of $M^{n(\zeta)} \equiv \left\langle \mathcal{B}(\zeta), f_{\lambda, n(\zeta)} \right\rangle$ is equivalent to $n(\zeta)$ transition periods of $M \equiv \langle \mathcal{B}(\zeta), f_\lambda \rangle$. Since $M^{n(\zeta)} \equiv \left\langle \mathcal{B}(\zeta), f_{\lambda, n(\zeta)} \right\rangle$ is then not only irreducible but also aperiodic (by construction), we obtain results on stationary distributions of states of $M^{n(\zeta)} \equiv \left\langle \mathcal{B}(\zeta), f_{\lambda, n(\zeta)} \right\rangle$. This subsection will deal with these issues.

PROPOSITION 4.1 *Let $M \equiv \langle \mathcal{B}(\zeta), f_\lambda \rangle$ be the Markov chain as described in Section 3. Define an integer function $n : \mathbb{Z}_+ \setminus \{1\} \to \mathbb{Z}_+, n : \zeta \mapsto n(\zeta)$ such that*

$$n(\zeta) = \arg\min_m \left\{ m \in \mathbb{Z}_+ \mid f_\lambda^m(\hat{h} \mid \hat{h}') > 0, \forall \hat{h}, \hat{h}' \in \mathcal{B}(\zeta) \right\}.$$

Then, (i) $n(\zeta)$ exists and (ii) $n(\zeta) = \zeta$.

Proof. (i) Denote by $F(\zeta) = \left[f_\lambda(\hat{h} \mid \hat{h}') \right]_{\hat{h}, \hat{h}' \in \mathcal{B}(\zeta)}$ the stochastic matrix of the Markov chain $M \equiv \langle \mathcal{B}(\zeta), f_\lambda \rangle$ and F^m its $m-$th power. We need to show the existence of $n(\zeta) \in \mathbb{Z}_+$ such that $F^{n(\zeta)} >> 0$, i.e., $f_\lambda^{n(\zeta)}(\hat{h} \mid \hat{h}') > 0, \forall \hat{h}, \hat{h}' \in \mathcal{B}(\zeta)$.

Let $(\rho_i)_{i=1}^r$ be the set of eigenvalues of $F(\zeta)$ and let $C(\rho) = \prod_{i=1}^r (\rho - \rho_i)$ be its characteristic function which can be rewritten as $C(\rho) = \rho^r + \rho^i \sum_{i=1}^{r-1} a_i \rho^{i-(r-1)}$ with $a_i \neq 0$ $\forall i = 1, 2, ..., r$. Define

$$b = \# \left\{ \rho_i \mid \nu(F(\zeta)) = |\rho_i| \, , \nu(F(\zeta)) = \max_i \{\rho_i\} \right\}.$$

Thus $\nu(F(\zeta))$ is the spectrum of $F(\zeta)$. Since $F(\zeta)$ is such that $\forall \zeta$, $trace(F(\zeta)) > 0$, by Lemma 4.10 in Graham (1987), $b = 1$. Since $b = 1$ is the definition for $F(\zeta)$ to be primitive (if $b = 1$, $A^k >> 0$ for some finite $k > 0$, see Graham 1987), our result follows.

(ii) Part (i) establishes that for any value of ζ, there exists an integer $n(\zeta)$ such that $f_\lambda^{n(\zeta)}(\hat{h} \mid \hat{h}') > 0, \forall \hat{h}, \hat{h}' \in \mathcal{B}(\zeta)$. Let $\zeta \in \mathbb{Z}_+ \setminus \{1\}$ and choose two constant $\zeta-$ length sequences $(a, a, ..., a)$ and $(a', a', ..., a')$ with $a, a' \in A$, $a \neq a'$. Since *ccomp* of these sequences equal ζ, they are elements of $\mathcal{B}(\zeta)$.

Without loss of generality, let $\hat{h} = (a, a, ..., a)$ and $\hat{h}' = (a', a', ..., a')$. It is straightforward to see that

$$\arg\min_{m} \left\{ m \in \mathbb{Z}_+ \mid f_\lambda^m(\hat{h} \mid \hat{h}') > 0 \right\} = \zeta.$$

Now choose any other pair of states $\hat{h}^*, \hat{h}^{**} \in \mathcal{B}(\zeta)$ such that

$$m^* = \arg\min_{m} \left\{ m \in \mathbb{Z}_+ \mid f_\lambda^m(\hat{h}^* \mid \hat{h}^{**}) > 0 \right\} > \zeta.$$

Since $\Pr(a \mid \hat{h}) > 0 \ \forall a \in A, \forall \hat{h} \in \mathcal{B}(\zeta)$, and $\forall m < m^*$ we have $f_\lambda^m(\hat{h}^* \mid \hat{h}^{**}) = 0$, it must be that $\ell(\hat{h}^*) = m^* > \zeta$ implying that $\hat{h}^* \notin \mathcal{B}(\zeta)$. Therefore, $n(\zeta) = \zeta$. ■

We will now deal with $M^\zeta \equiv \langle \mathcal{B}(\zeta), f_{\lambda,\zeta} \rangle$ which is irreducible and aperiodic and prove the following theorem. Let

$$\gamma_{\lambda,\zeta} = \min_{\hat{h}, \hat{h}' \in \mathcal{B}(\zeta)} f_{\lambda,\zeta}(\hat{h} \mid \hat{h}'). \tag{14}$$

THEOREM 4.2 $\forall \zeta \geq 2, M^\zeta \equiv \langle \mathcal{B}(\zeta), f_{\lambda,\zeta} \rangle$ *has a stationary distribution* $g_\lambda^*(\cdot)$ *such that* $\forall \hat{h}, \hat{h}' \in \mathcal{B}(\zeta)$,

$$\left| f_{\lambda,\zeta}^{(n)}(\hat{h} \mid \hat{h}') - g_\lambda^*(\hat{h}) \right| \leq (1 - \gamma_{\lambda,\zeta} \left| \mathcal{B}(\zeta) \right|)^n$$

Proof. Given Proposition 4.1, this is a direct application of Theorem 4.1. ■

Notice that $g_\lambda^*(\hat{h})$ is the stationary distribution of the $\zeta-$ step Markov chain $M^{(\zeta)} \equiv \langle \mathcal{B}(\zeta), f_{\lambda,\zeta} \rangle$. Since our main concern is with the long run distribution of the Markov chain $M \equiv \langle \mathcal{B}(\zeta), f_\lambda \rangle$, we need to interpret $g_\lambda^*(\hat{h})$ in terms of $M \equiv \langle \mathcal{B}(\zeta), f_\lambda \rangle$. In this sense, $g_\lambda^*(\hat{h})$ is the long run probability with which $\hat{h}$ occurs in every $\zeta-$ th period. Therefore the long run probability distribution of $M \equiv \langle \mathcal{B}(\zeta), f_\lambda \rangle$ is such that $f_\lambda^*(\hat{h})$ will be around $g_\lambda^*(\hat{h})$ and at every $\zeta-$ th period will exactly equal $g_\lambda^*(\hat{h})$. As far as our interests go, we can keep this informal interpretation in mind and proceed to characterize long run outcomes of Γ^∞.

5. Long run outcomes in Prisoners' Dilemma

We are now in a position to characterize convergence of play in infinitely repeated Prisoners' Dilemma. In the previous sections we abstracted from action profiles and dealt with the evolution of history segments. These history segments in turn are the information sets of the learning mechanism used in deciding upon the choice probabilities. Since choice of actions, in turn, determines the actual path of play and therefore the sequence of future history segments used as information sets, convergence in probability of information sets should also imply convergence of the choice probabilities. The case $\zeta = 1$ is rather simple as history segments are then singleton sets of realized actions and convergence of history segments immediately implies convergence of actions. However, when $\zeta \geq 2$, history segments may consist of more than one realized action pair and therefore we need to study convergence of actual play more carefully. As before, we will study these two cases separately. The following proposition shows that with $\zeta = 1$, highly rational probability choices converge to the stationary Nash equilibrium if defecting is highly rewarding but otherwise converge to the cooperative outcome.

Proposition 5.1 *Let* $\zeta = 1$.

(i) If $\alpha + \theta < 2\beta$ *and* $\beta + \varepsilon > 2\theta$, $\lim_{\lambda\to\infty} f^*_\lambda(C, C) = 1$.

(ii) If $\alpha + \theta > 2\beta$ *and* $\beta + \varepsilon < 2\theta$, $\lim_{\lambda\to\infty} f^*_\lambda(D, D) = 1$.

Proof. Consider the stochastic matrix $F_\lambda(\zeta = 1) \equiv \left[f_\lambda(\widehat{h} \mid \widehat{h}')\right]_{\widehat{h},\widehat{h}' \in \mathcal{B}(1)}$ of the Markov chain $M \equiv \langle \mathcal{B}(1), f_\lambda\rangle$. Since $f^*_\lambda = \left(f^*_\lambda(\widehat{h})\right)_{\widehat{h}\in\mathcal{B}(1)}$ is the stationary probability distribution, we have

$$F_\lambda(\zeta = 1) f^*_\lambda = f^*_\lambda \qquad \forall\lambda \in [0, \infty]\,. \tag{15}$$

(i) If $\alpha + \theta < 2\beta$ and $\beta + \varepsilon > 2\theta$, and states are arranged in the order $(C, C), (C, D), (D, C), (D, D)$,

$$\lim_{\lambda\to\infty} F_\lambda(\zeta = 1) = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}. \tag{16}$$

From Eqs. (15) and (16), we have

$$\begin{aligned} \lim_{\lambda\to\infty} f^*_\lambda(C, C) + \lim_{\lambda\to\infty} f^*_\lambda(D, D) &= \lim_{\lambda\to\infty} f^*_\lambda(C, C) \text{ and} \\ \lim_{\lambda\to\infty} f^*_\lambda(C, D) + \lim_{\lambda\to\infty} f^*_\lambda(D, C) &= \lim_{\lambda\to\infty} f^*_\lambda(D, D). \end{aligned}$$

Since $0 \leq f^*_\lambda(\widehat{h}) \leq 1 \ \forall\lambda \in [0, \infty]\,, \forall\widehat{h} \in \mathcal{B}(1)$, it follows that

$$\lim_{\lambda\to\infty} f^*_\lambda(C, C) = 1 \text{ and } \lim_{\lambda\to\infty} f^*_\lambda(\widehat{h}) = 0 \ \forall\widehat{h} \in \mathcal{B}(1), \widehat{h} \neq (C, C).$$

(ii) If $\alpha + \theta > 2\beta$ and $\beta + \varepsilon < 2\theta$ and the order of the states are preserved as in part (i),

$$\lim_{\lambda\to\infty} F_\lambda(\zeta = 1) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}. \tag{17}$$

From Eqs. (15) and (17), we have

$$\lim_{\lambda\to\infty} f^*_\lambda(C,C) = \lim_{\lambda\to\infty} f^*_\lambda(C,D) = \lim_{\lambda\to\infty} f^*_\lambda(D,C) = 0, \lim_{\lambda\to\infty} f^*_\lambda(D,D) = 1.$$

■

Proposition 5.1 says that in Prisoners' Dilemmas with $\lambda \to \infty$, if the average cooperative payoff is less than the defective payoff and the average defective payoff is higher than the cooperative payoff, sufficiently rational players individually converge to the Nash equilibrium and play D most of the time in the long run. On the other hand, if the average defecting payoff is less than the cooperative payoff and the average cooperative payoff is greater than the defective payoff, we expect rational play to converge to cooperation.

The intuition here is that with one period memory (and we will see that this intuition extends to any finite cognitive bound), whenever play moves away from at least one player defecting so that the most recent play was (C, C), players perceive that a payoff from a deviation to action D is only the average of the two possible payoff realizations from such a deviation and this average equals $\frac{\alpha+\theta}{2}$. Now, if β, the current individual payoff at (C, C), is greater than this average, then as rationality increases, players are less and less likely to deviate from (C, C). Similarly, whenever play moves away from at least one player cooperating so that the most recent play was (D, D), players perceive that a payoff from a deviation to action C is only the average of the two possible payoff realizations from such a deviation and this average equals $\frac{\beta+\varepsilon}{2}$. Now if θ, the current individual payoff at (D, D), is less than this average, then as rationality increases, players are more and more likely to deviate from (D, D) and play C and hence they hardly ever stay at (D, D). Moreover, (C, D) or (D, C) are never sustainable for long because ε is very small and hence the player who is currently at such states playing C deviates and plays D, a case where from thereon both players have a very high probability of playing C. These two forces work together (and get stronger and stronger as $\lambda \to \infty$) to make sure that the process converges to (C, C) for ever in the long run. This is more or less how part (i) of Proposition 5.1 works. Part (ii) is its mirror image where play converges instead to (D, D).

On the other hand, with infinite memory, players always remember some past experience with action D that produced the "highest possible" payoff of α, and therefore as $\lambda \to \infty$, they eventually deviate. Hence an important message of this paper is that for the generalized quantal response learning model studied here, the two limits, $\zeta \to \infty$ (as in case of McKelvey and Palfrey) and $\lambda \to \infty$ for a given but finite value of ζ (which is addressed in this paper) do not commute. In other words, to converge to the Nash equilibrium (D, D) in a Prisoners' Dilemma game, it is not enough to be fully rational. What is also required is full memory, falling short of which even full rationality does not guarantee Nash equilibrium play in the long run in a Logit model of choice.

The proposition shows that we can define two classes of Prisoners' Dilemmas, one where defecting is relatively more rewarding and we converge to the Nash equilibrium, and the other where cooperating is relatively more rewarding and we converge to the Pareto dominant outcome. There are other classes of Prisoners' Dilemmas which do not fall in either of the two classes mentioned above. However, Theorem 4.1 tells us that nevertheless, players will converge to some stationary probability distribution in any finite game. We now extend this proposition to any finite ζ.

PROPOSITION 5.2 *Let* $\infty > \zeta \geq 2$.

(i) If $\alpha + \theta < 2\beta$ *and* $\beta + \varepsilon > 2\theta$,

$$\lim_{\lambda\to\infty} g^*_\lambda(\Big\{(C,C), \overset{\zeta-\ times}{\cdots\cdots\cdots}, (C,C)\Big\}) = 1.$$

(ii) If $\alpha + \theta > 2\beta$ *and* $\beta + \varepsilon < 2\theta$,

$$\lim_{\lambda\to\infty} g^*_\lambda(\Big\{(D,D), \overset{\zeta-\ times}{\cdots\cdots\cdots}, (D,D)\Big\}) = 1.$$

Proof.
The strategy of the proof is as follows. We show that the system $\lim\limits_{\lambda\to\infty} F^\zeta_\lambda g^*_\lambda = g^*_\lambda$ satisfies what is claimed in part (i) and part (ii) of the above statement. In order to do this, we first compute F^ζ_λ and show that $\lim\limits_{\lambda\to\infty} F^\zeta_\lambda$ satisfies conditions under which our results are outcomes of unique solutions to the system $\lim\limits_{\lambda\to\infty} F^\zeta_\lambda g^*_\lambda = g^*_\lambda$. The proof involves a number of steps, each considering subsets of the state space of history segments that partition the set $\mathcal{B}(\zeta)$. For a given $\zeta \geq 2$, denote by $L_\zeta(a, a', ..., a'^{.....'}) \subset H^t(\tau_b, \tau_a)$ the collection of history segments with cognitive complexity between ζ and $\zeta - 1$ and containing all elements $a, a', ..., a'^{.....'}$ at least once and no other element.

(i) STEP 1. Consider $L_\zeta\{(C,D),(D,C),(D,D)\} \subset H^t(\tau_b,\tau_a)$. By the definition of the Prisoners' Dilemma, $\forall i = 1, 2$, we have

$$\lim_{\lambda\to\infty} \Big\{\sigma^C_i(t) \mid \hat{h}^t \in L_\zeta\{(C,D),(D,C),(D,D)\}\Big\} = 0.$$

STEP 2. Consider $L_\zeta\{(C,D),(D,D)\}$ and $L_\zeta\{(D,C),(D,D)\} \subset H^t(\tau_b,\tau_a)$. Given the definition of the Prisoners' Dilemma,

$$\begin{aligned}\lim_{\lambda\to\infty} \Big\{\sigma^D_i(t) \mid \hat{h}^t \in L_\zeta\{(C,D),(D,D)\}\Big\} &= \\ \lim_{\lambda\to\infty} \Big\{\sigma^D_i(t) \mid \hat{h}^t \in L_\zeta\{(D,C),(D,D)\}\Big\} &= 1.\end{aligned}$$

STEP 3. Consider $L_\zeta\{(C,D),(C,C)\} \subset H^t(\tau_b,\tau_a)$ and $L_\zeta\{(D,C),(C,C)\} \subset H^t(\tau_b,\tau_a)$. $\forall \hat{h}^t \in L_\zeta\{(C,D),(C,C)\}$, $\sigma^C_2(t) \to 0$ as $\lambda \to \infty$ (as $\alpha > \beta$) and $\sigma^D_1(t) \to 1$ as $\lambda \to \infty$ if $\hat{h}^t$ is such that (C,D) occurs sufficiently many times while $\sigma_1^D(t) \to 0$ as $\lambda \to \infty$ if $\hat{h}^t$ is such that (C, C) occurs sufficiently many times. Thus, $\forall \hat{h}^t \in L_\zeta \left\{(C, D), (C, C)\right\}$,

$$\sum_{\hat{h}^{t\prime} \in L_\zeta\{(C,D),(D,D)\}} \lim_{\lambda\to\infty} f_{\lambda,\zeta}(\hat{h}^{t\prime} \mid \hat{h}^t) = 1.$$

Take any $\hat{h}^{t\prime} \in L_\zeta \left\{(C, D), (D, D)\right\}$. Then, by STEP 2, $\forall i = 1, 2$, $\sigma_i^D(t) \to 1$ as $\lambda \to \infty$. The argument is symmetric $\forall \hat{h}^t \in L_\zeta \left\{(D, C), (C, C)\right\}$.

STEP 4. Let $L_\zeta \left\{(D, D), (C, C)\right\} \subset H^t(\tau_b, \tau_a)$ be collection of history segments containing both (C, C) and (D, D). $L_\zeta \left\{(D, D), (C, C)\right\} \subset \mathcal{B}(\zeta)$ if $\zeta \geq 3$. Since $\beta > \theta$, either

$$\lim_{\lambda\to\infty} f_{\lambda,\zeta}((C, C), \overset{\zeta-\text{times}}{\ldots\ldots\ldots\ldots}, (C, C) \quad | \quad \hat{h}^t \in L_\zeta \left\{(D, D), (C, C)\right\}) = 1 \text{ or}$$

$$\lim_{\lambda\to\infty} f_{\lambda,\zeta}((C, C), \overset{(\zeta-1)-\text{times}}{\ldots\ldots\ldots\ldots}, (C, C) \quad | \quad \hat{h}^t \in L_\zeta \left\{(D, D), (C, C)\right\}) = 1.$$

STEP 5. Let $L_\zeta \left\{(C, D), (D, C)\right\} \subset H^t(\tau_b, \tau_a)$ be collection of history segments containing both (C, D) and (D, C). $L_\zeta \left\{(C, D), (D, C)\right\} \subset \mathcal{B}(\zeta)$ if $\zeta \geq 3$. By the definition of the Prisoners' Dilemma,

$$\lim_{\lambda\to\infty} \left\{\sigma_i^D(t) \mid \hat{h}^t \in L_\zeta \left\{(C, D), (D, C)\right\}\right\} = 1.$$

STEP 6. Let $L_\zeta(a) \subset H^t(\tau_b, \tau_a)$ be collections of constant history segments, $a \in A$. If $a = (C, C)$ and $(\alpha + \theta)/2 < \beta$

$$\lim_{\lambda\to\infty} f_{\lambda,\zeta}((C, C), \overset{\zeta-\text{ times}}{\ldots\ldots\ldots\ldots}, (C, C) \mid \hat{h}^t \in L_\zeta \left\{(C, C)\right\}) = 1.$$

If $a = (D, D)$, and $(\beta + \varepsilon)/2 > \theta$,

$$\lim_{\lambda\to\infty} f_{\lambda,\zeta}((C, C), \overset{\zeta-\text{ times}}{\ldots\ldots\ldots\ldots}, (C, C) \mid \hat{h}^t \in L_\zeta \left\{(D, D)\right\}) = 1.$$

If $a = (C, D)$ or (D, C),

$$\lim_{\lambda\to\infty} f_{\lambda,\zeta}((D, D), \overset{\zeta-\text{ times}}{\ldots\ldots\ldots\ldots}, (D, D) \quad | \quad \hat{h}^t \in L_\zeta \left\{(C, D)\right\}) =$$

$$\lim_{\lambda\to\infty} f_{\lambda,\zeta}((D, D), \overset{\zeta-\text{ times}}{\ldots\ldots\ldots\ldots}, (D, D) \quad | \quad \hat{h}^t \in L_\zeta \left\{(D, C)\right\}) = 1.$$

STEP 7. Define $L_\zeta^{(C,C)} \left\{a, a', ..., a'^{....\prime}\right\} \subset H^t(\tau_b, \tau_a)$ such that $\exists a' = (C, C)$ and $\exists a'' \neq (C, C)$. For any $\hat{h}^t \in L_\zeta^{(C,C)} \left\{a, a', ..., a'^{....\prime}\right\}$, $\hat{h}^t$ cannot be repeated. To see this observe that given any $\hat{h}^t = \left\{a^1, ..., a^n\right\}$, for some n, if $a^k = (C, C)$, $k \leq n$, and if $\exists j \geq 1$ such that $a(t + j) = (C, C)$, then $\forall m \geq 1$, $a(t + j + m) = (C, C)$ if $(\beta + \varepsilon) > 2\theta$ and $(\alpha + \theta) < 2\beta$. Thus starting from any $\hat{h}^t \in L_\zeta^{(C,C)} \left\{a, a', ..., a'^{....\prime}\right\}$, we end up in any one of the above cases.

Now consider the stochastic matrix $F_\lambda^\zeta = \left[f_{\lambda,\zeta}(\hat{h} \mid \hat{h}')\right]_{\hat{h},\hat{h}' \in \mathcal{B}(\zeta)}$ of the $\zeta-$step Markov chain $M^\zeta \equiv \langle \mathcal{B}(\zeta), f_{\lambda,\zeta}\rangle$. Since $g_\lambda^* = \left[g_\lambda^*(\hat{h})\right]_{\hat{h}\in\mathcal{B}(\zeta)}$ is the stationary distribution $\forall\lambda \in [0,1], \forall 2 \leq \zeta < \infty$,

$$F_\lambda^\zeta g_\lambda^* = g_\lambda^*. \tag{18}$$

Given STEP 1 through STEP 7 above, consider the permutation of F_λ^ζ such that the states are ordered as

$$\begin{aligned}
\hat{h}_1 &= \left\{(C,C), \overset{\zeta-\text{ times}}{\cdots\cdots\cdots}, (C,C)\right\}, \\
\hat{h}_2 &= \left\{(C,C), \overset{(\zeta-1)-\text{ times}}{\cdots\cdots\cdots}, (C,C)\right\}, \\
\hat{h}_3 &= \left\{(D,D), \overset{\zeta-\text{ times}}{\cdots\cdots\cdots}, (D,D)\right\}, \\
\hat{h}_4 &= \left\{(D,D), \overset{(\zeta-1)-\text{ times}}{\cdots\cdots\cdots}, (D,D)\right\}, \\
\hat{h}_5 &= \ldots\ldots\ldots .
\end{aligned}$$

Then, as $\lambda \to \infty$, F_λ^ζ satisfies the following conditions:

$$\begin{aligned}
(i)\ \lim_{\lambda\to\infty} f_{\lambda,\zeta}(\hat{h}_1 &\mid \hat{h}_i, i = 1,2,3,4) = 1, \\
(ii)\ \lim_{\lambda\to\infty} f_{\lambda,\zeta}(\hat{h}_1 &\mid \hat{h}_i, i = 5, ..., |\mathcal{B}(\zeta)|) = 0, \text{ and} \\
(iii)\ \lim_{\lambda\to\infty} f_{\lambda,\zeta}(\hat{h}_i, i &= 5, ..., |\mathcal{B}(\zeta)| \mid \hat{h}_j, j = 1,2,....,|\mathcal{B}(\zeta)|) = 0.
\end{aligned}$$

Furthermore, since $f_{\lambda,\zeta}(\hat{h} \mid \hat{h}') \in [0,1] \forall \hat{h},\hat{h}' \in \mathcal{B}(\zeta)$ and $\sum\limits_{\hat{h}\in\mathcal{B}(\zeta)} f_{\lambda,\zeta}(\hat{h} \mid \hat{h}') = 1 \forall \hat{h}' \in \mathcal{B}(\zeta)$ and $\forall\lambda \in [0,\infty)$, it is easy to see that the unique solution to the system as $\lambda \to \infty$ is

$$\lim_{\lambda\to\infty} g_\lambda^*(\left\{(C,C), \overset{\zeta-\text{ times}}{\cdots\cdots\cdots}, (C,C)\right\}) = 1.$$

(ii) Following the proof in part (i) it can be shown that for the permutation of F_λ^ζ such that the states are arranged in the following order,

$$\begin{aligned}
\hat{h}_1 &= \left\{(D,D), \overset{\zeta-\text{ times}}{\cdots\cdots\cdots}, (D,D)\right\}, \\
\hat{h}_2 &= \left\{(D,D), \overset{(\zeta-1)-\text{ times}}{\cdots\cdots\cdots}, (D,D)\right\}, \\
\hat{h}_3 &= \left\{(C,C), \overset{\zeta-\text{ times}}{\cdots\cdots\cdots}, (C,C)\right\}, \\
\hat{h}_4 &= \left\{(C,C), \overset{(\zeta-1)-\text{ times}}{\cdots\cdots\cdots}, (C,C)\right\}, \\
\hat{h}_5 &= \ldots\ldots\ldots .
\end{aligned}$$

if $\alpha + \theta > 2\beta$ and $\beta + \varepsilon < 2\theta$, as $\lambda \to \infty$, F_λ^ζ satisfies

$$\begin{aligned}(i)\ \lim_{\lambda\to\infty} f_{\lambda,\zeta}(\hat{h}_1 \quad &| \quad \hat{h}_i, i = 1,2,3,4) = 1,\\ (ii)\ \lim_{\lambda\to\infty} f_{\lambda,\zeta}(\hat{h}_1 \quad &| \quad \hat{h}_i, i = 5, ..., |\mathcal{B}(\zeta)| = 0, \text{ and}\\ (iii)\ \lim_{\lambda\to\infty} f_{\lambda,\zeta}(\hat{h}_i, i \quad &= \quad 5, ..., |\mathcal{B}(\zeta)| \mid \hat{h}_j, j = 1,2, ..., |\mathcal{B}(\zeta)|) = 0\end{aligned}$$

and thus the result follows. ■

The above proposition shows that if $\alpha + \theta < 2\beta, \beta + \varepsilon > 2\theta$ and $\zeta < \infty$, the Markov chain M^ζ has stationary distribution g_λ^* such that the probability of realizing the history segment $\left\{(C,C), \overset{\zeta-\text{ times}}{\cdots\cdots\cdots}, (C,C)\right\}$ tends to 1 as $\lambda \to \infty$. Does this necessarily guarantee that for the Markov chain $M \equiv \langle \mathcal{B}(\zeta), f_\lambda \rangle$, the history segment $\left\{(C,C), \overset{\zeta-\text{ times}}{\cdots\cdots\cdots}, (C,C)\right\}$ also occurs at every period in the long run with probability one ? The following Corollary deals with this.

Corollary 5.1 *Let M and M^ζ be as defined above with long run stationary distributions f_λ^* and g_λ^* respectively. Consider any constant history segment $\left\{(a), \overset{\zeta-\ times}{\cdots\cdots\cdots}, (a)\right\}$ of length ζ with $a \in \{C,D\}^2$ such that*

$$\lim_{\lambda\to\infty} g_\lambda^*(\left\{(a), \overset{\zeta-\ times}{\cdots\cdots\cdots}, (a)\right\}) = 1.$$

Then,

$$\lim_{\lambda\to\infty} f_\lambda^*(\left\{(a), \overset{\zeta-\ times}{\cdots\cdots\cdots}, (a)\right\}) = 1.$$

Proof. Consider any $\zeta-$ length period starting at $t-\zeta$ and ending at $t-1$ for some t sufficiently large. Let π be the probability that $a(\tau) = a\ \forall\tau \in \{t-\zeta, ..., t-1\}$ and π_T be the probability that $a(\tau) = a\ \forall\tau \in \{t-\zeta, ..., T\}, T \leq t-1$ for some $a \in \{C,D\}^2$. It follows that $\pi_T \geq \pi\ \forall T \in \{t-\zeta, ..., t-1\}$. Since

$$\left\{\lim_{\lambda\to\infty} g_\lambda^*(\left\{(a), \overset{\zeta-\text{ times}}{\cdots\cdots\cdots}, (a)\right\}) = 1\right\} \Rightarrow \pi = 1 \text{ as } t \to \infty,$$

we have

$$\pi_T \geq 1\ \forall T \in \{t-\zeta, ..., t-1\} \text{ as } t \to \infty.$$

Since by definition, $\pi_T \leq 1$, we get

$$\pi_T = 1\ \forall T \in \{t-\zeta, ..., t-1\} \text{ as } t \to \infty$$

which implies

$$\lim_{\lambda\to\infty} f^*_\lambda(\Big\{(a), \overset{\zeta-\text{ times}}{\ldots\ldots\ldots\ldots}, (a)\Big\}) = 1.$$

■

The conditions $\alpha + \theta < 2\beta$ and $\beta + \varepsilon > 2\theta$ guarantee that in the long run (C, C) is played with probability one while $\alpha + \theta > 2\beta$ and $\beta + \varepsilon < 2\theta$ guarantee that (D, D) is played with probability one.

6. Application to other 2×2 games with $\zeta = 1$

In this section we apply our general results to a large class of 2×2 games which include Pure Coordination, Common Interest and the game of Chicken, under the restriction that the length of the memory cannot exceed unity. We then argue why the results in this section hold in principle even when the complexity of the history segments used exceeds one.

6.1. Coordination games

In this class of games, we would like to distinguish between two types. The first one is often called *pure coordination* games where $\beta > \theta > Max\{\varepsilon, \alpha\}$ and the second one is often called *common interest* games where $\beta > \alpha > \theta > \varepsilon$. Before we prove results on these games, let us highlight an important distiction between them which is also reflected in the next two propositions. Consider the following games $G1$ and $G2$ which are both coordination games:

$G1$:

		player 2	
		C	D
player 1	C	$4, 4$	$1, 1$
	D	$1, 1$	$2, 2$

and

$G2$:

		player 2	
		C	D
player 1	C	$4, 4$	$1, 3$
	D	$3, 1$	$2, 2$

Both games have two pure strategy Nash equilibria, namely (C, C) and (D, D) and in both games (C, C) is the only efficient outcome, hence the term "coordination". But the crucial difference is that while in $G1$, both players prefer to be in one of these Nash equilibria than not being in any, in $G2$, for the Nash equilibrium (D, D), player 1 prefers (D, C) while player 2 prefers (C, D) while they both prefer to be in (C, C) rather than in (D, C) or (C, D). However, notice that while from (D, D) , player 1 cannot deviate and implement (D, C) which he prefers, neither can player 2 deviate and implement (C, D) which he prefers. The game $G1$ is called pure coordination while game $G2$ is called common interest, which are terms used to capture this difference. And as seen from the statements of Propositions 6.1 and 6.2, for pure coordination games, the entry α does not appear while it does in an importantly specific manner in common interest games.

6.1.1. Pure coordination

Consider a game of Pure Coordination with two pure strategy Nash equilibria (C, C) and (D, D) with (C, C) Pareto dominating all other outcomes.

PROPOSITION 6.1 *Let $\zeta = 1$ and Γ be a game of Pure Coordination. If $2\theta < \beta + \varepsilon, (C, C)$ is the unique long run equilibrium for $\lambda \to \infty$.*

Proof. Following the procedure described in the proof of Proposition 5.1, if the states are arranged in the order $(C, C), (D, D), (C, D), (D, C)$, and if $2\theta < \beta$, we have

$$\lim_{\lambda \to \infty} F_\lambda(\zeta = 1) = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

and therefore the unique solution to the system $F_\lambda(\zeta = 1) f^*_\lambda = f^*_\lambda$ is as desired when $\lambda \to \infty$. ■

Our model therefore predicts that players will eventually converge with probability one to the Pareto dominant Nash equilibrium in a game of Pure Coordination if they have a unit cognitive bound and the Pareto dominant Nash equilibrium is sufficiently payoff rewarding.

6.1.2. Common interest

Consider a game of Common Interest with two pure strategy Nash equilibria (C, C) and (D, D) with (C, C) Pareto dominating all other outcomes.

PROPOSITION 6.2 *Let $\zeta = 1$ and let Γ be a game of Common Interest. If $\alpha > \left(\frac{\beta+\varepsilon}{2}\right) > \theta$, (C, C) is the unique long run equilibrium when $\lambda \to \infty$.*

Proof. If the states are arranged in the order $(C, C), (C, D), (D, C), (D, D)$, and if $\alpha > \left(\frac{\beta+\varepsilon}{2}\right) > \theta$, we have

$$\lim_{\lambda \to \infty} F_\lambda(\zeta = 1) = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

and therefore the unique solution to the system $F_\lambda(\zeta = 1) f^*_\lambda = f^*_\lambda$ is as desired when $\lambda \to \infty$. ■

Our model therefore predicts that players will eventually converge with probability one to the Pareto dominant Nash equilibrium in a game of Common

Interest if they face unit cognitive bounds and $\alpha > \left(\frac{\beta+\varepsilon}{2}\right) > \theta$. The intuition behind this result is that with $\left(\frac{\beta+\varepsilon}{2}\right) > \theta$, payoff experience of θ makes both players play C with very high probability and eventually both players start playing C. By the definition of a Common Interest game, once we observe play of (C, C), players keep playing C with high probabilities which converge to one as their rationality parameter approaches infinity. Moreover, if they ever observe play of (D, C) or (C, D), if $\alpha > \left(\frac{\beta+\varepsilon}{2}\right)$, the probability of both players playing D is very high. This further implies that with a very high probability (which tends to one as rationality is increased unboundedly) we observe play of (D, D) and then play converges to (C, C). If on the other hand $\alpha < \left(\frac{\beta+\varepsilon}{2}\right)$, one may show that there exists a long run equilibrium where players alternatively mis-coordinate.

6.2. Chicken and the fairness equilibrium

Consider Γ with $\varepsilon > \theta > \alpha > \beta > 0$. Then Γ is a game of Chicken with two pure strategy Nash equilibria (C, D) and (D, C). Following Rabin (1993) and Camerer (1997), suppose player 1 has a positive 'sympathy' coefficient when player 2 'kindly helps' player 1 and conversely a negative 'sympathy' coefficient when player 2 behaves 'meanly' by choosing an action that hurts player 1. Rabin assumes that such feelings add to the utility from money payoffs, but become relatively less important as money payoffs rise. These assumptions and a few others (see Rabin) lead to the concept of a *fairness equilibrium.* First let us study the following example to understand this concept of fairness. Let Γ be represented by the following payoff matrix.

		player 2	
		fight	accommodate
player 1	fight	0.01,0.01	6,2
	accommodate	2,6	4,4

In our model specification, $C \equiv$ fight , $D \equiv$ accommodate, $\varepsilon = 6, \theta = 4, \alpha = 2, \beta = 0.01$. The two pure strategy Nash equilibria are (C, D) and (D, C). Although in spirit it is a simultaneous move game, consider the following pre-play thinking on part of the two players. Suppose we are in (C, D). If player 1 deviates and 'politely' plays D, she sacrifices 2 to benefit player 2 an extra amount of 2. This 'nice' choice triggers reciprocal niceness in the behavior of player 2 and rather than exploiting over player 1 choosing D, he prefers to sacrifice to repay player 1's kindness and plays D. If player 2 also reasons in the same way, (D, D) is the unique outcome and is called the *fairness equilibrium.* Experimental evidence supports the fact that subjects tend to play fairness equilibrium strategies in a game of Chicken (see Camerer, 1997).

Proposition 6.3 *If $\theta > \left(\frac{\beta+\varepsilon}{2}\right) > \alpha$ and Γ is game of Chicken, players converge to the fairness equilibrium with probability one as $\lambda \to \infty$.*

Proof. It is easy to see that if the states are arranged in the order (D, D), (C, C), (C, D), (D, C), and if $\theta > \left(\frac{\beta+\varepsilon}{2}\right) > \alpha$, we have

$$\lim_{\lambda\to\infty} F_\lambda(\zeta = 1) = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

and therefore the unique solution to the system $F_\lambda(\zeta = 1)f^*_\lambda = f^*_\lambda$ is as desired when $\lambda \to \infty$. ■

From the above proposition we see that players converge to the *fairness equilibrium* (D, D) with probability one in the long run if $\theta > \left(\frac{\beta+\varepsilon}{2}\right) > \alpha$. Thus, in the numerical example above our theory predicts that players will converge with probability one to the *fairness equilibrium* (D, D). This result also supports in some sense the assumption of Rabin (1993) that utility out of sympathy is outweighed by increased money payoff incentives from daring against a chicken which is necessary for the existence of a fairness equilibrium. However, the intuition in our set up is that with unit cognitive bounds, since $\beta + \varepsilon < 2\theta$, if we observe the action profile (D, D), players keep playing D with very high probabilities which tend to one as their rationality tends to infinity. Furthermore, if the play ever leads to outcomes like (C, D) or (D, C), the player playing D deviates and plays C with very high probability while the player playing C keeps playing C with very high probability too. This implies that almost certainly, we may observe play of (C, C) in the subsequent period. By the definition of the Chicken game, with rationality close to infinity, players eventually start playing D with very high probability. At the limit, when the rationality goes to infinity, play gets stuck in (D, D). Mention must be made here that in the McKelvey and Palfrey (1995) formulation, players always converge to either (C, D) or (D, C).

Before ending this section, we would like to mention the basic intuition by which we claim (without providing formal proofs) that the results for this section would actually hold for any finite cognitive bound. As long as players are forced to use only most recent and finite history segments as information sets for deciding upon the choice probabilities, the above conditions would drive them to information sets wherefrom playing the long run outcome is enforced with very high probability. The relative weights of payoffs are in some sense stability conditions for the dynamic system generated by the behavior rules of the players.

7. A discussion on beliefs without experience

The results obtained in this paper can be easily supported in spirit in a model with general belief-probabilities over outcomes generated by strategies for which players have no experience. In the Prisoners' Dilemma game for example, suppose players believe that whenever they have no experience with a particular strategy, the opponent chooses C with probability $\omega \in (0,1)$. Consider a class of games where $\omega\alpha + (1-\omega)\,\theta < \beta$ and $\omega\beta + (1-\omega)\,\varepsilon > \theta$. Then, our theory would predict that players in the long run will play (C, C) with probability one. Similarly, consider another class of games where $\omega\alpha + (1-\omega)\,\theta > \beta$ and $\omega\beta + (1-\omega)\,\varepsilon < \theta$. Then, our theory would predict that players in the long run will play (D, D) with probability one. It follows from this discussion that beliefs "on the forgotten path" or "off the actual path" of play becomes central. However, our aim is not to provide a theory as to how such beliefs are formed. One may argue that there is some inconsistency or asymmetry in these beliefs as hypothesized here in the following sense. It seems that a player's belief regarding whether her opponent cooperates or defects depends upon the action the player herself chooses to play. For example, if the memory constrained information set contains only the outcome (C, C), the player believes that playing C results in the outcome (C, C) in the following period with probability one (which is possible only if the player believes that her opponent plays C with probability one), while playing D results in the outcome (D, C) with probability ω and the outcome (D, D) with probability $1-\omega$ (which is now possible only if the player believes her opponent plays C with probability ω). Firstly, in logit models like the one used here, it would be incorrect to interpret ω in terms of a player's belief regarding what her opponent will choose. Rather, it should be thought of as the relative frequency of success (which could arise either when both players cooperate or for the player who defects when her opponent still cooperates). Clearly then, while confronting a strategy either without experience or with forgotten outcomes, ω becomes arbitrary. Secondly, it is not unknown in social sciences and in particular in various studies in psychology that perceptions towards risks may be different depending upon whether an agent is herself in a cooperative or a defective mood. Also, a drastic change in mood of a player within a period before an action is actually implemented may disturb the ongoing beliefs over outcomes thereby leading to apparent inconsistencies.

8. Conclusion

We conclude with a summary of our results and relevant comparisons with the existing literature. We studied infinite repetition of the Prisoners' Dilemma. Players were assumed to use a version of Logistic Quantal Response Learning behavior. However, they face finite cognitive bounds in understanding histories of past play. We define two classes of Prisoners' Dilemma games: one in which the average defecting payoff is higher than the cooperative payoff and the average cooperative payoff is lower than the defective payoff; and the other where the average defecting payoff is lower than the cooperative payoff and the average cooperative payoff is higher than the defective payoff. As the degree of rationality goes to infinity, we show that as long as players face finite cognitive bounds, in the former class of games, play converges to the static Nash equilibrium which is Pareto dominated while in the latter, play converges to the Pareto dominant outcome where both players play the cooperative action. As the degree of assumed rationality is reduced, the convergence point moves away in both classes of games until it hits the centroid of the three-dimensional unit simplex of probability distributions over action pairs. Note that there are other classes of Prisoners' Dilemmas which do not fall in any of the classes mentioned above. However, we show that repetition of any finite game leads to some stationary long run distribution over the space of action profiles. Our theory calls for experiments where subjects play the Prisoners' Dilemma with an uncertain terminal period. One easy way of capturing the notion of cognitive bounds in such experimental set up would be to impose time restrictions within which subjects need to decide upon their current actions. We then apply the results obtained under the general framework to other classes of 2 × 2 games like Pure Coordination, Common Interest and Chicken. We show that as long as players face unit cognitive bounds, under relevant ordinal payoff restrictions, play converges to the Pareto dominant Nash equilibrium in both Pure Coordination and Common Interest games. In case of the Chicken game, we show that players converge to the fairness equilibrium if 'daring' is not sufficiently rewarding.

9. APPENDIX

9.1. Results with alternative formulation

The McKelvey and Palfrey result of convergence of Logit equilibrium to the Nash equilibrium when $\lambda \to \infty$ is proved with the alternative functional form of the choice probabilities as in Eq. (7). Here we show that our results hold with their formulation as well. In place of Eq. (6) or Eq. (7), consider a more general choice probability function

$$\sigma_i^k(t) = \frac{F(\pi_i^k(t), \lambda)}{\sum\limits_{k=1}^{N} F(\pi_i^k(t), \lambda)}. \tag{19}$$

If $F(\cdot)$ is continuous and bounded $\forall \lambda \in (0, \infty)$ and $F(\pi_i^k(t), \lambda) > 0 \; \forall (\pi_i^k(t), \lambda) \in \mathbb{R} \times (0, \infty)$, then these logistic choice probabilities are well defined. Furthermore, if $\Big(F(\pi_i^k(t), \lambda)/F(\pi_i^{k'}(t), \lambda)\Big) \to 0$ as $\lambda \to \infty$ whenever $F(\pi_i^k(t), \lambda) < F(\pi_i^{k'}(t), \lambda)$, convergence to Nash equilibrium as in McKelvey and Palfrey (1995) is ensured. Thus our formulation as in Eq. (6) guarantees convergence to (D, D)

with probability one when $\zeta = \infty$. As far as our results with $\zeta < \infty$ are concerned, they also hold good with Eq. (7) since our results depend on these conditions on $F(.)$ as well.

Acknowledgements

The author holds a visiting position at the Faculty of Economics, Microeconomics Unit, University of Warsaw, Poland.

The author remains grateful to Sandro Brusco for help and guidance, to Debabrata Datta, Laurence Kranich, and Prabuddha Sanyal for motivating discussions on this topic and thanks an anonymous referee of this journal for many detailed comments that have improved the paper both in clarity and content. The paper also benefitted from seminar presentations at the Indian Statistical Institute at the Calcutta and Delhi centers, and at the Institute of Computer Science of the Polish Academy of Sciences in Warsaw. The usual disclaimer applies.

References

Axelrod, R. (1984) *The Evolution of Cooperation.* New York, Basic Books.

Bendor J., Mookherjee, D. and Ray, D. (1995) Aspirations, Adaptive Learning and Cooperation in Repeated Games. *Discussion Paper*, Planning Unit, Indian Statistical Institute, New Delhi.

Billingsley, P. (1986) *Probability and Measure.* John Wiley & Sons.

Binmore, K. and Samuelson, L. (1992) Evolutionary Stability in Repeated Games Played by Finite Automata. *Journal of Economic Theory* **57**, 278-305.

Camerer, C.F. (1997) Progress in Behavioral Game Theory. *Journal of Economic Perspectives* **11** (4), 167-188.

Hsiao-Chi, Ch., Thisse, J.-F. and Friedman, J.W. (1997) Boundedly Rational Nash Equilibrium: A probabilistic choice approach. *Games and Econ. Behav.* **18**, 32-54.

Doob, J.L. (1953) *Stochastic Processes.* John Wiley, New York.

Gilboa, I. (1988) The Complexity of Computing Best-Response Automata in Repeated Games. *Journal of Economic Theory* **45** (2), 342-352.

Gilboa, I. and Schmeidler, D. (1995) Case-Based Decision Theory. *Quarterly Journal of Economics* **110**, 605-639.

Graham, A. (1987) *Nonnegative Matrices and Applicable Topics in Linear Algebra.* John Wiley & Sons.

Harsanyi, J.C. and Selten, R. (1988) *A General Theory of Equilibrium Selection in Games.* MIT Press.

Hopcroft, J. and Ullman, J. (1979) *Introduction to Automata Theory, Languages, and Computation.* Addison-Wesley.

Kalai, E. and Stanford, W. (1988) Finite Rationality and Interpersonal Complexity in Repeated Games. *Econometrica* **56** (2), 397-410.

Karandikar, R., Mookherjee, D., Ray, D. and Vega-Redondo, F. (1998) Evolving Aspirations and Cooperation. *Journal of Economic Theory* **80**, 292-331.

Kim, Y. (1999) Satisficing and optimality in common interest games. *Economic Theory* **13**, 365-375.

Luce, R.D. (1959) *Individual Choice Behavior: A Theoretical Analysis.* John Wiley, New York.

Maynard Smith, J. (1982) *Evolution and the Theory of Games. C*ambridge University Press.

McKelvey, R.D. and Palfrey, T.R. (1995) Quantal Response Equilibria for Normal Form Games. *Games and Econ. Behav.* **10**, 6-38.

McFadden, D. (1976) Quantal Choice Analysis: A Survey. *Ann. of Econ. and Soc. Measures* **5**, 363-390.

Mookherjee, D. and Sopher, B. (1997) Learning and Decision Costs in Experimental Constant Sum Games. *Games and Econ. Behav.* **19**, 97-132.

Neyman, A. (1985) Bounded Complexity justifies Cooperation in the Finitely Repeated Prisoners' Dilemma. *Economic Letters* **19**, 227-229.

Rabin, M. (1993) Incorporating Fairness into Game Theory and Economics. *American Economic Review* **83**, 1281-1302.

Simon, H. (1972) *Models of Man.* John Wiley, New York.

Sonsino, D. (1997) Learning to Learn, Pattern Recognition, and Nash Equilibrium. *Games and Economic Behavior* **18**, 286-331.