Stochastic adaptation in finite games played by heterogeneous populations

Jens Josephson a,b,*

a Department of Economics, Stockholm University, 106 91 Stockholm, Sweden
b Department of Economics and Business, Universitat Pompeu Fabra and Barcelona GSE, Ramon Trias Fargas 25-27, 08005 Barcelona, Spain

Abstract

We analyze stochastic adaptation in finite n-player games played by heterogeneous populations containing best repliers, better repliers, and imitators. Individuals select strategies by applying a personal learning rule to a sample from a finite history of past play. We give sufficient conditions for convergence to minimal closed sets under better replies and selection of a Pareto dominant such set. Finally, we demonstrate that the stochastically stable states are sensitive to the sample size by showing convergence to the risk-dominant equilibrium for sufficiently small sample size and to the Pareto-dominant equilibrium for sufficiently large sample size in 2 × 2 coordination games.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

In a seminal paper, Young (1993a) obtained long-run predictions for coordination games and weakly acyclic games in a model where homogeneous populations of best-repliers, who choose a myopic best reply to a random sample from a finite history of past play. Using a variant of a theorem of Hurkens (1995), he extended these results to generic finite games in Young (1998). Other papers have analyzed evolutionary dynamics in a similar framework assuming homogeneous populations of individuals using other types of adaptive behavior. In particular, Josephson and Matros (2004) have studied the case of homogeneous populations of imitators, who imitate the most successful strategy in a sample of past play, and Josephson (2008a) has considered the case of homogeneous populations of better repliers, who choose any strategy that gives a weakly higher expected payoff against a sample distribution of the opponents’ strategies than a sample of the own population’s past strategy choices.1

A natural question is whether any of the results in these papers extend to the case of heterogeneous populations, containing individuals of all of the three types (and possibly many others). In particular, one may ask how the long-run predictions depend on the population shares of different types of individuals and if any type of behavior has a greater impact on the outcome than others. This is the topic of this paper.
We first analyze stochastic learning in finite n-player games played by heterogeneous populations of individuals whose behaviors share a certain property which can be summarized as: “If all the strategy profiles of the finite history are in a minimal closed set under better replies (MCUBR set) of the game, then do not play a strategy outside this set.”

Better repliers, best repliers, and imitators all have this property, but so do many other behaviors such as randomizing over the strategies in a sample from a finite history of past play or playing the most prevalent strategy in a sample from a finite history of past play.

We prove that for any finite game, if the ratio between the sample and history size is sufficiently small for at least one better replier in each population, the resulting Markov chain converges with probability one to a MCUBR set of the game. This result is independent of population shares, as long as the share of better repliers is positive. Moreover, by Josephson (2008a), it implies convergence to a strict Nash equilibrium in a number of well-known classes of games. It is also similar to previous results in a different framework. Ritzberger and Weibull (1995) show that for a large class of deterministic selection dynamics in continuous time, including the replicator dynamics, a product set is asymptotically stable if and only if it is a MCUBR set.

We then restrict the populations to contain only better-repliers, best repliers, and imitators. We also perturb the stochastic process in a standard fashion by assuming that with a small probability, the individuals make mistakes or experiment and play a pure strategy at random according to some fixed probability distribution with full support. This assumption makes the process irreducible and aperiodic, and thus implies a unique stationary distribution. We calculate the support of this distribution as the probability of mistakes tends to zero using the techniques of Freidlin and Wentzell (1984).

If the sample size is sufficiently large, the ratio between the sample and history size sufficiently small for at least one better replier in each population, and one MCUBR set strictly Pareto dominates all strategy-tuples outside the set, then the perturbed Markov chain puts probability one on this set as the level of noise tends to zero. A corollary of this is that the strictly Pareto-dominant equilibrium is selected in games of common interest and 2 \times 2 coordination games.

However, the minimum sample size required for this result depends on the payoffs of the game. In fact, if all individuals have the same sample size, then for sample sizes below a certain critical level, a unique risk-dominant equilibrium is often selected in 2 \times 2 coordination games (and always in symmetric such games).

The basic setting in this paper is similar to that in Young (1993a, 1998), but the paper is perhaps closest in spirit to Hurkens (1995). The latter proves convergence to minimal closed sets under best replies (MCUBR sets) when individuals best-reply to a sample drawn with replacement and also allows for the presence of either imitators, sophisticated players, or other types of behaviors. Our paper is different in that it introduces better repliers in all populations, allows for a large class of behaviors, assumes sampling without replacement, and also contains an analysis of the perturbed Markov chain.

There are also other papers on stochastic learning with heterogeneous populations, but they generally focus on special classes of games and restrict the number of behaviors in each population.

In Sáez Martí and Weibull (1999), Young’s (1993b) evolutionary version of Nash’s demand game is played by one population of myopic best repliers only, and another population of best repliers and clever agents who play a best reply to the best reply. The authors show that Young’s predictions are still valid in the presence of such clever agents. Matros (2000) extends this result to finite two-player games.

Kaarbøe and Tieman (1999) use the Kandori et al. (1993) framework, where all individuals of a finite population are matched in each period, to study strictly supermodular games played by myopic best repliers and imitators and show selection of the Pareto efficient equilibrium.

Gale and Rosenthal (1999), analyze the dynamics when imitators and experimenters play a symmetric game with a unique equilibrium. They show global convergence to a compact neighborhood of the equilibrium, but also find that the local behavior depends on the details of the model.

Kaniovski et al. (2000) study adaptive dynamics in 2 \times 2 coordination games played by heterogeneous populations of myopic best repliers, conformists (who do what the majority does), and nonconformists (who do the opposite of what the majority does). They show that the resulting process may have limit cycles even when the proportion of non-best repliers is arbitrary small.

Schipper (2007) analyzes a symmetric Cournot game with heterogeneous populations of imitators and myopic best repliers, who interact in each period. He shows that the stochastically stable states of the process correspond to Stackelberg equilibria where the imitators are better off than the best repliers. On the other hand, Hehenkamp and Kaarbøe (2008) show, in the setting of a symmetric two-player game with an imitator and a myopic best replier, that there is evolutionary pressure against imitators in a changing environment under certain conditions.

A few papers also allow individuals to switch behavior as a function of their past performance. Conlisk (1980) analyze a quadratic setting with optimizers, who react to a correct perception of society-wide conditions, and imitators, who react to a convention. He finds that in the presence of information costs, imitators may have as high fitness as optimizers.

Brock and Hommes (1997) analyze a demand-supply cobweb model with heterogeneous populations of individuals who choose between costless naive expectations and costly rational expectations. They show that highly irregular equilibrium prices may result if the intensity of choice between the two predictors is large.

\footnote{This set concept was introduced by Basu and Weibull (1991) and Ritzberger and Weibull (1995). For a definition, see Section 2.}
Droste et al. (2002) study a Cournot symmetric duopoly played by individuals drawn from heterogeneous populations with a finite set of learning rules with different information costs. Assuming the fractions of different rules are updated according to noisy replicator dynamics, the authors find that the long-run behavior of the system may be complicated and that endogenous fluctuations may arise.

Juang (2002) studies a $2 \times 2$ coordination game in a Kandori et al. (1993) setting with myopic best replier and imitators. If agents cannot change rules, then equilibrium selection is determined by the relative frequencies or the rules. However, if agents can change rule by adaptation, then the Pareto efficient equilibrium is selected.

Thijssen (2005) considers a symmetric Cournot game where firms are myopic optimizers with different conjectures about how other firms will respond to quantity changes. Assuming the different populations fractions are updated sufficiently infrequently, he shows convergence to the Walrasian equilibrium.

Josephson (2008b) studies a general class of parametric learning rules. He uses simulations to investigate which learning rules are evolutionarily stable, in the sense of being robust to a small invasion by a different learning rule in the class.

Finally, there are a couple of related papers with homogeneous populations. Antoci et al. (2008) consider a deterministic $2 \times 2$ coordination game in a Kandori et al. (1993) setting with myopic best replier and imitators. Antoci et al. (2008) also study a $2 \times 2$ coordination game with a finite set of learning rules with different information costs. Assuming the fractions of different rules are updated under what conditions analytical concepts from evolutionary game theory such as stochastic stability and continuous-time dynamics are useful in predicting the behavior of several interacting finite populations of agents. He shows that there are instances where continuous-time dynamics give poor predictions and where stochastic stability has no bite.

This paper is organized as follows. We start by describing the model in Section 2. In Section 3, we present general results for finite games. In Section 4, we give sufficient conditions for convergence to Pareto-dominant sets and in Section 5, we illustrate how the long-run distribution depends crucially on the sample size in the setting of $2 \times 2$ coordination games. Section 6 concludes. All omitted proofs can be found in the Appendix.

2. The model

The basic setting is similar to that of Young (1993a, 1998), although our notation is slightly different. Let $I$ be a finite $n$-player game in strategic form. Let $X_i$ be the finite set of pure strategies $x_i$ available to player $i \in \{1, \ldots, n\} = N$ and let $A(X_i)$ be the set of probability distributions $p_i$ over these pure strategies. Define the (full dimensional) product sets $X = \prod_{i \in N} X_i$ and $\Box(X) = \prod_{i \in N} A(X_i)$ with typical elements $x$ and $p$, respectively. Let $p_i(x_i)$ denote the probability mass on pure strategy $x_i$ and let $p(x) = \prod_{i \in N} p_i(x_i)$. We write $x_{-i} \in \prod_{j \neq i} X_j = X_{-i}$ and $p_{-i} \in \prod_{j \neq i} A(X_j) = \Box(X_{-i})$ to represent the pure strategies and the distributions of pure strategies of player $i$'s opponents.

Let $C_1, \ldots, C_n$ be $n$ finite and non-empty populations of individuals using personal learning rules, to be defined below. The populations need not be of equal size, nor do they necessarily have equal shares of individuals using different learning rules. Each member of population $C_i$ is a candidate to play role $i$ in the game $I$ and has payoffs represented by the Bernoulli function $\pi_i : X \to \mathbb{R}$, and expected payoffs represented by the function $u_i : \Box(X) \to \mathbb{R}$. In slight abuse of notation, we write $u_i(x_i, p_{-i})$ instead of $u_i(p_i, p_{-i})$ if $p_i(x_i) = 1$.

Let $t = 1, 2, \ldots$ denote successive time periods. The stage game $I$ is played once in each period. In period $t$, one individual is drawn at random from each of the $n$ populations and assigned to play the corresponding role. The individual in role $i$ chooses a pure strategy $x_i$ from her strategy space $X_i$, with a probability determined by her personal learning rule. The realized pure-strategy profile $x^t = (x_1^t, \ldots, x_n^t)$ is recorded and referred to as the play at time $t$. The history of play up to time $t$ is the sequence $h^t = (x^1, \ldots, x^t)$.

Strategies are chosen as follows. Fix a positive integer $m > 1$, the memory size of all individuals. For each $i \in N$, let $F^m_i$ be the set of learning rules, the class of functions $f^m_i : X^m \to A(X_i)$ that map the set of vectors of pure-strategy profiles of length $m$ to the set of mixed strategies corresponding to player role $i$. Each individual $k$ in population $C_i$ is endowed with a learning rule $f^m_i \in F^m_i$. If drawn to play in a given period $t + 1 > m$, her strategy choice $x_i^{t+1}$ is the realization of a random variable with conditional probability distribution $f^m_i(h^m_i)$, where $h^m_i = (x^t, \ldots, x^t)$ is the history of the most recent realizations of strategy profiles.

We will make two assumptions about the learning rules in the populations. The first assumption is that a positive fraction of each population consists of better-repliers. Let $\gamma_i : \Box(X) \to X$ be the better-reply correspondence, defined by

$$\gamma_i(p) = \{x_i \in X_i | u_i(x_i, p_{-i}) \geq u_i(p_i)\}. \quad (1)$$

Also, let $s$ be an arbitrary positive integer, the sample size of the individual. Finally, let $f^{m,s}_i(x_i|h)$ be the conditional probability that a better replier with sample size $s$ drawn to play in role $i$ chooses pure strategy $x_i$, given the history $h$ of the

---

3 When we henceforth refer to product sets, i.e. sets of the form $\prod_{i \in M} Y_i$ for some $M \subseteq N$, we will always assume that they are full dimensional in the sense that $M = N$.

4 The better-reply correspondence has a number of interesting properties, as discussed in Josephson (2008a). First, the image of the best-reply correspondence $\gamma_i(p)$ is always included in the image of the better-reply correspondence $\gamma_i(p)$. Second, if a strategy $x_i$ is a better reply to some mixed-strategy profile $p$, then it is also a better reply to a pure-strategy profile $y$ in the support of $p$. Third, a pure-strategy profile $x$ with corresponding mixed-strategy representation $p$ is a strict Nash equilibrium if and only if $\gamma_i(p) = x_i$ for all $i \in N$. 

m most recent realizations of strategy profiles. We assume \( f_{ij}^m(x_i|h_0) > 0 \) if and only if there exists a sample of size \( s \leq m \), consisting of \( n \) independent draws without replacement of all player roles' \( m \) most recent strategy realizations, with corresponding empirical distribution \( \hat{p} \in \square(X) \), such that \( x_i \in \gamma_i(\hat{p}) \).

The second assumption about that the learning rules in the populations is that they all share a particular property. Let \( \mathcal{A} \) be the collection of all non-empty product sets \( Y \subseteq X \). Let \( \mathcal{A}(Y) \) be the set of probability distribution with support in \( Y \), and let \( \square(Y) = \prod_{i \in n} \mathcal{A}(Y_i) \) be the corresponding product set. Following Basu and Weibull (1991), and Ritzberger and Weibull (1995), we say that a set \( Y \in \mathcal{A} \) is closed under better replies (CUBR) if \( \gamma(\square(Y)) \subseteq Y \), where \( \gamma(p) = \prod_{i \in n} \gamma_i(p) \). A set \( Y \in \mathcal{A} \) is a MCUBR if it is CUBR and contains no proper subset with this property.

We will assume that all learning rules represented in the populations prescribe play of a strategy in a MCUBR set if the history of the \( m \) most recent strategy profiles contains strategy profiles only from that set. Formally, if some individual \( k \) in some population \( C_i \) is using the learning rule \( f_{ik}^m \in \mathcal{A}^m \), then, for any MCUBR set \( \gamma \) and any history \( h \) of length \( m \) that contains only strategy profiles in \( Y \), \( f_{ik}^m(h) \in \mathcal{A}(Y) \). We will call learning rules with this property returning.\(^5\)

It is obvious that the learning rule of the better repliers have this property. However, so do many other rules such as those of imitators, who play the strategy with the highest (average) realized payoff in a sample from the state, conformists, who play the most prevalent strategy in a sample from the state, randomizers, who choose any strategy at random of the strategies in a sample from the state, best repliers, who play a myopic best reply to the empirical sample distribution, and sophisticated players, who play a best reply to a myopic best reply to the sample distribution by the other players.

Starting from an arbitrary initial sequence of \( m \) plays, \( h^m = (x^1, \ldots, x^m) \), this defines a Markov chain on the finite state space \( H = X^m \) of histories of length \( m \). We will denote this process by \( P^m \) and generally refer to it as heterogeneous adaptive play or the unperturbed process. Given a state \( h = (x^1, \ldots, x^m) \) at time \( t \), the process moves to a state of the form \( h' = (x^{t+1}, \ldots, x^m) \) in the next period. Such a state is called a successor of \( h \). For each \( x_i \in X_i \), let \( P_i(x_i|h_i) \) be the conditional probability that the individual drawn to play in role \( i \) chooses pure strategy \( x_i \), given state \( h \). Our behavioral assumptions imply that \( P_i(x_i|h_i) \) is independent of \( t \) and that \( P_i(x_i|h_i) > 0 \) if and only if there is a \( k \in C_i \) such that \( f_{ik}^m(h) \) puts positive probability on \( x_i \). If \( x \) is the rightmost element of \( h' \), the probability of moving from \( h \) to \( h' \) is \( P_{ih}^{m,n} = \prod_{i=1}^{n} P_i(x_i|h_i) \) if \( h' \) is a successor of \( h \) and 0 if \( h' \) is not it.

Several concepts from Markov-chain theory are used below. A recurrent class \( E_h \) is a set of states such that there is zero probability of moving from any state in the class to any state outside, and a positive probability of moving from any state in the class to any other state in the class in a finite number of periods. A state \( h \) is absorbing if it constitutes a singleton recurrent class. A state is monomorphic if it consists of a single pure-strategy profile \( x \in X \) repeated \( m \) times. We denote a monomorphic state by \( h_{ma} = (x, \ldots, x) \). The basin of attraction of a state \( h' \) is the set of states \( h \) such that there is a positive probability of moving from \( h \) to \( h' \) in a finite number of periods.

We also define a perturbed process in a manner completely analogous to Young (1993a, 1998). Formally, in each period there is some small probability \( \varepsilon \) that the individual in role \( i \) experiments or makes a mistake and chooses a pure strategy at random from \( X_i \), instead of according to her learning rule. The event that the individual in role \( i \) experiments is assumed to be independent of the event that the individual in role \( j \) experiments for every \( j \neq i \). For every role \( i \), let \( Q_i(x_i|h_i) \) be the conditional probability that the individual in role \( i \) chooses \( x_i \), given that she experiments and the process is in state \( h \). We assume that \( Q_i(x_i|h_i) \) is independent of \( t \) and decision rules, and that \( Q_i(x_i|h_i) > 0 \) for all \( x_i \in X_i \). Suppose that the process is in state \( h \) at time \( t \). Let \( J \) be a subset of \( j \) players. The probability is \( \varepsilon/(1 - \varepsilon)^{n - \varepsilon} \) that exactly the players in \( J \) experiment and the others do not. Conditional on this event, the transition probability of moving from \( h \) to \( h' \) is \( Q_{nh} = \prod_{i \in J} Q_i(x_i|h_i) \prod_{j \notin J} P_i(x_j|h_j) \) if \( h' \) is a successor of \( h \) and \( x \) is the rightmost element of \( h' \) and 0 if \( h' \) is not a successor of \( h \). This gives the following transition probability of the perturbed Markov chain:

\[
P_{nh}^{m,n} = (1 - \varepsilon)^{nm} P_{nh}^{m,n} + \sum_{j \in H \neq h} \varepsilon^{j} (1 - \varepsilon)^{n - \varepsilon} Q_{nh}^{m,n}.
\]

We denote this process \( P_{nh}^{m,n} \), and generally refer to it as heterogeneous adaptive play with mistakes or the perturbed process.

This process is irreducible and aperiodic, and thus, has a unique stationary distribution \( \mu'\). We study this distribution as \( \varepsilon \) tends to zero. In our analysis, we use the following standard definitions, due to Freidlin and Wentzell (1984) and Foster and Young (1990). A state \( h \) is stochastically stable if \( \lim_{\varepsilon \to 0} \mu' (h) \) exists and is positive. A mistake in the transition from a state \( h = (x^1, \ldots, x^m) \) to a successor \( h' = (x^2, \ldots, x^m, y) \) is a component \( y \) of \( y \) such that \( P(y|h) = 0 \). For any two states \( h, h' \), the resistance \( r_{hh'} \) is the total number of mistakes involved in the transition from \( h \) to \( h' \) if \( h' \) is a successor of \( h \), and otherwise \( r_{hh'} = \infty \). For each pair of distinct recurrent classes, \( E_h \) and \( E_i \), a kl-path is a sequence of states \( \zeta = (h^1, h^2, \ldots, h^k) \) beginning in \( E_h \) and ending in \( E_i \). The resistance of this path is the sum of the resistances on the edges that compose it. Let \( r_{kl} \) be the least resistance over all kl-paths. Construct a complete directed graph with one vertex for each recurrent class. The weight on the directed edge \( E_h \to E_i \) is \( r_{kl} \). A tree rooted at \( E_i \) is a set of directed edges such that, from every vertex different from \( E_i \), there is a unique directed path in the tree to \( E_i \). The resistance of a rooted tree is the sum of the resistances on the edges that compose it. The stochastic potential \( \rho(E_i) \) of a recurrent class \( E_i \) is the minimum resistance over all trees rooted at \( E_i \).

\(^5\) The returning property is here defined with respect to the game \( \Gamma \). However, the results in this paper would remain unaltered if we strengthened the assumption by requiring that the learning rules have this property for all finite games.
3. General results

In this section, we will present our main theorem on the asymptotic distributions of the perturbed and unperturbed processes. In order to state our results, one more definition is needed. Let $H' \subseteq H$ be an arbitrary set of states. We say that $H'$ spans $Y \in X$ if $Y$ is the product set of all strategies that appear in some state in $H'$.

**Theorem 1.** If there is at least one better replier with $s/m \leq 1/|X|$ in each population and all individuals have returning learning rules:

(i) The unperturbed process $P^m$ converges almost surely to a set of states that spans a MCUBR set.

(ii) The set of stochastically stable states of $P^m$ is a (union of) set(s) of states that (each) spans a MCUBR set and has minimum stochastic potential.

The first part of this theorem is analogous to Theorem 1 in Josephson (2008a), which states that in a homogeneous setting, where the share of better repliers in each population is one, for a sufficiently low ratio of sample size and memory size $(s/m \leq 1/|X|)$, the unperturbed process converges with probability one to a set of states that spans a MCUBR set. The proof of Theorem 1 is based on this result, the fact that if the process reaches a state involving strategies in a MCUBR set, returning learning rules would never play a strategy outside this set, and the observation that in each period, there is a positive probability that only better repliers with the required sample size will be drawn to play from the heterogeneous populations. This is also what makes the predictions different from those in models where all members of a single population are matched in every period, such as Schipper (2007).

The second part of Theorem 1 is a mere application of Theorem 3.1 in Young (1998) (restated as Lemma 2 in the Appendix), which says that the set of stochastically stable states are those contained in the recurrent classes with minimum stochastic potential.

Theorem 1 is also similar to previous results for deterministic continuous-time selection dynamics. Ritzberger and Weibull (1995) show that a set is MCURB if and only if it is asymptotically stable for regular selection dynamics that are sign preserving. This is a large class of selection dynamics, which contain several well known dynamics, such as the replicator dynamics.

It is worth noting that Theorem 1 makes no reference to population mixtures. As long as the share of better repliers with the required sample size is positive in all populations, the span of the recurrent sets of the unperturbed process is independent of the shares of different learning rules and sample sizes in the populations. However, it is clear that this does not imply that the exact shape of the asymptotic distribution is independent of the population mixtures. Moreover, as will be shown below, the set of stochastically stable states may depend upon which learning rules are present in the populations.

In games where all MCUBR sets are singleton, and hence also strict Nash equilibria, Theorem 1 implies convergence to a convention, a monomorphic state which is a repetition of a strict Nash equilibrium (Young, 1993a, 1998). By Josephson (2008a), the set of games with this property contains a number of well-known classes of games such as weakly acyclic games, finite dominance solvable games, super-modular games where all pure-strategy Nash equilibria are strict, (ordinal) potential games, transversal quasi-concave two-player games, and almost all transversal $n$-player aggregative games.

4. Convergence to Pareto-dominant MCUBR sets

The second result of this paper characterizes the stochastically stable states of heterogeneous play in finite $n$-player games with a Pareto optimal MCUBR set when all populations consist of better repliers and imitators.

If an imitator is drawn, she observes a sample of size $s \in \{1, \ldots, m\}$ of population-specific strategy and payoff realizations and chooses a pure strategy which has maximum average payoff among the strategies included in the sample. The average payoff is computed by summing all the realized payoffs when the strategy was used in the sample and dividing by the number of instance of the strategy in the sample.

We say that a non-empty set of strategy-tuples $Z \subset X$ strictly Pareto dominates a pure-strategy profile $y \in X \setminus Z$ if, for all $i \in N$,

$$\min_{x \in Z} \pi_i(x) > \pi_i(y).$$

6. If there are no better repliers with a sufficiently small sample size compared to the memory size, certain strategies of the MCUBR sets may never be played. See the example in Appendix B.
7. The case when all individuals are best-repliers is covered by Young (1998), who shows convergence to a set of states that spans a MCURB set in generic games.
8. The result holds also if there are also best repliers in the populations.
9. The results in this section would also hold if some or all of the imitators instead chose a strategy with maximum empirical payoff in the sample.
Theorem 2. Suppose the game has a MCUBR set, Y, that strictly Pareto dominates all other pure-strategy profiles. If all populations consist of imitators and better repliers, at least one better replier in each population has a sample size such that 

s/m ≤ 1/|X|,

and all individuals have sufficiently large sample size, then the set of stochastically stable states of \( P^m \) spans Y.

Theorem 2 gives sufficient conditions for convergence to a Pareto-dominant MCUBR set. In *games of common interest*, which have a strict Nash equilibrium that strictly Pareto dominates all other strategy-tuples, it implies convergence to a Pareto-dominant equilibrium. The result is similar to Theorem 4, for a homogeneous populations of imitators, in Josephson and Matros (2004).

The proof of Theorem 2 uses the following two observations. First, a state containing only strategy profiles in a strictly Pareto-dominant set \( Y \) can be reached from any state outside the set if all individuals drawn to play simultaneously make a mistake and play a pure strategy in \( Y \), and a sequence of imitators thereafter pick samples including this mistake. Second, the resistance of the reverse transition can be made arbitrarily large by choosing a sufficiently large sample and memory size.

Theorem 2 does not hold under the weaker condition that an MCUBR set strictly Pareto dominates the pure-strategy profiles of all other MCUBR sets. Consider the game in Fig. 1 and assume for simplicity that all individuals have a common sample size \( s \).

In this game, there are two MCUBR sets, \( \{(A, a)\} \) and \( \{(C, c)\} \), and the first strictly Pareto dominates \( (C, c) \). However, \( \{(A, a)\} \) does not strictly Pareto dominate \( (B, b) \), and for a sufficiently large sample size and sufficiently small ratio between the sample and memory size, the stochastic potential is two for both \( h_{(A,a)} \) and \( h_{(C,c)} \). This follows since the process will make the transition from \( h_{(A,a)} \) to \( h_{(B,b)} \), if the two players simultaneously make mistakes in period \( t \) and play \( (B, b) \), a sequence of \( s - 1 \) imitators thereafter are drawn to play in the position of the row player and sample only \( x^t \) and earlier strategy realizations, and finally imitators in both populations sample only from plays more recent than \( x^{t-s} \) for \( m - s \) periods. Since state \( h_{(B,b)} \) clearly is in the basin of attraction of \( h_{(C,c)} \), this implies that the stochastic potential of \( h_{(C,c)} \) is two at most.

5. Sample size and stochastic stability in 2 × 2 coordination games

In this section, we will study the predictions in the special class of 2 × 2 coordination games in order to illustrate how the set of stochastically stable states may depend on the sample size.\(^{10}\) We will maintain the assumption that the populations all consist of better replier and imitators and for simplicity assume that all individuals have a common sample size, \( s \).\(^{11}\)

Consider the game in Fig. 2. This game is a 2 × 2 coordination game if \( (A, a) \) and \( (B, b) \) are strict Nash equilibria. It is a symmetric 2 × 2 coordination game if, in addition, the diagonal payoffs are equal for the two players, \( \pi_1(B, a) = \pi_2(A, b) \), and \( \pi_1(A, b) = \pi_2(B, a) \).

An equilibrium \( (A, a) \) of a 2 × 2 coordination game is risk dominant if its Nash product exceeds that of \( (B, b) \):

\[
(\pi_1(A, a) - \pi_1(B, a))(\pi_2(A, a) - \pi_2(B, a)) > (\pi_1(B, b) - \pi_1(A, b))(\pi_2(B, b) - \pi_2(B, a)).
\]

This definition (with a strict inequality) is originally due to Harsanyi and Selten (1988). Recall that if \( x \) is a strict Nash equilibrium, then \( h_x \) is called a convention. If \( x \) is risk (strictly Pareto) dominant, we say that the convention \( h_x \) is risk (strictly Pareto) dominant. From Theorem 2, the following corollary immediately follows.

Corollary 1. In 2 × 2 coordination games with a strictly Pareto-dominant equilibrium, for a sufficiently large sample size, and sufficiently low ratio between the sample and memory size, the strictly Pareto-dominant convention is a unique stochastically stable state.

\(^{10}\) For an analysis of dependency on memory size in a homogeneous Cournot setting see Alós-Ferrer (2004).

\(^{11}\) The results in this section hold also if all or some of the better repliers are replaced by best repliers. If the imitators are replaced by individuals choosing a strategy with maximum empirical payoff in the sample, the only modification is that the threshold in Theorem 3(ii) becomes \( 2/q_{\text{max}} \).
For certain payoffs, the sample size must be very large for this result to hold. Consider the game in Fig. 3, where the equilibrium \((A, a)\) strictly Pareto dominates the equilibrium \((B, b)\), but where \((B, b)\) is the unique risk-dominant equilibrium.

If the process is in state \(h_{(B,B)}\) and the sample size \(s>1\), then two simultaneous mistakes, followed by a sequence of imitators, are required to reach \(h_{(A,A)}\). One mistake is sufficient to make the reverse transition if this either makes the expected payoff to playing strategy \(B\) higher than that of \(A\) or the expected payoff to playing strategy \(b\) higher than that of \(a\) for a better replier, or formally if at least one of the following two conditions hold:

\[
\frac{2 \cdot (s - 1) - 997}{s} \leq \frac{0 \cdot (s - 1) + 1}{s} \iff s \leq 500,
\]

\[
\frac{2 \cdot (s - 1) - 497}{s} \leq \frac{0 \cdot (s - 1) + 1}{s} \iff s \leq 250.
\]

Hence, only the risk-dominant equilibrium is selected, in the sense that the corresponding convention is stochastically stable, for \(s\) such that \(1 < s \leq 500\). In order to ensure that only the strictly Pareto-dominant equilibrium is selected, the sample size must be so large that both of the following two conditions hold:

\[
\frac{2 \cdot (s - 2) - 997 \cdot 2}{s} > \frac{0 \cdot (s - 2) + 1 \cdot 2}{s} \iff s > 999,
\]

\[
\frac{2 \cdot (s - 2) - 497 \cdot 2}{s} > \frac{0 \cdot (s - 2) + 1 \cdot 2}{s} \iff s > 499.
\]

For intermediate sample sizes, i.e. \(s\) such that \(500 < s \leq 999\), both equilibria are selected.

More generally, define the probabilities \(q_A, q_a\), and \(q_{\text{min}}\) by

\[
q_A = \frac{\pi_2(B, b) - \pi_2(B, a)}{\pi_2(A, a) - \pi_2(A, b) + \pi_2(B, b) - \pi_2(B, a)},
\]

\[
q_a = \frac{\pi_1(B, b) - \pi_1(A, b)}{\pi_1(A, a) - \pi_1(A, b) + \pi_1(B, b) - \pi_1(B, a)},
\]

and

\[
q_{\text{min}} = \min\{q_A, q_a, 1 - q_A, 1 - q_a\}.
\]

In other words, \(q_A\) is the probability of strategy \(A\) which makes the expected payoff to strategies \(a\) and \(b\) identical for player \(2\), and \(q_a\) is the probability of strategy \(a\) which makes the expected payoff to strategies \(A\) and \(B\) identical for player \(1\). \(q_{\text{min}}\) is the lowest probability required to make a player indifferent between her pure strategies. Note that in symmetric coordination games \(q_A = q_a\) and that (see Young, 1998) an equilibrium \((A, a)\) is risk dominant if and only if

\[
\min\{q_A, q_a\} \leq \min\{1 - q_A, 1 - q_a\}.
\]

Let \(\mathbb{N}_+\) be the set of positive integers, \(\mathbb{N}\) the set of non-negative integers, and let \([y]\) denote the smallest integer greater than or equal to \(y\) for any real \(y\). Further, define the function \(s_{\text{diff}} : \mathbb{N}_+ \rightarrow \mathbb{N}\) and the set \(S^{RD} \subset \mathbb{N}_+\) as follows:

\[
s_{\text{diff}}(s) = [|\{s \min\{q_A, q_a\}\}] - [|\min\{1 - q_A, 1 - q_a\}|],
\]

\[
S^{RD} = \{s \in \mathbb{N}_+: s_{\text{diff}}(s) \geq 1\} \text{ and } s \leq 1/q_{\text{min}}.
\]

In words, \(s_{\text{diff}}(s)\) is the absolute difference in the number of mistakes necessary to make a better replier switch from each of the two equilibria. \(S^{RD}\) is the set of sample sizes such that this difference is at least one and such that the minimum number of mistakes necessary for a switch is at most one. Note that \(s_{\text{diff}}(s)\) is increasing in \(s\), implying that if there are integer sample sizes \(s, s',\) and \(s''\) such that \(s < s' < s''\), and \(s, s'' \in S^{RD}\), then also \(s' \in S^{RD}\). Note also that \(S^{RD}\) may be empty.

Finally, for \(i = 1, 2\) define

\[
\zeta_i(x) = \frac{2\pi_i(x) - \pi_i(x_i, x_{-i}) - \pi_i(x_i, x'_{-i})}{\pi_i(x) - \pi_i(x_i, x_{-i})},
\]

where \(x' \neq x_i\) and \(x' \neq x_{-i}\). For a strictly Pareto-dominated convention \(h_i\), \(\zeta_i(x)\) is the maximum sample size such that an imitator in player position \(i\) can switch strategy after only two previous mistakes, one in each population.
Theorem 3. Let $\Gamma$ be a $2 \times 2$ coordination game and assume that $s/m \leq \frac{1}{2}$.

(i) From any initial state, the unperturbed process $P^n$ converges with probability one to a convention.
(ii) If the game has a strictly Pareto-dominant equilibrium $x$ and $s > \max\{\hat{s}_i(x), \hat{s}_f(x), 2/q_{\min}\}$, then only the Pareto-dominant convention is stochastically stable.
(iii) If the game has a unique risk-dominant equilibrium and $s \in S^0$, then only the risk-dominant convention is stochastically stable.
(iv) For any game and sample size covered neither by (ii) nor (iii), both conventions are stochastically stable.

Theorem 3 gives sufficient and necessary conditions for the selection of different equilibria. The intuition behind the proof of this theorem is the same as in the above example. If the process is in state $h_{(A,A)}$ and $(B,B)$ is a risk-dominant equilibrium, then, for sample sizes such that $q_{\min} \leq 1/s$, or equivalently $s \leq 1/q_{\min}$, only one mistake is necessary to make a sequence of subsequent better repliers switch to the risk-dominant pure strategy such that $h_{(B,B)}$ is reached. The condition that $s_{\text{diff}}(s) \geq 1$ implies that the transition from $h_{(B,B)}$ to $h_{(A,A)}$ cannot be made with only one mistake followed by a sequence of better repliers. Second, if the process is in state $h_{(B,B)}$ and $(A,A)$ is strictly Pareto dominant, then two simultaneous mistakes, followed by a sequence of imitators in one of the populations, are sufficient to reach the basin of attraction of $h_{(A,A)}$. The condition that $s > \max\{\hat{s}_i(x), \hat{s}_f(x), 2/q_{\min}\}$ implies that the reverse transition will require more than two mistakes.

The following corollary shows that in symmetric coordination games, there always exists a range of sample sizes such that the risk-dominant equilibrium is selected.

Corollary 2. In symmetric $2 \times 2$ coordination games with a unique risk-dominant equilibrium, if $s/m \leq 1/2$ and $2 \leq s \leq 1/q_{\min}$, only the risk-dominant convention is stochastically stable.

Proof. If the game has a unique risk-dominant equilibrium, then $q_{\min} \leq \frac{1}{2}$. This implies that if the sample size is such that $2 \leq s \leq 1/q_{\min}$, then $q_{\min} \leq 1$, and $s(1 - q_{\min}) > 1$; hence, the requirement that $s_{\text{diff}}(s) \geq 1$ is redundant, and the claim follows from Theorem 3(iii). \(\square\)

Theorem 3 differs from the results in Juang (2002) in that the selection of equilibria does not depend on the shares of imitators and better repliers, as long as they are positive in all populations. This is mainly due to the fact that a single player is sufficient to reach the basin of attraction of any convention. Second, if the process is in state $h_{(B,B)}$ and $(A,A)$ is strictly Pareto dominant, then two simultaneous mistakes, followed by a sequence of imitators in one of the populations, are sufficient to reach the basin of attraction of $h_{(A,A)}$. The condition that $s > \max\{\hat{s}_i(x), \hat{s}_f(x), 2/q_{\min}\}$ implies that the reverse transition will require more than two mistakes.

The following corollary shows that in symmetric coordination games, there always exists a range of sample sizes such that the risk-dominant equilibrium is selected.

6. Conclusions

In this paper, we analyze stochastic adaptation in finite $n$-player games played by heterogeneous populations containing better repliers. We show that for a large class of learning rules, if a least one better replier in each population has sufficiently low ratio between the sample and memory size, and independently of the population shares, the recurrent classes of the resulting unperturbed Markov chain correspond one to one with the MCUBR sets of the game. Such sets are also asymptotically stable under a large class of deterministic continuous-time selection dynamics, containing the replicator dynamics. Moreover, in large classes of games, they correspond to strict Nash equilibria.

We also demonstrate that when populations consist of imitators and better repliers and one MCUBR set is Pareto efficient, the set of stochastically stable states spans this set. This result requires a sufficiently large sample size, determined by the payoffs of the game.

Finally, we show that in all symmetric $2 \times 2$ coordination games, and many asymmetric such games, although the Pareto efficient equilibrium is selected for a sufficiently large sample size, the risk-dominant equilibrium is selected for a range of sample sizes.

The main contribution of this paper is to analyze the long-run outcome when multiple rules of adaptation, previously analyzed only in isolation, are present in the populations. Most of our results do not hinge on the population fractions using different learning rules, as long as the shares of better repliers and imitators are positive. This is in contrast to a model where all individuals are matched in each period, such as Juang (2002). However, generally the exact shape of the limiting distribution will depend on these fractions, and so will the expected payoffs to the individuals employing the various rules. A next step would be to allow the population shares to vary over time as a function of realized payoffs and possibly to allow for different information cost of different learning rules along the lines of Conlisk (1980), Brock and Hommes (1997), and Droste et al. (2002).
Acknowledgments

I thank the editor Cars H. Hommes, three anonymous referees, Alexander Matros, Marı́a Sáez Martí, and Jörgen W. Weibull as well as seminar audiences at Universitat Pompeu Fabra and Universidad de Alicante for helpful comments. Financial support from Bankforskningstitet, Handelsbankens forskningsstifelter, the Spanish Ministry of Education and Science through grants SEJ2004-03149 and SEJ2006-09993, the Barcelona GSE Research Network, and the Government of Catalonia is gratefully acknowledged.

Appendix A

Let $P^0$ be a stationary Markov process on a finite state space. Following Young (1998), we call $P^0$ a regular perturbed Markov process if (i) $P^0$ is irreducible for every $\epsilon$ in some interval $(0, \epsilon^*)$, (ii) for every $h, h' \in H$, $P^0_{hh'}$ approaches $P^0_{hh}$ at an exponential rate, i.e. $\lim_{\epsilon \to 0} P^0_{hh'} = P^0_{hh}$, and if $P^0_{hh} > 0$ for some $\epsilon > 0$, then $0 < \lim_{\epsilon \to 0} P^0_{hh} / \epsilon^\omega < \infty$ for some $r_{hh}^0 > 0$.

**Lemma 1.** $P^{m,\epsilon}$ is a regular perturbed Markov process.

**Proof.** The proof of Lemma 1 is completely analogous to the proof in Young (1998, p. 55). □

**Lemma 2.** (Young, 1998, p. 153). Let $P^e$ be a regular perturbed Markov process and let $\mu^e$ be the unique stationary distribution of $P^e$ for each $\epsilon > 0$. Then $\lim_{\epsilon \to 0} \mu^e = \mu^0$ exists, and $\mu^0$ is a stationary distribution of $P^0$. The stochastically stable states are precisely those states that are contained in the recurrent classes of $P^0$ having minimum stochastic potential.

**Proof of Theorem 1.** Part (i) of Theorem 1 will be proved in two steps. In step (A), we will prove that from any initial state, $P^m$ converges to a set of states that span a MCUBR set. In step (B), we will prove that for any MCUBR set, there exists a recurrent class that consists of states that span the set. In what follows, we will use $S(E)$ to denote the set spanned by the set of states $E$.

(A) Let the stochastic better-reply dynamics be the induced Markov chain when the share of better repliers with $s/m < 1/|X|$ in each population is one. According to Theorem 1 in Josephson (2008a), for a sufficiently large sample size the span of each recurrent class of the stochastic better-reply dynamics corresponds one to one with a MCUBR set. Since, in each period, there is a positive probability that only better repliers will be selected to play, there is also a positive probability that heterogeneous adaptive play, from any initial state, and in a finite number of periods, will end up in a state $h$ that belongs to a recurrent class $E^0_h$ of the stochastic better-reply dynamics, and thus only involves play of pure-strategy profiles in a corresponding MCUBR set $Y$. Since all learning rules are returning, it follows that the process will never play a pure strategy outside $Y$, once it has reached such a state. If $E^0_h$ is singleton, this means that it is a recurrent class also under heterogeneous adaptive play $P^m$. If $E^0_h$ is not singleton, there is a positive probability that the process makes the transition from $h$ to any other state $h'$ of $E^0_h$ in a finite number of periods and, naturally, without playing any pure strategy outside $Y$. This follows since in each period, there is a positive probability that only better repliers will be selected to play. Hence, there exists a recurrent class $E^m_Y$ of $P^m$, such that $E^m_Y \subseteq E^0_Y$ and $S(E^0_Y) \subseteq Y = S(E^m_Y)$.

(B) Conversely, we will prove that for any MCUBR set, heterogeneous adaptive play has a recurrent set that spans this set. If $Y$ is a MCUBR set, then, by Theorem 1 in Josephson (2008a), there exists a unique recurrent class $E^m_Y$ of the stochastic better-reply dynamics such that $S(E^m_Y) = Y$. If the Markov chain $P^m$ is in a state involving no pure strategy outside $Y$, there is a positive probability that it will reach the state which belongs to $E^0_Y$ in a finite number of periods. This follows since $P^m$, by (A), from any initial state and in a finite number of periods will end up in a state $h$, which belongs to a recurrent class of the better-reply dynamic, and since, because all learning rules are returning, the process will never play a pure strategy outside $Y$, once it has reached a state only involving pure-strategy profiles of $Y$. It thereafter follows from (A) that there exists a recurrent class $E^m_Y$ of $P^m$, such that $S(E^0_Y) = Y$.

Part (ii) of Theorem 1 follows directly from Lemma 2 in the Appendix since $P^{m,\epsilon}$, by Lemma 1 in the Appendix, is a regular perturbed Markov process. □

**Proof of Theorem 2.** Theorem 2 will be proved in three steps. In step (A), we show that the transition from any recurrent class to a set of states that span a Pareto-dominant MCUBR set can always be made with at most $n$ mistakes. In step (B), we prove that for a sufficiently large sample size, the transition from a set of states that span a Pareto-dominant MCUBR set to any other recurrent class requires at least $n + 1$ mistakes. In step (C), we use (A) and (B) to prove that the recurrent class that spans the Pareto-dominant MCUBR set must have minimum stochastic potential.

(A) Assume that $s/m < 1/|X|$ for at least one better replier in each population so that, by Theorem 1, the span of each recurrent class of $P^m$ corresponds one to one with a MCUBR set. Let $Y \subseteq X$ be a strictly Pareto-dominant MCUBR set, and let $E^m_Y$ be the corresponding recurrent class. Assume there exists at least one other recurrent class (otherwise Theorem 2 holds trivially). The transition from such a recurrent class to $E^m_Y$ can always be made with a probability of the order $\epsilon^N$ (or higher). This is, for instance, the case if the individuals in all roles experiment and play a pure-strategy profile $y^t \in Y$ in period $t$, and a sequence of $m - 1$ imitators thereafter are drawn to play in all roles, and all sample $y^t$. 
(B) Let $E_2$ be an arbitrary recurrent class, different from $E_Y$. We claim that if all individuals have sufficiently large samples, the probability of the transition from $E_Y$ to $E_2$ is at least of the order $\epsilon^{N+1}$. To make a better replier with sample size $s$ in role $i$ switch to a pure strategy $x_i \neq Y_i$ after at most $n$ mistakes, the expected payoff to that pure strategy must be greater than for any pure strategy $y_i \in Y_i$,

$$
\frac{s-n}{s} u_i(y_i, p^{Y_i}) + \frac{n}{s} u_i(y_i, p^{y_i}) \leq \frac{s-n}{s} u_i(x_i, p^{Y_i}) + \frac{n}{s} u_i(x_i, p^{x_i}),
$$

which $p^{Y_i} \in \square(Y_{-i})$, and $p^{x_i} \in \square(x_{-i})$. By the boundedness of payoffs and the strict Pareto dominance of $Y$, the right-hand side of this inequality is clearly bounded for any $i$, $y_i \in Y_i$, $x_i \neq Y_i$, $p^{Y_i} \in \square(Y_{-i})$, and $p^{x_i} \in \square(x_{-i})$. Hence, there exists some finite $\hat{s}$, such that for $s > \hat{s}$, strictly more than $n$ mistakes are necessary for a better replier to play a pure strategy outside $Y$.

Similarly, in order to make an imitator with sample size $s$ maximizing the average realized payoff switch to a pure strategy $x_i \neq Y_i$ after at most $n$ mistakes,

$$
\frac{(s-n)u_i(y_i, p^{Y_i}) + (n-1)u_i(y_i, p^{y_i})}{s-1} \leq \pi_i(x_i, x_{-i}),
$$

where $p^{Y_i} \in \square(Y_{-i})$, and $p^{x_i} \in \square(x_{-i})$. By the boundedness of payoffs and the strict Pareto dominance of $Y$, the right-hand side of the last inequality is bounded for any $i$, $y_i \in Y_i$, $x_i \neq Y_i$, $p^{Y_i} \in \square(Y_{-i})$, and $p^{x_i} \in \square(x_{-i})$. Hence, there exists some finite $\hat{s}$, such that if $s > \hat{s}$, strictly more than $n$ mistakes are necessary for an imitator to play a pure strategy outside $Y$.\footnote{An imitator who simply picks a pure strategy with the maximum realized payoff can, of course, never switch to a pure strategy $x_i \neq Y_i$ with less than $s$ mistakes.} Thus, for if all individuals have sample size larger than $\max(\hat{s}, \hat{s}, N)$, the resistance of the transition from $E_Y$ to $E_2$ must be greater than $N$.

(C) Consider the minimum resistance tree rooted at an arbitrary recurrent class $E_D$, different from $E_Y$. In this tree, there must be a directed edge from $E_Y$ to some other recurrent class $E_Z$ (possibly identical to $E_D$). Assume that the sample size is so large that the resistance of the transition from $E_Y$ to $E_Z$ is greater than $N$ (this is possible by (B)), and that the stochastic potential of $E_D$ is smaller than or equal to that of $E_Y$. Create a new tree by deleting the edge from $E_Y$ to $E_Z$, and adding an edge from $E_D$ to $E_Y$. The resistance of the deleted edge is, by assumption, greater than $N$, and the resistance of the added edge is (by (A)) smaller than or equal to $N$. Hence, the total resistance of the new tree is less than that of the tree rooted at $E_D$, contradicting the assumption that the stochastic potential of $E_D$ is smaller than or equal to that of $E_Y$. This proves that the stochastic potential of the Pareto-dominant recurrent class $E_Y$ is lower than for any other recurrent class and, by Lemmas 1 and 2 in the Appendix, Theorem 2 follows.

Proof of Theorem 3. (i) Convergence with probability one to a convention. By replacing best-replies by better-replies in the proof of Theorem 4.2 in Young (1998, pp. 68–70), it follows that if the share of better repliers is one in all populations and $s/m < \frac{1}{2}$, then the recurrent classes of the unperturbed process are the two monomorphic states $h_{A,A}$ and $h_{B,B}$. It is clear that these states are absorbing also when the shares of imitators and better repliers are positive. Moreover, since in each period, there is a positive probability that only better repliers are drawn to play, there is a positive probability of reaching one of these two states in a finite number of periods from any other state. Hence, for $s/m \leq \frac{1}{2}$ the recurrent classes of $P^m$ are two states $h_{A,A}$ and $h_{B,B}$.

(ii) Selection of strictly Pareto-dominant equilibrium. Without loss of generality, assume that $(A,a)$ is a strictly Pareto-dominant equilibrium. Then, the transition from $h_{A,B}$ to the basin of attraction of $h_{A,A}$ can always be made with two simultaneous mistakes in period $t$ followed by a sequence of $s-1$ imitators in both populations, who all sample $x_{-i}$ or a more recent strategy-uptake. The reverse transition requires at least three mistakes if the following two conditions are fulfilled. First, in state $h_{A,A}$, better repliers in one of the populations should not be able to switch to strategy $B$ when the sample contains less than three mistakes by the other population. This is prevented if $s > \frac{2}{q_{\min}}$ or equivalently if $s > \frac{2}{q_{\min}}$. Second, imitators in the role of the row player should not be able to switch to strategy $B$ with less than three mistakes. They can only do this if two mistakes, one after the other by different populations, make the average payoff to strategy $B$ at least as large as that of $A$, or formally, if

$$
\frac{(s-2)\pi_1(A,a) + \pi_2(A,b)}{s-1} \leq \pi_1(B,a),
$$

\label{eq:inequality}

(ii) Selection of strictly Pareto-dominant equilibrium. Without loss of generality, assume that $(A,a)$ is a strictly Pareto-dominant equilibrium. Then, the transition from $h_{A,A}$ to the basin of attraction of $h_{A,A}$ can always be made with two simultaneous mistakes in period $t$ followed by a sequence of $s-1$ imitators in both populations, who all sample $x_{-i}$ or a more recent strategy-uptake. The reverse transition requires at least three mistakes if the following two conditions are fulfilled. First, in state $h_{A,A}$, better repliers in one of the populations should not be able to switch to strategy $B$ when the sample contains less than three mistakes by the other population. This is prevented if $s > \frac{2}{q_{\min}}$ or equivalently if $s > \frac{2}{q_{\min}}$. Second, imitators in the role of the row player should not be able to switch to strategy $B$ with less than three mistakes. They can only do this if two mistakes, one after the other by different populations, make the average payoff to strategy $B$ at least as large as that of $A$, or formally, if

$$
\frac{(s-2)\pi_1(A,a) + \pi_2(A,b)}{s-1} \leq \pi_1(B,a),
$$

\label{eq:inequality}

(ii) Selection of strictly Pareto-dominant equilibrium. Without loss of generality, assume that $(A,a)$ is a strictly Pareto-dominant equilibrium. Then, the transition from $h_{A,A}$ to the basin of attraction of $h_{A,A}$ can always be made with two simultaneous mistakes in period $t$ followed by a sequence of $s-1$ imitators in both populations, who all sample $x_{-i}$ or a more recent strategy-uptake. The reverse transition requires at least three mistakes if the following two conditions are fulfilled. First, in state $h_{A,A}$, better repliers in one of the populations should not be able to switch to strategy $B$ when the sample contains less than three mistakes by the other population. This is prevented if $s > \frac{2}{q_{\min}}$ or equivalently if $s > \frac{2}{q_{\min}}$. Second, imitators in the role of the row player should not be able to switch to strategy $B$ with less than three mistakes. They can only do this if two mistakes, one after the other by different populations, make the average payoff to strategy $B$ at least as large as that of $A$, or formally, if
It is clear that for $s > \hat{s}_1(A, a)$ this inequality does not hold, and a similar critical sample size $\hat{s}_2(A, a)$ can be computed for individuals in the role of the column player. Hence, for $s > \max\{\hat{s}_1(A, a), \hat{s}_2(A, a), 2/q_{\min}\}$, the transition to $(B, b)$ requires at least three mistakes, whereas the reverse transition requires exactly two mistakes (if all imitators choose the pure strategy with maximum sample payoff, i.e. nobody chooses the pure strategy with the maximum average payoff in the sample, then it is sufficient that $s > 2/q_{\min}$).

(iii) Unique risk-dominant equilibrium. The transition from $h_{(B, b)}$ to the basin of attraction of $h_{(A, a)}$ can be made with $k$ mistakes if individuals in one of the populations, say population $C_1$, by mistake plays $A$ $k$ times in a row, and $k/s \geq q_A$. This follows since there is a positive probability that better repliers are drawn to play in the other population for the next $s$ periods, and that these individuals all sample the maximum the $k$ mistakes. Without loss of generality, assume that $(A, a)$ is a risk-dominant equilibrium. Then, the transition from $h_{(B, b)}$ to $h_{(A, a)}$ requires only one mistake if $q_{\min} \leq 1/s$. The requirement that $s_{\text{diff}}(s) \geq 1$ ensures that the reverse transition requires at least two mistakes.

(iv) Both conventions stochastically stable. Consider a general $2 \times 2$ coordination game. First note that if $s_{\text{diff}}(s) = 0$ and $s_{\text{min}} \leq 1$, then the transition from any convention to the basin of attraction of the other can be done with just one mistake in the same fashion as in the proof of (iii). If, on the other hand, $s_{\text{min}} > 1$, then at least two mistakes are necessary to make this kind of transition. If the game has a strictly Pareto-dominant equilibrium, but $s \leq \max\{\hat{s}_1, \hat{s}_2, 2/q_{\min}\}$, then the basin of attraction of the Pareto dominated convention can be reached from the other convention either by two mistakes in one of the populations followed by better repliers sampling these mistakes in the other population, or by one mistake in one population and another mistake in the other followed by imitators sampling these mistakes (as described in the proof of (ii)). If the game does not have a strictly Pareto-dominant equilibrium, then the basin of attraction of each convention can be reached from the other with two simultaneous mistakes followed by a sequence of imitators, in the position of the player who is weakly better off in the other equilibrium, picking a sample including the play with these mistakes.

Hence, in all cases not covered by (ii) and (iii), the stochastic potential is identical for the two conventions.

Theorem 3 now follows by Lemma 1 and Lemma 2. □

Appendix B

In this section, we will demonstrate why the ratio $s/m$ needs to be bounded for at least one better replier in each population for Theorem 1 to hold. For simplicity we may assume that all populations consist of better repliers with $s = 2$ and $m = 3$ (similar arguments hold for any $m \geq 1$ and $s \in \{1, \ldots, m\}$, such that the ratio $s/m$ is sufficiently large).

Consider the three-player game in Fig. 4, where player 1 chooses row, player 2 column, and player 3 matrix.

This game has a unique MCUBR set consisting of the entire strategy space. However, if the process starts in the state $h_{(A, a, a)}$, then strategy $\beta$ will never be played. To see this, note that for an individual in player position 3 to start playing $\beta$, he needs to draw a sample containing only strategy $C$ for player position 1 and only strategy $c$ for player position 2. However, a state containing $s$ instances of $C$ and $s$ instances of $c$ cannot be reached from the state $h_{(A, a, a)}$. The reason is that if a state contains $s - 1$ or more instances of $C$ (c) and less than $s$ instances of $C$ (c), then there is no sample such that strategy $C$ (c) is a better reply for player 2 (1).

References