Economic Theory Bulletin (2021) 9:209–224
https://doi.org/10.1007/s40505-021-00204-0

RESEARCH ARTICLE

Evolution of conventions in games between behavioural rules

Abhimanyu Khan[1]

Received: 26 February 2020 / Accepted: 19 April 2021 / Published online: 6 May 2021

Abstract

I examine when, how, and which conventions arise in N-player games. Each player draws a random sample of strategies used in the recent past, and then chooses a strategy in response to this sample. A player's response is determined by a behavioural rule, which maps from the set of recently used strategy profiles to a subset of his own strategy set, and each element in the latter set is chosen with positive probability. A random sample of strategies is *monomorphic* if it contains only one distinct strategy for each of the other players. The behavioural rule of a player is *responsive* if, on drawing a monomorphic sample, there is a positive probability of playing a best-response to the other players' strategy profile that is induced by their respective strategies in that sample; in addition, if the said induced strategy profile supports a strict Nash equilibrium, then a strategy played by him in the recent past is chosen with the complementary probability. A game is *weakly acyclic* if there exists a 'best-response path' from each outcome that is not a strict Nash equilibrium to a strict Nash equilibrium. I show that: (i) a convention forms whenever the players' behavioural rules are responsive, and the game is weakly acyclic, (ii) in bi-matrix games, individuals described by the behavioural rule of *extreme optimism*—whereby, conditional on the random sample, they play a best-response to the most optimistic belief about the other player's strategy choice—perform better than individuals described by any other responsive behavioural rule in the sense that the convention that is most preferred by the former is always in the stochastically stable set, and (iii) in bi-matrix pure coordination games, the said convention is the uniquely stochastically stable state if the other player's behavioural rule is 'mildly different' from extreme optimism.

Keywords Evolution · Conventions · Behavioural rules · Responsiveness · Weakly acyclic games · Extreme optimism

JEL Classification C73

I would like to thank an anonymous referee and the editor for their helpful comments on this paper.

✉ Abhimanyu Khan
abhimanyu.khan.research@gmail.com

[1] Shiv Nadar University, Gautam Buddha Nagar, Greater Noida, Uttar Pradesh, India

1 Introduction

A convention is described as a situation where individuals customarily act in a predictable fashion. The focus of this paper is on how conventions arise in a decentralised manner when individuals recurrently interact with each other. In any environment with strategic uncertainty, an individual's behaviour depends on how the game has unfolded in the past, on how he expects the other individuals to behave, and on how he conditions his action in response to these considerations. A convention is sustained by the expectation that others will act as per the prevailing custom; this results in every individual also preferring to do the same, as choosing to do otherwise would presumably lead to a less desirable outcome. In this paper, I assume that each individual's expectation about the other individuals is shaped purely by their historical behaviour, and each individual is associated with a behavioural rule that specifies his response to the history of the game. In this setting, I pose the following questions: What are the conditions on the nature of the strategic environment and on the behavioural rules under which one would find that the individuals themselves end up settling on a convention? Furthermore, in an environment where different conventions may arise, the one that is most likely to be observed depends on the interaction between the behavioural rules of the players; in this context, is a particular behavioural rule more advantageous than others in the sense that the most preferred convention of the individuals who behave according to that rule is also the convention that is most likely to emerge?

I study these issues in the adaptive play framework of Young (1993, 1998). I consider N-player games where each player's strategy set is finite, and he only needs to know his binary preference relation between each pair of outcomes of the game. In order to obtain information about the history of the game, and to form an expectation about how other individuals may play the game, each individual draws a random sample from the strategies used in the recent past. His strategy choice is obtained from his behavioural rule, which maps from the set of recently used strategies to a subset of his own strategy set, and each element in this latter subset is chosen with positive probability.

A random sample of strategies is *monomorphic* if it contains only one distinct strategy for each of the other players. The behavioural rule of a player is *responsive* if, on drawing a monomorphic sample, there is a positive probability of playing a best-response to the other players' strategy profile that is induced by their respective strategies in that sample; in addition, if the said induced strategy profile supports a strict Nash equilibrium, then a strategy played by him in the recent past is chosen with the complementary probability. I emphasise that responsiveness does not impose any restriction on an individual's response to non-monomorphic samples. Since individuals' expectation about others' actions are formed by the strategies observed in the randomly drawn sample, and because each individual knows his binary preference relation, any reasonable behavioural rule should arguably be responsive.

I interpret a convention to be a strict Nash equilibrium of a game, and show that when the behavioural rule of each player is responsive and the N-player game is weakly acyclic, then, irrespective of the initial history of the game, the individuals settle on a convention. A game is *weakly acyclic* if it has at least one strict Nash equilibrium, and there is a 'best-response path' from any outcome that is not a strict Nash equilibrium to a strict Nash equilibrium.[1] It follows that there may not be complete certainty about the particular convention that may arise in weakly acyclic games with multiple strict Nash equilibria. In order to study the influence of behavioural rules on the convention that is most likely to emerge in such situations, I examine bi-matrix (or two-player) games where players described by one particular behavioural rule interact with players described by another behavioural rule. I focus on a particular responsive behavioural rule, namely *extreme optimism*; conditional on the randomly drawn sample, an extremely optimistic individual plays a best-response to the most optimistic belief about the strategy choice of the other player. I show that the convention that is most preferred by extremely optimistic individuals is always in the set of long-run outcomes. Furthermore, in a sub-class of weakly acyclic games, namely pure coordination games, it is the unique long-run outcome whenever the behavioural rule of the other individual is mildly different from extreme optimism (in a sense made precise later). This leads me to suggest that extreme optimism outperforms almost any other behavioural rule.

This paper contributes to the literature on the formation of conventions in a decentralised environment. Previous work that has examined this issue in the context of all players being described by similar behavioural rules include Young (1993), Kandori et al. (1993), Hurkens (1995), Sáez-Marti and Weibull (1999), Matros (2003), and Khan and Peeters (2014) for the best-response rule or 'clever/sophisticated' versions thereof;[2] Josephson (2008) for better-reply rules; Karandikar et al. (1998) for satisficing and aspirational play in 2×2 games; Robson and Vega-Redondo (1996), Josephson and Matros (2004), and Bergin and Bernhardt (2009) for imitation.[3] Blume (1993, 1997) analyses a logit-response dynamic in a particular class of games (i.e. potential games) and under an asynchronous strategy revision protocol; in such cases, the outcome of the game converges to a subset of the Nash equilibrium. Alós-Ferrer and Netzer (2010) show that this result depends critically on both the revision protocol (i.e. independent revision vs asynchronous revision) and the particular class of games; this motivates them to characterize the long-run outcome for arbitrary normal form games when the individuals' strategy revision process is described more broadly by a generalization of the logit-response dynamics. Hwang and Newton (2017) also analyse a class of coordination games under a dynamic that includes the logit choice rule.

On the other hand, Kaniovski et al. (2000), Juang (2002), and Josephson (2009) study heterogeneous behavioural rules, where the heterogeneity is not across individuals but about the same individual using different rules (for example, probabilistically best-responding and imitating). In this paper, I broaden the understanding of how,

[1] Alós-Ferrer and Netzer (2017) also show convergence to the set of strict Nash equilibrium in weakly acyclic games with the logit-response dynamic.

[2] Bilancini and Boncinelli (2018) add an additional layer by examining the outcome of a coordination game when individuals choose both the action and the set of individuals they interact with whereas Bilancini and Boncinelli (2020) study how the interaction between the perturbation in the response of an individual and the persistence of interaction affects the outcome.

[3] Vega-Redondo (1997), Alós-Ferrer and Ania (2005), and Hedlund (2015) examine the relationship between imitative behaviour and perfectly competitive outcomes in the context of markets and firm behaviour.

when, and which conventions may arise when an individual's decision-making process is fairly unrestricted.[4] In addition, I analyse a question that, to the best of my knowledge, has not been previously examined in the literature: which behavioural rule is most successful in the sense of delivering the most preferred convention to the individuals who adopt that rule?[5]

2 Unperturbed adaptive play: framework and analysis

There are N distinct finite populations, which are indexed by the first N natural numbers. Population i is assigned the role of player i in the N-player game under consideration. X_i denotes the finite strategy set of the individuals in population i, $\forall i \in \{1, \ldots, N\}$. Time is discrete. In every time period, one individual from each population is randomly chosen, and these randomly chosen N individuals play the game in that period. $x_i(t) \in X_i$ is the strategy chosen in period t by the individual from population i. $X_{-i} = \Pi_{j=1,\ j \neq i}^{N} X_j$ is the strategy space of all players other than i, and $x_{-i}(t) \in X_{-i}$ is their period t strategy profile. $X = \Pi_{i=1}^{N} X_i$ denotes the set of outcomes, and the period t outcome is denoted by $x(t) = (x_1(t), \ldots, x_N(t))$. Whenever convenient, I write (x_i, x_{-i}), with $x_i \in X_i$ and $x_{-i} \in X_{-i}$, to denote the outcome obtained when, for a particular $i \in \{1, \ldots, N\}$, the population i player chooses $x_i \in X_i$, and strategy profile of the other players is $x_{-i} \in X_{-i}$. The binary rational preference relation of the individuals in population i, that is defined on the set X, is represented by $\succeq_i$. For any $x, y \in X$, $x \succeq_i y$ implies that the individuals in population i consider the outcome x to be at least as good as the outcome y. Individuals need only be aware of their binary preference relation—while individuals may know more about their preferences, I do not make any such assumption. I emphasise the recurrent nature of the game, and abstract away from considerations that may arise out of repeated interactions.

The state of the game at the beginning of period $t + 1$—before each individual has chosen his strategy—is given by the finite H period history of strategies $\omega(t + 1) = (x(t), x(t - 1), \ldots, x(t - H + 1))$. I assume that the initial history of the game (i.e. the strategies played in the first H periods) is specified exogenously. $\omega_i(t + 1) = (x_i(t), \ldots, x_i(t - H + 1))$ denotes the vector of strategies used by population i in the preceding H periods. Similarly, $\omega_{-i}(t + 1) = (x_{-i}(t), \ldots, x_{-i}(t - H + 1))$ is the vector of the strategy profiles of the populations other than population i in the previous H periods. The state space is denoted by $\Omega = (X_1 \times \cdots \times X_N)^H$.

In the beginning of period $t + 1$, the randomly chosen individual from each population $i \in \{1, \ldots, N\}$ draws a random sample of S strategies without replacement from $\omega_j(t + 1)$, $\forall j \in \{1, \ldots, N\}$ That is, he draws S strategies from the strategies used by each population in the last H time periods. The sample size $S \leq H$ is fixed

[4] On a related note, Khan (2021) examines the evolutionary stability of behavioural rules in the context of a bargaining game.

[5] While Axelrod (1984) studies the success of various strategies in the repeated Prisoner's Dilemma, the novelty of the question in this paper lies in its focus on the most successful behavioural rule in a broad class of games—rather than the most successful strategy in a particular game (i.e. the Prisoner's dilemma)—and I abstract from repeated game considerations in order to isolate the effect of the behavioural rules.

exogenously. Every feasible sample has positive probability of being drawn; however, it is possible that particular strategies in the history are more likely than others to be drawn. $s_i(t+1)$ denotes the sample drawn in period $t+1$ by the individual from population i. $supp(s_{i,j}(t+1))$ is the set of strategies of population j that appear in the sample $s_i(t+1)$. I define the sample $s_i(t+1)$ to be *monomorphic* if $supp(s_{i,j}(t+1))$ is a singleton for each of the other populations $j \in \{1, \ldots, N\} \backslash \{i\}$. Thus, $supp(s_{i,i}(t+1))$ need not be a singleton for the sample to be monomorphic. If $s_i(t+1)$ is monomorphic, then $x_{-i}(s_i(t+1)) \in X_{-i}$ denotes the strategy profile induced by the monomorphic sample $s_i(t+1)$. That is, $x_{-i}(s_i(t+1))$ is the strategy profile formed by taking the strategy in the singleton $supp(s_{i,j}(t+1))$ for each of the other populations $j \in \{1, \ldots, N\} \backslash \{i\}$.

Individuals in population i are associated with a behavioural rule $R_i : \Pi_{i=1}^{N} X_i^S \rightrightarrows P(X_i)$ that goes from $\Pi_{i=1}^{N} X_i^S$, the space of feasible samples, to $P(X_i)$, the power set of the strategy set X_i that contains all the non-empty subsets of X_i. Each element of the response set $R_i(s_i(t+1))$ has strictly positive probability of being played in period $t+1$. $SoBR_i(x_{-i})$ represents the set of best responses of the individuals in population i to the strategy profile x_{-i} of the other players, i.e. $SoBR_i(x_{-i}) = \{x_i \in X_i : (x_i, x_{-i}) \succeq_i (z_i, x_{-i}), \forall z_i \in X_i\}$. I use $BR_i(x_{-i})$ to refer to an element in $SoBR_i(x_{-i})$ whenever further identification of the element is not necessary; so, $BR_i(x_{-i})$ denotes a best response to the strategy profile x_{-i}.

A *best-response path* $x' \rightarrow x''$ exists from the outcome $x' \in X$ to another outcome $x'' \in X$ if there exists $i \in \{1, \ldots, N\}$ such that $x' = (x'_i, x'_{-i})$, $x'' = (x''_i, x'_{-i})$, and $x''_i \in SoBR_i(x'_{-i})$. That is, along a best-response path from x' to x'', the strategy chosen by exactly one player differs, and the strategy chosen by him in x'' is a best-response to the strategy profile of the other players in x'. An outcome is defined to be a *sink* if there does not exist any best-response path from it; thus, a sink is equivalent to a strict Nash equilibrium. I will say that a strategy profile $x_{-i} \in X_{-i}$ *supports a strict Nash equilibrium* if there exists $x_i \in X_i$ such that (x_i, x_{-i}) is a strict Nash equilibrium. A game is *weakly acyclic* if it contains at least one sink, and if there exists a finite sequence of best-response paths from every non-sink outcome to a sink. Hence any outcome $x^0 \in X$ is either a sink, or there exists a finite sequence of best-response paths $x^0 \rightarrow x^2 \cdots x^{k-1} \rightarrow x^k$ from x^0 to a sink x^k. Some prominent examples of weakly acyclic games include the Nash bargaining/demand game, coordination games, and the Prisoner's dilemma.

The behavioural rule of population i is *responsive* if $SoBR_i(x_{-i}(s_i(t+1))) \subset R_i(s_i(t+1))$ for any monomorphic sample $s_i(t+1)$; in addition, if $x_{-i}(s_i(t+1))$ supports a strict Nash equilibrium, then $SoBR_i(x_{-i}(s_i(t+1))) \subset R_i(s_i(t+1)) \subset SoBR_i(x_{-i}(s_i(t+1))) \cup \omega_i(t+1)$. That is, the set of best-responses to the other players' strategy profile that is induced by a monomorphic sample is always in his response set. Furthermore, if the induced strategy profile of the other players supports a strict Nash equilibrium, then the response set may contain the strategies played by the individuals in population i in any of the previous H periods. I underline that in the event that $x_{-i}(s_i(t+1))$ does not support a strict Nash equilibrium, $R_i(s_i(t+1))$ may contain any other strategy in addition to the ones in $SoBR_i(x_{-i}(s_i(t+1)))$, and that responsiveness does not impose any restriction on the response set corresponding to any non-monomorphic sample. I suggest that responsiveness is the weakest condition that must be satisfied by any behavioural rule that responds to information. Conditional on an individual's response depending on the obtained sample, the event where a monomorphic sample is drawn is the simplest decision making situation for an individual. If the individual does not play a best-response in such a situation in spite of being aware of his preference relation, then his behavioural rule is arguably completely unresponsive to the information contained in the sample. In this sense, responsiveness is a very mild restriction on the behavioural rules, and I present few examples to illustrate its generality:

(i) the best-response rule: an individual always plays a best-response to the empirical distribution of strategies of the other individuals in the random sample.
(ii) better-response rules/aspirational play: an individual has a reference strategy that may either be specified exogenously or obtained from the history of strategies played by the individuals in his population in the recent past; then, the response set corresponding to a monomorphic sample $s_i(t+1)$ comprises of the strategies that lead to outcomes that are at least as preferred as the outcome obtained by playing the reference strategy (when the other individuals' strategy profile is assumed to be $x_{-i}(s_i(t+1))$).
(iii) stochastic imitation and best/better response: with positive probability, an individual plays a best-response/better-response (as in (i) and (ii) above); with the complementary probability, he imitates either the strategy that was most commonly/least commonly used by the individuals in his population in the past H periods or the strategy that yielded the highest payoff/highest average payoff to the individuals in his population in the last H periods.
(iv) behavioural rules with inertia: with positive probability, an individual plays a strategy that is dictated by a combination of the rules in (i), (ii), or (iii) above; with the complimentary probability, he plays the same strategy that was chosen by the individual in his population in the previous period.

I also present two examples of behavioural rules that are not responsive. Firstly, a behavioural rule that induces an individual to always play the same strategy, irrespective of the sample, is not responsive. Another such example is a purely imitation based behavioural rule that causes an individual to play either: (a) the strategy that was most commonly/least commonly used in the previous H periods by the individuals in his population, or (b) the strategy that led to the highest payoff/highest average payoff to the individuals in his population in the previous H periods.

This recurrent strategic situation—where individuals respond to a sample of strategies played in the recent past—can be described by a Markov process Q on the state space Ω. A *convention* is a state $\omega \in \Omega$ where the same sink outcome, say $\bar{x} \in X$, has been realised in all the preceding H periods. This particular convention is denoted by $\omega_{\bar{x}}$. $Co(G)$ denotes the set of conventions of the game G. $Co(G)_i = \{\omega_x \in Co(G)$: $x \succeq_i y, \forall \omega_y \in Co(G)\}$ is the set of conventions of the game G that is most preferred by the individuals in population i. Without loss of generality, I assume $Co(G)_i$ to be a singleton—if not, then the individuals in population i are indifferent amongst all conventions in $Co(G)_i$. The game is said to *converge almost surely* to a convention if there is positive probability of a transition from any state to a convention in finite time, and if that convention persists in all succeeding periods.

In Proposition 1 below, I state that the game converges almost surely to a convention if the sampling is sufficiently incomplete, the game is weakly acyclic, and the behavioural rules are responsive. The reason is as follows. Consider any initial state that does not represent a convention. Then, with positive probability, individuals respond to the history of the game in such a manner that over time, the state is described in turn by the outcomes that lie on the sequence of best-response paths that originates from that initial state to a sink; thus, with positive probability, the game transitions from any state to a convention in finite time. Responsiveness of all the behavioural rules implies that once a convention is formed, it persists then on. This is because once a convention is formed, any feasible sample that may be drawn by any individual in any population is monomorphic, and the induced strategy profile of the other players supports the same strict Nash equilibrium; responsiveness of each behavioural rule then implies that the response set of any individual comprises only of the strategy played by him in that strict Nash equilibrium. I refer to the proof of the proposition in the Appendix for a more formal presentation of these arguments.

Proposition 1 *Let sampling be sufficiently incomplete so that* $\frac{S}{H} \leq \frac{1}{3}$ *if* $N = 2$*, and* $\frac{S}{H} \leq \frac{1}{\Sigma_{i=1}^{N}|X_i| - min\{|X_1|,\ldots,|X_N|\} - N + 1}$ *if* $N > 2$*. Suppose that the* N*-player game is weakly acyclic game, and the behavioural rule of the individuals in each population is responsive. Then, the game converges almost surely to a convention.*

Incomplete sampling, responsiveness of the behavioural rules, and weak-acyclicity of the games are the three salient features of Proposition 1. I will underline the importance of each feature by arguing that the game may not converge to a convention if any one of the three conditions do not hold.

Firstly, in order to highlight the importance of incomplete sampling, I borrow the example of a two-player coordination game from Young (1998). The row player [the column player] can choose strategy U or strategy D [strategy L or strategy R], and the payoff matrix is depicted in Fig. 1 below. This game is weakly acyclic as it has at least one strict Nash equilibrium—in this case, (U, L) and (D, R)—and because there exists a best-response path from each of the other outcomes, i.e. (U, R) and (D, L), to each of the two strict Nash equilibria. Suppose that individuals always sample the entire history of play, and that they mis-coordinate on the outcome (U, R) in the very first period. Also suppose the individuals always best-respond to the empirical distribution of the other player's strategies; recall that this is a responsive behavioural rule. Let $r(t)$ and $c(t)$ represent the relative frequency with which strategy U and strategy L has been played at the end of time period t by the row player and the column player, respectively. The row player [column player] plays strategy U [strategy L] in period $t + 1$ if and only if $\sqrt{2}(1 - c(t)) \leq c(t)$ [$\sqrt{2}\, r(t) \geq 1 - r(t)$]. Since the outcome in the first period is (U, R), $r(1) = 1$ and $c(1) = 0$; so, the row and the column player play D and L, respectively, and mis-coordinate in the next period as well. Continuing in this fashion, it may be seen that they mis-coordinate in each and every period. Thus, complete sampling results in the game not converging to a convention, even though the game is weakly acyclic and the behavioural rules are responsive. The reason is that with complete sampling, individuals best-respond instantaneously to the strategy chosen

Fig. 1 Example 1

$$\begin{array}{c|cc|} & L & R \\ \hline U & 1, \sqrt{2} & 0, 0 \\ D & 0, 0 & \sqrt{2}, 1 \\ \hline \end{array}$$

by the other individual in the previous period; hence, if they begin by mis-coordinating, then the mis-coordination persists thereafter. In contrast, incomplete sampling creates the opportunity for individuals to exhibit 'inertia' in their response, and this opens up the possibility of coordinating on a convention even if they mis-coordinate in any particular period.

Secondly, in order to convey the salience of responsiveness of the behavioural rules, I consider the same game but now suppose that the response set of each player always comprises of both strategies, irrespective of the sample; clearly, this behavioural rule is not responsive. I assume that sampling is arbitrarily incomplete. It follows that when the individuals behave in accordance to this behavioural rule, then the game never converges to a convention in spite of incomplete sampling and the game being weakly acyclic. Thus, responsiveness of the behavioural rules ensures both that the game eventually transitions along a finite sequence of best-response paths to a convention, and that it stays locked into that convention thereafter.

Finally, I underscore the criticality of the weakly acyclicity property by considering a two-person game where $\{U, M, D\}$ and $\{L, C, R\}$ is the strategy set of the row player and the column player, respectively. The payoff matrix is presented in Fig. 2 below. I assume that the sampling is arbitrarily incomplete, and that the individuals always best-respond to the empirical distribution of the other player's strategies in their randomly drawn sample. This game has a strict Nash equilibrium (D, R) but it is not weakly acyclic as there does not exist a best-response path from any of the outcomes (U, L), (U, C), (M, L) and (M, C) to the sink (D, R). Suppose that the initial H period history is specified by the row player [column player] playing U and M [L and C] in any exogenously fixed proportion. It follows from their behavioural rule that they will continue to play one of these two strategies thereafter; in particular, the row [column] player never plays strategy D [strategy R] as it is never a best response to any distribution of the other player's strategies that never contains strategy R [strategy D]. Hence, the absence of a best-response path to the sink outcome results in the game not converging to a convention even though the behavioural rules are responsive and sampling is incomplete.

Fig. 2 Example 2

$$\begin{array}{c|ccc|} & L & C & R \\ \hline U & 2, 1 & 1, 2 & 0, 0 \\ M & 1, 2 & 2, 1 & 0, 0 \\ D & 0, 0 & 0, 0 & 3, 3 \\ \hline \end{array}$$

3 Perturbed adaptive play: framework and analysis

The analysis in the previous section reveals that under certain conditions, the individuals are expected to eventually settle on a convention. In this context, it is instructive to develop an understanding of how behavioural rules influence the convention that is most likely to arise in games that have more than one strict Nash equilibrium. In this section, I posit that the strategic interaction may be punctuated with the possibility of exogenous shocks in the choices made by the individuals. That is, in any period $t + 1$, there is an independent probability ε of the individual in each population $i \in \{1, \ldots, N\}$ making a mistake (or experimenting) with his strategy choice. An individual is said to *experiment*, or make a *mistake*, if he plays a strategy that does not belong to the response set corresponding to *any* feasible sample that can be drawn from the H period history of strategies. These mistakes now make possible the transition from one convention to another, and, at an intuitive level, the conventions that are most likely to be observed in the long-run are the ones that are relatively hard to displace via mistakes but, at the same time, relatively easy to transit into from the complementary set of conventions via mistakes. Formally, the stochastically stable set of this perturbed Markov process is the set of states that receive positive weight in the limiting stationary distribution (see Foster and Young 1990; Kandori et al. 1993; Young 1993, 1998 for details). In what follows, I make use of an implication of particularly useful result from Ellison (2000): if it is possible to transit into a particular convention from any other convention with a single mistake, then the former convention is in the stochastically stable set; if, in addition, a transition from that particular convention needs strictly more than one mistake, then it is the uniquely stochastically stable state.

The question I pose is, does there exist a behavioural rule that outperforms any other behavioural rule in the sense that the convention that is most preferred by the individuals who follow the former rule is also the convention that is likely to be observed in the long-run. The nature of this question causes me to restrict attention to two-player games, as this allows me to pit individuals who follow one behavioural rule against individuals who follow another behavioural rule. I use i and $-i$ to denote the population that assumes the role of the row player and column player, respectively, in the game under consideration.

I focus on one particular responsive behavioural rule, namely extreme optimism. The individuals in population i are said to be *extremely optimistic* if, on drawing any sample $s_i(t + 1)$ in any period $t + 1$, they assess that the population $-i$ player will choose a strategy from the *assessment set* $AS_i(t + 1) = \{x_{-i} \in supp(s_{i,-i}(t + 1)) : (BR_i(x_{-i}), x_{-i}) \succeq_i (BR_i(x'_{-i}), x'_{-i}), \forall x'_{-i} \in supp(s_{i,-i}(t + 1))\}$; then, $R_i(s_i(t + 1)) \subset \bigcup_{x_{-i} \in AS_i(t+1)} SoBR_i(x_{-i})$. That is, an extremely optimistic individual believes that he is going to face only the most favourable of circumstances—amongst all the strategies of the other player that appear in his randomly drawn sample, he believes that his co-player will play a strategy that will lead to the most favourable outcome for him; he then plays a best-response to this belief. It is easily verified that this behavioural rule is responsive.

As an illustration of this behavioural rule, consider the game illustrated in Fig. 1, and suppose that the population of row players is extremely optimistic. Then, whenever a row player's randomly drawn sample contains both L and R, he believes that the column player will play R. This is because $(BR_i(R), R) = (D, R) \succ_i (BR_i(L), L) = (U, L)$. So, he responds by playing D. On the other hand, he believes that the column player will play strategy L [strategy R] when his sample comprises only of this strategy of the column player; in this case, he plays strategy U [strategy D] in response. Similarly, in context of the game in Fig. 2, a row player who is extremely optimistic believes that the column player will play: (i) strategy R whenever his random sample contains at least one instance of the column player playing strategy R (since $(BR_i(R), R) \succ_i (BR_i(C), C) \sim_i (BR_i(L), L)$), (ii) either strategy L or strategy C whenever the sample contains both of these strategies, but does not contain strategy R (since $(BR_i(C), C) \sim_i (BR_i(L), L)$), and (iii) strategy L [strategy C] when the sample contains only this strategy of the column player. Correspondingly, he plays (i) strategy D, (ii) strategy U or strategy M, and (iii) strategy U [strategy M].

I recall that $Co(G)_i$ is the set of conventions that is most preferred by the individuals in population i, and I assume that $Co(G)_i$ is a singleton. Let $supp_i(Co(G)_i)$ and $supp_{-i}(Co(G)_i)$ denote the strategy of the individuals in population i and population $-i$, respectively, that support the strict Nash equilibrium of the convention in $Co(G)_i$. Since $Co(G)_i$ is a singleton, $SoBR_i(supp_{-i}(Co(G)_i)) = \{supp_i(Co(G)_i)\}$, and $SoBR_{-i}(supp_i(Co(G)_i)) = \{supp_{-i}(Co(G)_i)\}$. For instance, in the game described in Fig. 1, $Co(G)_i = \{\omega_{D,R}\}$, $supp_i(Co(G)_i) = D$ and $supp_{-i}(Co(G)_i) = R$; furthermore, $SoBR_i(supp_{-i}(Co(G)_i)) = SoBR_i(R) = \{D\} = \{supp_i(Co(G)_i)\}$, and $SoBR_{-i}(supp_i(Co(G)_i)) = SoBR_{-i}(D) = \{R\} = \{supp_{-i}(Co(G)_i)\}$.

The behavioural rule of the individuals in population $-i$ is said to be *mildly different* from extreme optimism if the following condition holds. Consider any randomly drawn sample where the strategy $supp_i(Co(G)_i)$ appears *at least* $S - 1$ times while another strategy $x_i' \neq supp_i(Co(G)_i)$ appears *at most* once; then the only element in the corresponding response set is $BR_{-i}(supp_i(Co(G)_i)) = supp_{-i}(Co(G)_i)$. The reason why this behavioural rule *differs* from extreme optimism is that in response to the afore-mentioned sample, an extremely optimistic individual *may* play a strategy other than $BR_{-i}(supp_i(Co(G)_i))$ but if and only if $(x_i', BR_{-i}(x_i')) \succeq_{-i} Co(G)_i$. I underline that the proportion with which the strategy x_i' occurs in the sample can be arbitrarily small because the sample size S can be arbitrarily large, and that the definition of mildly different imposes a restriction only for the particular sample comprised of $S - 1$ instances of $supp_i(Co(G)_i)$ and a single occurrence of another strategy $x_i' \neq supp_i(Co(G)_i)$. There is no other restriction on these other rules, and hence, their response sets may mimic that of extreme optimism in all other situations—it is in this sense that these behavioural rules are mildly different from extreme optimism.

As an example, consider the game depicted in Fig. 1. The population of extremely optimistic individuals, who assume the role of the row player, play strategy D whenever they observe at least once occurrence of strategy R. In contrast, the individuals in the other population, who assume the role of the column player, play R when they observe strategy U appearing at most once and strategy D appearing at least $S-1$ times in their sample. Here, I note that responsiveness of their behavioural rule causes them to play R when they observe only strategy D in their sample; hence, the only implication of their behavioural rule being mildly different from extreme optimism is that they also play R when they observe strategy U appearing exactly once and strategy D appearing exactly $S-1$ times.

Finally, a strategic situation represents a *pure coordination game* if, for every strategy $x_i \in X_i$ [$x_{-i} \in X_{-i}$], there exists a unique strategy $x_{-i} \in X_{-i}$ [$x_i \in X_i$] such that the outcome (x_i, x_{-i}) is a strict Nash equilibrium. That is, $SoBR_i(x_{-i})$ [$SoBR_{-i}(x_i)$] is a singleton for any $x_{-i} \in X_{-i}$ [$x_i \in X_i$], and the outcome $(BR_i(x_{-i}), x_{-i})$ [$(x_i, BR_{-i}(x_i))$] is a strict Nash equilibrium. It is easily verified that pure coordination games are weakly acyclic.

The main result in this section is that in bi-matrix games, the most preferred convention of population i—that is comprised of extremely optimistic individuals—is always in the stochastically stable set if the behavioural rule of the individuals in the other population is responsive, the game is weakly acyclic, and the sampling is sufficiently incomplete. This is because a single mistake is sufficient to effect a transition from any convention in the set $Co(G)\backslash Co(G)_i$ to the convention in $Co(G)_i$. The intuition is as follows. Suppose that the state is described by a convention in $Co(G)\backslash Co(G)_i$, and, in a particular time period, the individual from population $-i$ makes a mistake by playing the strategy $supp_{-i}(Co(G)_i)$. Then, with positive probability, $supp_{-i}(Co(G)_i)$ is always contained in the samples that are drawn by the extremely optimistic individuals in the subsequent periods, and extreme optimism leads them to play $supp_i(Co(G)_i)$ in all of these periods. In response to the individuals in population i recurrently playing $supp_i(Co(G)_i)$, the individuals in population $-i$ eventually start playing the strategy $supp_{-i}(Co(G)_i)$, thereby resulting in a transition to the convention in $Co(G)_i$. Hence, this convention is always in the stochastically stable set.

In addition to the above, if the behavioural rule of the individuals in the other population is mildly different from extreme optimism, and the sample size $S > 1$, then the convention in $Co(G)_i$ is the uniquely stochastically stable state in pure coordination games. To show this, it is sufficient to argue that a transition out of the convention $\omega_{(x_i,x_{-i})} \in Co(G)_i$ is not possible by a single experimentation. So, suppose that the state in period t is $\omega_{(x_i,x_{-i})}$, and either an individual in population i experiments with $x_i' \neq x_i$, or a individual in population $-i$ experiments with $x_{-i}' \neq x_{-i}$. In either case, since $S > 1$, any sample that a population i player can draw comprises of at least one instance of strategy x_{-i}; similarly, any sample that a population $-i$ can draw comprises of at least $S-1$ instances of strategy x_i. It follows from the definition of the corresponding behavioural rules that the individuals in population i [population $-i$] continue to play $x_i[x_{-i}]$. Hence, it is not possible to transit from $\omega_{(x_i,x_{-i})}$ with one experimentation, and so, this is the only stochastically stable state. These results are summarised in Proposition 2; the formal proofs are presented in the appendix.

Proposition 2 *Consider any weakly acyclic bi-matrix game. Suppose that the individuals in one of the two populations are extremely optimistic. If sampling is sufficiently incomplete so that* $\frac{S}{H} \leq \frac{1}{3}$, *then the most preferred convention of these individuals is:* (i) *always in the stochastically stable set if the behavioural rule of the individuals in the other population is responsive, and* (ii) *the uniquely stochastically stable state in pure coordination games if the behavioural rule of the individuals in the other population is mildly different from extreme optimism, and* $S > 1$.

Corollary 1 *Consider any bi-matrix pure coordination game where a Pareto-efficient coordination outcome exists. Suppose that* $\frac{S}{H} \leq \frac{1}{3}$, $S > 1$, *and that individuals in one of the two populations are extremely optimistic but the behavioural rule of the individuals in the other population is mildly different from extreme optimism. Then, in the stochastically stable state of this game, the individuals coordinate on the Pareto-efficient coordination outcome.*

Corollary 2 *Consider the bi-matrix pure coordination game where two individuals bargain over the division of a pie of fixed size in the Nash demand game framework. Suppose that* $\frac{S}{H} \leq \frac{1}{3}$, $S > 1$, *and that individuals in one of the two populations are extremely optimistic but the behavioural rule of the individuals in the other population is mildly different from extreme optimism. Then, in the stochastically stable state of this game, the extremely optimistic individuals obtain almost the entire pie.*

4 Conclusion

I show that when individuals recurrently play an N-player game, then they end up settling on a convention if the game is weakly acyclic, and the behavioural rules of the players satisfies a very general condition, namely responsiveness. In the specific context of two-player weakly acyclic games, the most preferred convention of players described by the behavioural rule 'extreme optimism' is always in the set of long-run outcomes; furthermore, this particular convention is the unique long-run outcome in pure coordination games whenever the behavioural rule of the individuals in the other population is mildly different from extreme optimism. So, in two-player pure coordination games where the interests of the two populations are oppositely aligned (for eg. games where individuals bargain over a pie of fixed size, or Battle of the Sexes type coordination games), the other population is constrained to its least preferred convention; in such cases, extreme optimism outperforms almost any other behavioural rule. However, in two-player pure coordination games where there is no conflict of interest (for eg. Stag-hunt games or minimum effort games), the other population also shares the spoils of the most preferred convention of the population comprised of extremely optimistic individuals emerging as the long-run outcome.

Declaration

Conflict of Interest The author declares that no funding has been received for this study. The author further declares that there is no conflict of interest.

Appendix

Proof of Proposition 1 The history of play in the first H periods is assumed to be given exogenously. Consider the state $\omega(t+1)$, for any $t+1 > H$. Suppose that $\omega(t+1)$ is not a convention; otherwise, there is nothing left to prove. Now, in each time period from $t+1$ to $t+S$, with positive probability, the randomly chosen individual from each population draws the strategies played in the S time periods from $t-S+1$ to t. Then, with positive probability, the randomly chosen individual from population i plays the same strategy $\bar{x}_i \in X_i$ in all these periods, and this holds for all $i \in \{1, \ldots, N\}$. So, with positive probability, the outcome$\bar{x} = (\bar{x}_1, \ldots, \bar{x}_N)$ obtains in all the S time periods from $t+1$ to $t+S$. Then, there are two mutually exclusive and exhaustive possibilities:

(i) Suppose that the outcome $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_N)$ is a sink. Then, with positive probability, in each time period from $t+S+1$ onwards, individuals from each population draw a sample that comprises of the strategies used in the S periods immediately preceding it. Each such sample is monomorphic, and the induced strategy profile of the other players supports the strict Nash equilibrium $\bar{x}$. As a result, a population i player plays $\bar{x}_i \in X_i$, and this holds for each $i \in \{1, \ldots, N\}$. Hence, the H period history of play from $t+1$ to $t+H$ is described by the same sink outcome $\bar{x}$. Thus, the game converges to a convention almost surely.

(ii) Suppose, on the other hand, that the outcome $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_N)$ is not a sink. Then, there exists a sequence of best-response paths $x^0 \rightarrow x^1 \cdots x^{k-1} \rightarrow x^k$ where $x^0 = \bar{x}$, x^k is a sink, and x^l is the outcome $(x_1^l, \ldots, x_N^l)\ \forall l \in \{0, \ldots, k\}$. The best-response path $x^0 \rightarrow x^1$ implies that $x_i^0 \neq x_i^1$ for some $i \in \{1, \ldots, N\}$, $x_j^0 = x_j^1$ for all the other populations $j \in \{1, \ldots, N\} \backslash \{i\}$, and $x_i^1 \in SoBR(x_{-i}^0)$. Now, with positive probability:

  (a) The sample drawn in all the time periods from $t+S+1$ to $t+2\,S$ by the randomly chosen individual from population i comprises of the strategies used in the periods $t+1$ to $t+S$. Since this sample is monomorphic, the set of best-responses to the induced strategy profile of the other players belongs to his response set; hence, the randomly chosen individuals from population i play x_i^1 in all the time periods from $t+S+1$ to $t+2\,S$.

  (b) The sample drawn in all the time periods from $t+S+1$ to $t+2\,S$ by the randomly chosen individuals from all the other populations $j \in \{1, \ldots, N\} \backslash \{i\}$ comprises of the strategies used in the periods $t-S+1$ to t; so, for each $j \in \{1, \ldots, N\} \backslash \{i\}$, the individuals in population j play $x_j^0 \in X_j$ in all the time periods from $t+S+1$ to $t+2\,S$.

Then, at the end of period $t+2\,S$, the S-period history immediately preceding it comprises only of the outcome $x^1 = (x_i^1, x_{-i}^0)$. If this outcome is a sink, then (i) above applies. If not, then using the argument outlined above, there is positive probability that the outcome of the S periods immediately succeeding it corresponds to x^2 in the sequence of best-response paths. With positive probability, the process proceeds along the sequence in this manner till there is a S-period run of the sink outcome x^k, at which point (i) above applies. Hence, the game converges to a convention almost surely.

The final part of the proof relates to the incompleteness of sampling. The argument in (ii) above relies on the fact that: (a) as the state transits from one outcome of the sequence of best-response paths to another, only one population changes its strategy while the other populations do not, and (b) a S-period run of each outcome in the sequence is sufficient to move along this sequence. So, in order to move along the sequence of best-response paths, the population that changes its strategy must do so for S consecutive periods, and in these S periods, the individuals in each of the other populations do not change their strategy; in order for this to occur, the latter must have access to the sample that causes them to choose the same strategy as before. I now present on a sufficient condition that enables this.

Without loss of generality, consider the transition from the state described by the outcome x^0, and that has been outlined in (ii) above. Starting from period $t + S + 1$, exactly one population changes its strategy every S periods. At period $t + S + 1$, any population $i \in \{1, \ldots, N\}$ can play $\mid X_i \mid -1$ number of other strategies. I consider the following two mutually exclusive and exhaustive cases:

(a) When $N = 2$, then the population that is the last to change its strategy does so in period $t + 2\,S + 1$; this implies that till time period $t + 2\,S$, this population plays its strategy that supports the outcome x^0. A sufficient condition for this is that till period $t + 2$ S, the individuals in this population must have access to the strategies used as far back as period $t - S + 1$, and $\frac{S}{H} \leq \frac{1}{3}$ is sufficient to enable this.
(b) When $N > 2$, the population that is the last to change its strategy from the one that supports x^0 may actually do so after all the other populations have played all the strategies in their strategy sets. Suppose that population i is the last to change its strategy; then, the total number of strategies that may be played by the other populations before the strategy of population i changes is $Q = \Sigma_{j=1, j \neq i}^{N} \mid X_j \mid -N = \Sigma_{j=1}^{N} \mid X_j \mid - \mid X_i \mid -N$, where N is subtracted because a strategy is played by each of the N populations in the initial state x^0. Hence, the maximum of the total number of strategies that may be played by the other populations before the last population changes its strategy is $Q = \Sigma_{i=1}^{N} \mid X_i \mid -min\{\mid X_1 \mid, \ldots, \mid X_N \mid\} - N$. Now, for a sufficient upper bound on S, I consider the case where all of these Q strategies are played before the last population changes its strategy; if each of these Q strategies is played S times consecutively, then the game reaches the time period $t + Q\,S$. So, in order for individuals in this population to play the same strategy as in the outcome x^0 till period $t + Q\,S$, it may have to obtain the strategies used as far back as period $t - S + 1$, and $\frac{S}{H} \leq \frac{1}{\Sigma_{i=1}^{N}|X_i| - min\{|X_1|, \ldots, |X_N|\} - N + 1}$ is sufficient to ensure this. □

Proof of Proposition 2 (i) It is sufficient to show that there is positive probability of a transition from any convention $\omega_{(x_i, x_{-i})} \in Co(G) \backslash Co(G)_i$ to the convention $\omega_{(x'_i, x'_{-i})} \in Co(G)_i$ when an individual in population $-i$ makes a mistake in his choice of strategy. So, suppose that state is $\omega_{(x_i, x_{-i})}$, and the population $-i$ player experiments with $x'_{-i} \in X_{-i}$ in some time period t. Then, with positive probability, in all the time periods from $t + 1$ to $t + H$, the population i individual draws a sample that contains this strategy x'_{-i}; then, extreme optimism implies that $AS_i(t + k) = \{x'_{-i}\}$, and so, $R_i(s_i(t + k)) = \{x'_i\}$ for all $k = 1, \ldots, H$. Also, with positive probability, in each time period from $t + S + 1$ to $t + H$, the population $-i$ player's sample comprises of the strategies used in S periods immediately preceding it; this sample is monomorphic, and comprises only of x'_i; because the behavioural rule of the individuals in the other population is responsive and because the outcome (x'_i, x'_{-i}) is a sink, these individuals play x'_{-i} in all the time periods from $t+S+1$ to $t+H$. Hence, the same outcome (x'_i, x'_{-i}) obtains in all periods from $t+S+1$ to $t+H$.

Now, in all the $H-S$ periods from $t+H+1$ to $t+2H-S$, there is a positive probability of the sample drawn by the players from either population comprising of the strategies used in the previous S periods. These randomly drawn samples are monomorphic, and the individuals in population i [population $-i$] observe the individuals in the other population playing only x'_{-i} [x'_i]; consequently, they play x'_i [x'_{-i}] in all periods from $t+H+1$ to $t+2H-S$. Then, the H period history at the end of time period $t+2H-S$ is comprised only of the sink outcome (x'_i, x'_{-i}). Thus, a transition to the convention $\omega_{(x'_i, x'_{-i})}$ from any other convention is possible by a single mistake. Consequently, this convention is always in the stochastically stable set.

(ii) The argument for this part has been provided in the main body of the paper (just before the statement of Proposition 2). □

References

Alós-Ferrer, C., Ania, A.B.: The evolutionary stability of perfectly competitive behavior. Econ. Theory **26**(3), 497–516 (2005)

Alós-Ferrer, C., Netzer, N.: The logit-response dynamics. Games Econ. Behav. **68**(2), 413–427 (2010)

Alós-Ferrer, C., Netzer, N.: On the convergence of logit-response to (strict) nash equilibria. Econ. Theory Bull. **5**(1), 1–8 (2017)

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

Bergin, J., Bernhardt, D.: Cooperation through imitation. Games Econ. Behav. **67**(2), 376–388 (2009)

Bilancini, E., Boncinelli, L.: Social coordination with locally observable types. Econ. Theory **65**(4), 975–1009 (2018)

Bilancini, E., Boncinelli, L.: The evolution of conventions under condition-dependent mistakes. Econ. Theory **69**(2), 497–521 (2020)

Blume, L.E.: The statistical mechanics of strategic interaction. Games Econ Behav **5**(3), 387–424 (1993)

Blume, L.E.: Population Games. In: Arthur, B., Durlauf, S., Lane, D. (eds.) The Economy as an Evolving Complex System II, pp. 425–460. Addison-Wesley, Boston (1997)

Ellison, G.: Basins of attraction, long run stochastic stability, and the speed of step-by-step evolution. Rev. Econ. Stud. **67**(1), 17–45 (2000)

Foster, D., Young, H.P.: Stochastic evolutionary game dynamics. Theor. Popul. Biol. **38**(2), 219–232 (1990)

Hedlund, J.: Imitation in cournot oligopolies with multiple markets. Econ. Theory **60**(3), 567–587 (2015)

Hurkens, S.: Learning by forgetful players. Games Econ. Behav. **11**(2), 304–329 (1995)

Hwang, S.H., Newton, J.: Payoff-dependent dynamics and coordination games. Econ. Theory **64**(3), 589–604 (2017)

Juang, W.T.: Rule evolution and equilibrium selection. Games Econ. Behav. **39**(1), 71–90 (2002)

Josephson, J., Matros, A.: Stochastic imitation in finite games. Games Econ. Behav. **49**(2), 244–259 (2004)

Josephson, J.: Stochastic better-reply dynamics in finite games. Econ. Theory **35**(2), 381–389 (2008)

Josephson, J.: Stochastic adaptation in finite games played by heterogeneous populations. J. Econ. Dyn. Control **33**(8), 1543–1554 (2009)

Kandori, M., Mailath, G., Rob, R.: Learning, mutation, and long run equilibria in games. Econometrica **61**(1), 29–56 (1993)

Kaniovski, Y.M., Kryazhimskii, A.V., Young, H.P.: Adaptive dynamics in games played by heterogeneous populations. Games Econ. Behav. **31**(1), 50–96 (2000)

Karandikar, R., Mookerjee, D., Ray, D., Vega-Redondo, F.: Evolving aspirations and cooperation. J. Econ. Theory **80**(2), 292–331 (1998)

Khan, A.: Evolutionary stability of behavioural rules in bargaining. J. Econ. Behav. Organ. (2021) (**Forthcoming**)

Khan, A., Peeters, R.: Cognitive hierarchies in adaptive play. Int. J. Game Theory **43**(4), 903–924 (2014)

Matros, A.: Clever agents in adaptive learning. J. Econ. Theory **111**(1), 110–124 (2003)

Robson, A.J., Vega-Redondo, F.: Efficient equilibrium selection in evolutionary games with random matching. J. Econ. Theory **70**(1), 65–92 (1996)
Sáez-Marti, M., Weibull, J.: Clever agents in Young's evolutionary bargaining model. J. Econ. Theory **86**(2), 268–279 (1999)
Vega-Redondo, F.: The evolution of Walrasian behavior. Econometrica **65**(2), 375–384 (1997)
Young, H.P.: The evolution of conventions. Econometrica **61**(1), 57–84 (1993)
Young, H.P.: Individual Strategy and Social Structure: An Evolutionary Theory of Institutions. Princeton University Press, Princeton (1998)