Reinforcement Learning in Evolutionary Games*

Ratul Lahkar† and Robert M. Seymour‡

October 5, 2010

Abstract

We study an evolutionary model in which strategy revision protocols are based on agent specific characteristics rather than wider social characteristics. We assume that agents are primed to play a mixed strategy, with the weights on each pure strategy modifiable on the basis of experience. At any time, the distribution of mixed strategies over agents in a large population is described by a probability measure on the space of mixed strategies. In each round, a pair of randomly chosen agents play a symmetric game, after which they update their mixed strategies using a reinforcement learning rule based on payoff information. The resulting change in the distribution over mixed strategies is described by a non-linear continuity equation — in its simplest form a first order partial differential equation associated with the classical replicator dynamics. We provide a general solution to this equation in terms of solutions to an associated finite-dimensional dynamical system. We use these results to study in detail the evolution of mixed strategies in various classes of symmetric games, and in a simple model of price dispersion. A key finding is that, when agents carry mixed strategies, distributional considerations cannot be subsumed under a classical approach such as the deterministic replicator dynamics.

Keywords: Reinforcement Learning; Continuity Equation; Replicator Dynamics.
JEL classification: C72; C73.

*The support of the UK's Economic and Social Research Council (ESRC) for both authors is gratefully acknowledged. This work was part of the programme of the ESRC Research Centre for Economic Learning and Social Evolution (ELSE) at UCL. We thank Bill Sandholm, for comments and suggestions.

†Institute for Financial Management and Research, 24, Kothari Road, Nungambakkam, Chennai, 600 034, India. email: r.lahkar@ifmr.ac.in.

‡Department of Mathematics and ELSE centre, University College London, Gower Street, London, WC1E 6BT, UK. email: rms@math.ucl.ac.uk.

1 Introduction

It is well accepted that individuals arrive at most economic decisions through a learning process that is based on their personal experience. This is well recognized in the extensive literature on learning by individual players in finite player games (see Young, 2007 for a review of this literature). Perhaps the simplest experienced-based learning protocol is *reinforcement learning*, a procedure under which each of several possible actions open to an agent is rewarded (or penalized) according to its performance against some learning criterion. Reinforcement protocols have considerable empirical support (e.g. Erev and Roth, 1998). Under reinforcement learning, an individual carries an internal mixed strategy, construed as the agent's *behavioral disposition*, and it is this that is modified by reinforcement through adjustment of the weights on each pure strategy. Such learning has been extensively investigated, beginning with Cross (1973), who proposed a simple rule that reinforcement should be proportional to a player's payoff. This suggestion was later taken up and elaborated by Börgers and Sarin (1997, 2000), Hopkins (1998) and Börgers et al (2004), who showed that there is a close relationship between this type of learning rule and the classical replicator dynamics of evolutionary game theory (Taylor and Jonker, 1978).

It is therefore reasonable to believe that individuals will also use experience based learning to guide their behavior when they play as members of large populations. Yet, in the literature on evolutionary game theory, which studies the evolution of social behavior in large populations, there is scant recognition of this feature of human learning. Instead, agents are assumed to arrive at decisions on which pure strategy to play on the basis of certain observations about some social characteristics that are external to themselves. For example, in the imitative revision protocols that generate the replicator dynamic, an individual imitates the strategy of a more successful rival. In the best response protocol, an agent observes the current social state (the distribution in the population of different pure strategy players) and plays the strategy yielding the best expected payoff. We therefore believe that it is meaningful to ask what shape the evolution of social behaviour takes when individual agents are allowed to interact on the basis of experience based learning. This paper develops a formal approach to analyze this question.

The conditioning of revision protocols in evolutionary game theory on observations of external social characteristics is a legacy of the origins of the field in biology. The evolutionary process in biology is modeled as an automatic process, driven by births and deaths, working to increase the frequency of better performing strategies. In economics, this has been directly translated into imitative revision protocols yielding the replicator dynamic (Börnerstedt and Weibull, 1996; Schlag, 1998). These protocols have the advantage that they require very little information for their implementation; calling for nothing more than the ability to observe the strategy of a randomly chosen member of the population. However, the range of situations of an economic or social interest over we may expect naive imitative behavior to apply is likely to be limited. In order to allow for more sophisticated behavior, evolutionary game theory has also incorporated more elaborate revision procedures like best response or its smoothened version, the perturbed best response.[1] However, these revision protocols have the disadvantage of making extremely onerous informational demands: an agent needs to observe the entire social state before calculating the payoffs of each strategy as an expectation over the social state. It is unlikely that in a decentralized environment agents would be privy to such detailed information about the social state, thereby rendering the implementation of such revision protocols essentially unfeasible.

In contrast, the evolutionary framework that we construct is free from such onerous informational requirements, being based on strategy revision protocols that require only knowledge of *agent-specific* characteristics rather than of wider social characteristics. This allows a greater range of behaviors—whether resulting from conscious deliberation or from essentially subconsciously processes—to be feasibly implemented, thereby doing justice to a greater range of the cognitive abilities of human agents. In particular, we use revision protocols based on reinforcement learning, since, as noted earlier, these are both simple and extremely parsimonious in the information required for their implementation, as well as having empirical support, as documented by Erev and Roth (1998). However, our general theoretical framework can be adapted to other learning mechanisms like regret matching (Hart and Mas-Colell, 2000) or stochastic fictitious play (Fudenberg and Levine, 1998). The application of rules from the learning literature to the study of evolution of social behavior also formally links together the work in these two fields.

Within reinforcement learning, we focus on a specific rule—the Cross (1973) learning rule, as developed by Börgers and Sarin (1997). Under this rule, the decision maker increases the probability of the action he chose in the last round in proportion to the payoff received, while reducing the probability of the other actions proportionately. Our focus on the Cross learning rule allows us to adapt the axiomatic learning framework developed in Borgers et. al. (2004) into the analysis of large population models. In this approach, learning rules are evaluated not the basis of any heuristic plausibility but according to their confirmity to certain fundamental principles; namely *absolute expediency* and *monotonicity*. The former condition requires that the expected payoff obtained from a learning rule strictly increase over time whereas the latter demands that the probability assigned to the best actions increase over time. This axiomatic approach clearly has the advantage of establishing rigorous benchmarks for the admissibility of a strategy revision rule. Traditionally, evolutionary game theory, instead of axiomatizing individual behavior, has adopted certain group characteristics such as *monotone percentage growth rate* or *Nash stationarity* as the desiderata that a dynamic model should satisfy.[2] However, in a model of atomistic individuals guided by self interest, we believe that the foundational principles used to validate a model should apply at the level of individual behavior. The Cross rule, as shown in Börgers et. al. (2004), is the prototype of the class of rules that satisfy their two fundamental axioms. Our approach therefore allows us

[1] These revision protocols generate respectively the best response dynamic (Gilboa and Matsui, 1991) and the perturbed best response dynamic (Hofbauer and Sandholm, 2005). The prototypical perturbed best response dynamic is the logit dynamic (Fudenberg and Levine, 1998). See Weibull (1995), Hofbauer and Sigmund (1988, 1998), Samuelson (1997), and Sandholm, 2009, for book level reviews of evolutionary game theory.

[2] See Sandholm (2009) for a review of these conditions. Monotone percentage growth rate, or its weaker counterpart, positive correlation ensures a positive relationship between the payoff of a strategy and the growth rate of its population share while Nash stationarity ensures that the population is at rest at a Nash equilibrium.

to provide more rigorous foundations to evolutionary game theory.

The analysis in Börgers and Sarin (1997) also shows that under the Cross learning rule, the expected change in the mixed strategy of a player is given by the replicator dynamic. Since they focus on learning in finite player models, this finding does not have any immediate implication concerning the evolutionary consequences of Cross-like rules in large population models. Nevertheless, this does suggest the possibility that the application of Cross' rule in large population models should lead to an evolutionary process that is associated to the replicator dynamic. We explore this question in this paper and find that this is indeed so: under Cross' rule (as deployed by Börgers and Sarin, 1997), the evolution of the proportion of players playing a particular pure action is given by an adjusted form of the replicator dynamic.

We begin by formalizing the application of a general experience based learning rule to a large population model. The first key point is that, to apply reinforcement type learning to population games, agents must be assumed to use individual mixed strategies. This contrasts sharply with almost all developments of evolutionary game models, which assume that agents are primed to play pure strategies, and that it is the proportions of players that play different pure strategies that changes over time[3]. With agents playing mixed strategies, the *population state* is now specified by a probability measure over the space of mixed strategies (a simplex), and it is this population distribution that changes over time in response to agents' learning. The main challenge is to develop a technique to analyze the evolution of this population distribution. We consider a situation in which players from the same population are randomly matched to play a two player symmetric game. Matchings last for one period and in each new matching, players revise their mixed strategies using some general experience based strategy revision rule such as reinforcement learning. We note that this introduces a radical form of agent heterogeneity into the population, extending the classical setting in which all agents use a fixed mixed strategy or, equivalently, a fixed mixture of pure strategies. As players revise their strategies, the population state changes in a specifiable way that depends on the form of the learning rule. By making the time difference between successive matches go to zero, we are able to track the change in the population state by using a generalization to a probability measure setting of a first-order partial differential equation system akin to the *continuity equations* commonly encountered in physics in the study of conserved quantities, such as bulk fluids.[4] Under plausible assumptions, we construe this continuous-time limit of the discrete-time

[3] Hines (1980) and Zeeman (1981) are two examples from the early biology literature on the evolution of mixed strategies. These papers use straightforward adaptations of different versions of the replicator dynamic (see footnote 13) to study the evolution of mixed strategies using the standard biological motivation that the growth of the population share using a strategy is proportional to the advantage of that strategy over the mean strategy. This approach cannot be directly translated into the study of human interaction since the analogous motivation of imitating successful strategies is inapplicable due to the non-observability of mixed strategies. Our reinforcement based approach, of course, does not suffer from this drawback.

[4] In physics, the continuity equation is a linear partial differential equation that describes the rate of change in the mass of fluid in any part of the medium through which it is flowing. See, for example, Margenau and Murphy (1962). However, our continuity equations concern the change in probability mass of agents in any part of mixed-strategy space, and differ from classical versions encountered in physics in that they contain non-linearities. See Ramsza and Seymour (2009) for an application of continuity equation techniques to track the evolution of fictitious play updating weights in a population game. Our paper provides a more general method of constructing continuity equations that

matching routine as equivalent to the limit as the population size becomes infinite.

From the general continuity equation, we generate one particular form—the *replicator continuity equation*—using the Börgers and Sarin (1997) version of the Cross learning rule. The name we have chosen for this dynamic reflects the connection between this rule and the classical replicator dynamic. The continuity equation can also be applied to follow the evolution of the mean of the population state which is the *aggregate social state*: the proportion of agents playing different pure actions. We show that the mean dynamic corresponding to the replicator continuity equation is the classical replicator dynamic adjusted for a covariance term. This confirms the intuition obtained from Börgers and Sarin (1997) that the application of the Cross learning rule to a population would generate an evolutionary process closely related to the classical replicator dynamic. We reiterate, however, that the microfoundations of the continuity dynamic is very different from the imitation type protocols usually invoked to generate the classical replicator dynamic. Further, by applying learning algorithms explicitly to large population models, our work provides a more general perspective on the link between learning and evolution.

We solve the replicator continuity equation using standard methods based on Liouville's formula.[5] To characterize solutions explicitly requires us to derive an associated ODE system whose solutions describe trajectories of certain aggregate quantities associated to the population means. We call this ODE system the *distributional replicator dynamics*. We show that the continuity replicator dynamics has many stationary solutions, in particular any probability distribution over mixed strategies having mean that is a Nash equilibrium. Thus, equilibrium populations can be very heterogeneous, with individuals playing any mixed strategy with positive probability, but with population mean a Nash equilibrium.

We then apply the replicator continuity equation to the analysis of evolution in three classes of symmetric games that have been extensively studied in evolutionary game theory literature: negative definite and semi-definite games, positive definite games and doubly symmetric games. We focus on the evolution of the mean of the population distribution. Even though agents are playing mixed strategies, the observable social state is the distribution across pure actions. By the law of large numbers, this is identical to the mean social state or the aggregate social state. In addition, focusing on the aggregate state provides a natural way to compare our results with the the results from conventional evolutionary game theory.

Under weak assumptions on the initial population distribution we show that, for negative-definite games the unique Nash equilibrium is globally asymptotically stable. In contrast, for positive definite games we show that any interior mixed Nash equilibrium is unstable. For doubly symmetric games, we show that the mean state converges to some Nash equilibrium. These results are of course consistent with the larger set of stability results under conventional evolutionary

can be used for a variety of learning algorithms.

[5]This formula expresses the time evolution of the probability density function as a function of the initial probability density and the deterministic trajectories of the underlying characteristic ODE system, which describes the motion of individual agents in the population– see, section 4 below. The classical Liouville formula describes the change in volume along flow lines of an underlying dynamical system see, for example, Hartman (1964). Related versions are discussed in Weibull (1995) and Hofbauer and Sigmund (1998).

dynamics in these classes of games (for a review, see Sandholm 2009). Nevertheless, these results are important since they hold even under the different evolutionary paradigm that we have developed.

However, even when the aggregate social state converges to a Nash equilibrium of the game under the continuity replicator dynamics, it is not necessarily the case that the associated limiting population probability distribution is a mass point at a Nash equilibrium. Equilibrium population distributions over mixed strategies are often complex objects, defining highly heterogeneous behaviors within an equilibrium population, though the mean behavior is a Nash equilibrium. We go on to characterize such limiting distributions.

We then discuss some simple examples. First, we characterize convergence of the mean population state for generic 2×2 symmetric games. We find that the limiting aggregate social state is the same Nash equilibrium that would result under the classical replicator dynamic, provided the initial point in the latter case is identical to the initial aggregate state in the former case. However, this conclusion does not hold for $n \times n$ symmetric games with $n > 2$, and we provide a counter-example for $n = 3$ in which the two dynamics converge to different pure equilibria. Hence, expanding the behavioral flexibility of agents to allow use of mixed strategies in evolutionary contexts has real consequences, in that it can lead to radically different conclusions about the equilibrium social state.

Finally, we provide an illustrative economic application of our model using a simplified case of the Burdett and Judd (1983) model of price dispersion. We show that mixed equilibria (the dispersed price equilibria) of this simplified model are unstable under the replicator continuity equation. Results of a similar nature have been obtained in Hopkins and Seymour (2002) and Lahkar (2010) under the replicator dynamic and logit dynamic, respectively. However, in our model, players reinforce any particular price based on whether that price led to a sale or not in the previous round of the game. In light of our comments about the greater plausibility of experience based learning rather than observation based learning, we believe that this is a more realistic way to model seller behavior.

The remainder of this paper is organized as follows. In section 2 we set out the formal evolutionary model based on an abstract form of reinforcement learning for 2 player symmetric games, and show how the continuous-time limit of this general model leads to a continuity equation. Section 3 presents the Cross (1973) learning rule and shows that this rule generates the replicator continuity equation. We also characterize the rest points under this dynamic. In section 4, we introduce Liouville's formula in a general context, and use this formula to solve a generalized form of the replicator continuity equation in section 5. In section 6, we introduce the distributional replicator dynamics, a system of autonomous ODEs and show how their solutions determine the distributional solution of the replicator continuity equation. Section 7 contains applications to the three classes of games mentioned above, and section 8 characterizes limiting population distributions for dynamics whose mean converges to a Nash equilibrium. Section 9 considers the simple 2×2 and 3×3 examples, and section 10 discusses the price dispersion example. Section 11 contains a concluding discussion. Certain proofs and additional technical material are presented in the appendix.

2 The General Continuity Equation for Population Games

We derive the continuity equation in the setting of population games. We consider the case in which two players, chosen from the same population, are randomly matched to play a symmetric normal form game.[6] The game has set of pure actions $\mathbf{n} = \{1, 2, \cdots, n\}$. Each agent in the population carries a mixed strategy which they use to determine their play when called upon to do so. The state space for individual agents is therefore the simplex $\Delta = \Delta[\mathbf{n}] \subset I\!R^n$ whose elements are the possible mixed strategies:

$$\Delta = \left\{ x \in I\!R^n : x_i \geq 0 \text{ for each } i, \text{ with } \sum_{i=1}^{n} x_i = 1 \right\}. \tag{1}$$

We assume that the population state (at a given time) is characterized by a Borel probability measure P defined on the state space Δ of mixed strategies[7] Thus, if $B \subseteq \Delta$ is a Borel set, then $P(B)$ can be regarded as the proportion of agents in the population playing mixed strategies in B. The population mean strategy, denoted $\langle P \rangle \in \Delta$, is given by

$$\langle P \rangle = \int_{\Delta} xP\,(dx)\,. \tag{2}$$

We interpret the mean $\langle P \rangle$ as the *aggregate social state* generated by the measure P. Even though agents are actually playing mixed strategies, the observable aggregate social state is the proportion of agents playing different pure actions. By the law of large numbers, this distribution over pure actions is equal to $\langle P \rangle$. We make use of this concept of the aggregate social state in our later applications of the continuity equation.

More generally, given a real, vector-valued continuous function $\phi(x)$ on Δ, we define its expectation with respect to P by

$$\langle \phi \mid P \rangle = \int_{\Delta} \phi(x) P(dx). \tag{3}$$

In particular, $\langle P \rangle = \langle \iota \mid P \rangle$, where ι is the identity map on Δ.

2.1 Matching and updating

In each time interval of length τ, two agents from the population are randomly matched to play the symmetric normal form game. We denote by P_t the probability measure characterizing the population state at time $t \geq 0$. Our objective is to track the evolution of the measures P_t over time.

Suppose the two chosen players use the mixed strategy pair $(x, y) \in \Delta \times \Delta$. The probability

[6] We confine ourselves to two-player symmetric games merely for notational convenience. All the ideas involved can be extended easily to multi-player symmetric as well as asymmetric games at the cost of more cumbersome notation.

[7] That is, P is a non-negative measure of total mass 1, defined on the σ-field of Borel sets in Δ, the smallest σ-field containing the closed sets of Δ – see, for example, Dunford and Schwatz (1964), p 137.

that they play the action profile $(i, j) \in \mathbf{n} \times \mathbf{n}$ is given by

$$\pi_{ij}(x, y) = x_i y_j. \tag{4}$$

Of course, $\sum_{i,j} \pi_{ij}(x, y) = 1$ for all (x, y). After a play of the game, and during the time interval τ, a player updates his mixed strategy according to some revision protocol of the following general form. Given that the action profile (i, j) has been played, we assume that the row player updates his strategy $x \in \Delta$ to x' given by an updating rule of the form:

$$x' = x + \eta f_{ij}(x), \tag{5}$$

where $f_{ij}(x)$ defines the potential revision that can be realized by this play. The parameter $\eta = \eta(\tau)$ is the proportion of this potential change that is realized in the revision exposure time τ. Alternatively, $\eta(\tau)$ can be regarded as the probability that the change $f_{ij}(x)$ is delivered in the given time interval. We assume that $\eta(\tau)$ satisfies: (i) $\eta(\tau)$ is increasing in τ with $\eta(\tau) \leq 1$ for all $\tau > 0$; (ii) $\eta(0) = 0$; (iii) $\eta'(0) > 0$. Condition (i) says that more of the potential revision is realized the longer the players are exposed to the revision stimulus consequent on the play of the game. Condition (ii) says that no revision takes place if there is no exposure to the consequences of play, and condition (iii) says that any positive exposure time delivers some of the potential revision Without loss of generality we can (and will) assume that $\eta'(0) = 1$. To ensure that (5) defines a mixed strategy, we require that $\sum_{r=1}^{n} f_{ij,r}(x) = 0$ for each $x \in \Delta$. We assume also that $x_r + f_{ij,r}(x) \geq 0$ for each $1 \leq r \leq n$. Then $x_r + \eta f_{ij,r}(x) \geq \eta\left(x_r + f_{ij,r}(x)\right) \geq 0$ for all $0 \leq \eta \leq 1$.

The reinforcement potentials $f_{ij}(x)$ define a function $f_{ij} : \Delta \to \mathbb{R}_0^n$, where

$$\mathbb{R}_0^n = \{z \in \mathbb{R}^n : \sum_{r=1}^{n} z_r = 0\}. \tag{6}$$

We call this the *forward state change* function: it specifies how the players' states change going forward in time. The associated *backward state change* function specifies where current states came from, going backward in time. Thus the backward state change is a function $b_{ij} : \Delta \to \mathbb{R}_0^n$ which satisfies:

$$x = u + \eta f_{ij}(u) \iff u = x - \eta b_{ij}(x). \tag{7}$$

Thus, the transformations $u \leftrightarrow x$ define a continuous bijection $\Delta \leftrightarrow \Delta$.

Between times t and $t + \tau$, the population state makes the transition from P_t to $P_{t+\tau}$, during which time the row player's mixed strategy changes from $u = x - \eta b_{ij}(x)$ to x, if the action profile (i, j) has been played at time t against a random opponent. Using (4), the relationship between the probability measure at the two time periods is given by

$$P_{t+\tau}(dx) = \sum_{i,j \in \mathbf{n}} \int_{y \in \Delta} u_i y_j P_t(du) P_t(dy), \tag{8}$$

Now multiply (8) by a smooth, real-valued, but otherwise arbitrary 'test function' $\phi(x)$, then integrate over x and use (2), (3) and (7) to obtain:

$$\langle \phi \mid P_{t+\tau} \rangle = \sum_{i,j \in \mathbf{n}} \int_{u \in \Delta} \phi\left(u + \eta f_{ij}(u)\right) P_t\left(du\right) \langle P_t \rangle_j, \tag{9}$$

This defines a general form of the discrete-time updating dynamics, determining the evolution of P_t.

2.2 The continuous time limit

We shall be interested in the continuous-time updating dynamics obtained by taking the limit as the exposure time $\tau \to 0$. To provide some justification for taking this limit, we suppose that the matching rate $1/\tau$ is an increasing function of the population size N. This reflects the idea that, as the population gets larger, opportunities for interactions between agents increase, and the more activity there is amongst agents, the faster do decisions have to be made and their consequences (in the form of strategy revisions) delivered and absorbed. Thus, on this interpretation, the limit $\tau \to 0$ is equivalent to the infinite population limit, $N \to \infty$. We can therefore construe the subsequent discussion as relating to the evolutionary dynamics within a very large population.

We show in Appendix A.1 that the continuous-time limit $\tau \to 0$, applied to the discrete-time dynamics (9), yields the following *weak form of the continuity equation*[8]:

$$\frac{d}{dt} \langle \phi \mid P_t \rangle = \int_{x \in \Delta} \nabla \phi(x) \cdot \left[\mathcal{F}(x) \langle P_t \rangle \right] P_t(dx), \tag{10}$$

where $\mathcal{F}(x)$ is the $n \times n$ matrix whose (i, j)-th entry is:

$$\mathcal{F}_{ij}(x) = \sum_{r=1}^{n} x_r f_{rj,i}(x). \tag{11}$$

The weak continuity equation (10) provides the dynamical equation that describe the evolution of the probability measure P_t from a specified initial measure P_0. Note that (10) is non-linear in P_t.

2.3 Absolute Continuity and Strong Form of the Continuity Equation

If a probability measure P is absolutely continuous with respect to Lebesgue measure, then we can write $P(dx) = p(x)dx$ for some L_1 probability density function $p(x)$. We show in Appendix A.2 that if P_t is a solution to (10) with absolutely continuous initial measure P_0, then P_t is absolutely continuous for all $t \geq 0$. In this case, the weak continuity equation (10) can be expressed in terms of probability densities $p_t(x)$.

[8] For a differentiable scalar function $\phi : \mathbb{R}^n \to \mathbb{R}$, the *gradient* of ϕ, $\nabla \phi$, is the vector field on $\mathbb{R}^n$ defined by $\nabla \phi(x) = (\partial \phi(x)/\partial x_1, \ldots, \partial \phi(x)/\partial x_n)$. See, for example, Margenau and Murphy (1962) for a general discussion of the differential operator ∇.

If we assume in addition that the density function $p_t(x)$ is differentiable in both x and t, then we may obtain the *strong form of the continuity equation*. Thus, taking $\phi(x) = 0$ for $x \in \partial\Delta$ (the boundary of Δ), using (3) and integrating by parts on the right-hand side in (10), we obtain[9]:

$$\int_{\Delta} \phi(x) \left\{ \frac{\partial p_t(x)}{\partial t} + \nabla \cdot [p_t(x)\mathcal{F}(x)\langle p_t \rangle] \right\} dx = 0.$$

Since this holds for all differentiable test functions $\phi(x)$ which vanish on $\partial\Delta$, we obtain the differential form of the continuity equation:

$$\frac{\partial p_t(x)}{\partial t} + \nabla \cdot [p_t(x)\mathcal{F}(x)\langle p_t \rangle] = 0, \qquad x \in \operatorname{int}\Delta,\ t > 0. \tag{12}$$

This is the strong form of the continuity equation, which applies to differentiable density functions.

The strong continuity equation for smooth densities, (12) gives the dynamical equation that describes the evolution of the probability density p_t. Intuitively, $\mathcal{F}(x)\langle p_t \rangle$ represents the adaptation 'velocity' of mixed strategy x.[10] That is, $[\mathcal{F}(x)\langle q_t \rangle]\,\tau$ is the expected change in mixed strategy x in the small time interval τ in response to a play of the game. Since the mass of x is represented by $p_t(x)$, $[\mathcal{F}(x)\langle q_t \rangle]\,p_t(x)$ gives the probability mass flow at x. The divergence of this vector field therefore gives the rate at which the probability mass in a small neighbourhood of x is expanding or contracting.

3 Replicator Continuity Equations

Equation (10) gives the general (weak) form of the continuity equation for a symmetric, 2-player game. In this section, we derive a particular form of this equation—the *replicator continuity equation*. We generate this equation using a particular form of reinforcement learning—the Cross (1973) learning rule as developed in Borgers and Sarin (1997)—as our forward state change rule. This rule therefore provides the microfoundations for this particular continuity equation.

Reinforcement models have been widely studied in the learning literature. A group of players, one in each role in the game, employ mixed strategies in each round of a game. Reinforcement models are based on the idea that if the action currently employed obtains a high payoff, then the probability assigned to it increases in the next round of play. Reinforcement models are therefore

[9] The formal argument has the following form. For X a vector field on the domain Δ, we use the identity $\nabla \cdot [\phi X] = \phi \nabla \cdot X + \nabla \phi \cdot X$ to obtain

$$\int_{\Delta} \nabla \phi \cdot X dV = \int_{\Delta} \nabla \cdot [\phi X]\, dV - \int_{\Delta} \phi \nabla \cdot X dV$$

Now use the divergence theorem (Margenau and Murphy, 1962) together with the assumption that $\phi = 0$ on $\partial\Delta$ to obtain:

$$\int_{\Delta} \nabla \cdot [\phi X]\, dV = \int_{\partial\Delta} (u \cdot X)\, \phi dA = 0.$$

[10] In the next section, we provide two strategy updating rules in which this velocity is given by the replicator dynamic.

extremely naive models of learning. Agents mechanically respond to stimuli from their environment without seeking to create any model of the situation or strategically evaluate how they are doing. Hence, they do not seek to exploit the pattern of opponents' past play and predict the future behaviour of their opponents.[11] In this sense, agents are boundedly rational.

We describe the Cross rule now. If the row player plays action $i \in \mathbf{n}$ and the column player plays action $j \in \mathbf{n}$, the payoff to the row player is u_{ij}. The expected payoff to the row player's play of i against the column player's mixed strategy x is $(Ux)_i$, where U is the $n \times n$ payoff matrix $U = (u_{ij})$. That is

$$e_i \cdot Ux = (Ux)_i = \sum_{j=1}^{n} u_{ij} x_j. \tag{13}$$

Here, e_i is the i-th standard basis vector of $\mathbb{R}^n$: the vector whose r-th co-ordinate is δ_{ir}. We consider a player in a 2-player game who employs strategy $x \in \Delta$, uses action i, and encounters an opponent who uses action j in the current round. The player then updates her strategy to x' according to an updating rule $f_{ij}(x)$, as in (5). The Cross rule assumes that all payoffs satisfy $0 \leq u_{ij} \leq 1$. Since it is always possible to rescale payoffs to meet these conditions without affecting incentives, we do not consider this a severe restriction. Under this rule, the mixed strategy x' and the forward state change vector take the form

$$x'_r = \delta_{ir} u_{ij} \eta + (1 - u_{ij}\eta) x_r, \tag{14}$$

$$f_{ij,r}(x) = (\delta_{ir} - x_r) u_{ij}. \tag{15}$$

Note that $\sum_r f_{ij,r}(x) = 0$ and $x_r + f_{ij,r}(x) \geq 0$ because $0 \leq u_{ij} \leq 1$. Hence, (14) defines an allowable updating rule of the form (5). Clearly, strategy i is always positively reinforced if $u_{ij} > 0$ and $x_i < 1$, and strategy $r \neq i$ is negatively reinforced if $x_r > 0$.

Recalling the notation of (13), we introduce the following family of operators $R(x) : \mathbb{R}^n \to \mathbb{R}^n_0$, with $\mathbb{R}^n_0$ as in (6), defined for $x \in \Delta$ by:

$$R_i(x)\, y = x_i \left\{ (Uy)_i - x \cdot Uy \right\} \qquad x \in \Delta, y \in \mathbb{R}^n. \tag{16}$$

Clearly, the vector field $v(x) = R(x)x$ defined for $x \in \Delta$ is identical to the vector field generated by the symmetric replicator dynamic on Δ (Taylor, 1979). Hence, we call the $n \times n$ matrix operator $R(x)$ the symmetric *replicator operator* defined by the payoff matrix U.

Under the Cross rule, the specific form of the operator $\mathcal{F}(x)$ in (10) is simply the replicator operator defined in (16). We establish this in the following lemma.

Lemma 3.1 *Under the Cross learning rule (14)-(15), $\mathcal{F}(x) = R(x)$.*

[11] Börgers and Sarin (1997) provide some justification of why agents respond to very limited information in these models–only their own payoffs. They argue that the acquisition or processing of new information may be too costly relative to benefits. Hence, they say, reinforcement models may be more plausible if agents' behaviour is habitual rather than the result of careful reflection.

Proof. We show that for $f_{ij}(x)$ given by (15), $[\mathcal{F}(x)y]_r = R_r(x)y$, for $1 \leq r \leq n$ and $x, y \in \Delta$. From (11) we have

$$\begin{aligned}
[\mathcal{F}(x)\, y]_r &= \sum_{i,j=1}^{n} x_i f_{ij,r}(x) y_j \\
&= \sum_{i,j=1}^{n} x_i y_j \left(\delta_{ir} - x_r\right) u_{ij} \\
&= x_r \left(\sum_{j=1}^{n} u_{rj} y_j - \sum_{i,j=1}^{n} x_i u_{ij} y_j \right) \\
&= x_r \left\{ [Uy]_r - x \cdot Uy \right\} \\
&= R_r(x) y. \ \blacksquare
\end{aligned}$$

The application of the Cross learning rule allows us to extend the axiomatic framework developed in Borgers et. al. (2004) to the analysis of evolution in a large population model. Their first axiom, absolute expediency, requires that in expected terms, payoff obtained from a learning rule strictly increase over time. Their second axiom, monotonicity, demands that, again in expected terms, the probability assigned to the best actions increase over time. Together, the two properties imply that in an environment of uncertainty (for example, where rivals' strategies are not fixed), the players move towards better choices monotonically in expected terms.[12] Borgers et. al. (2004) argue that among the wide variety of learning rules, those that possess these properties are the ones that are most appropriate as models of experience based learning. Extending this axiomatic approach to population games has the clear advantage of allowing the evaluation of any individual behavioral norm in evolutionary game theory on the basis of rigorous principles rather than on the grounds of heuristic plausibility.

The following proposition is now immediate.

Proposition 3.2 *Under the forward state change rule (14)-(15), the continuity equation (10) is given by*

$$\frac{d}{dt}\langle \phi \mid P_t \rangle = \int_{x \in \Delta} \nabla \phi(x) \cdot \left[R(x) \langle P_t \rangle \right] P_t(dx). \tag{17}$$

We call (17) the *replicator continuity equations.* This is the weak form of the replicator continuity equation. There is an obvious strong form corresponding to (12) for measures characterized by differentiable density functions.[13]

[12] Example 1 in Borgers et. al. (2004) establishes that the Cross rule satisfies these two axioms. While establishing absolute expediency requires some additional calculations, monotonicity follows from the fact that the expected change in the probability of any action is given by the replicator dynamic.

[13] Consider a symmetric game with payoff matrix U and let $\mu_t = \langle p_t \rangle$ be the mean. The strong form of the replicator continuity equation is $\frac{\partial p_t(x)}{\partial t} + \nabla \cdot [R(x)\mu_t p_t(x)] = 0$. We note that this dynamic is very different from those used in the early biology literature concerning the evolution of mixed strategies. For example, Zeeman (1981) uses a straightforward adaptation of the replicator dynamic having the form $\frac{\partial p_t(x)}{\partial t} = p_t(x)\,(x \cdot U\mu - \mu \cdot U\mu)$, and Hines (1980) uses the mean payoff adjusted replicator dynamic, $\frac{\partial p_t(x)}{\partial t} = \frac{p_t(x)}{\mu \cdot U\mu}\,(x \cdot U\mu - \mu \cdot U\mu)$ introduced in Maynard Smith (1982) for pure strategies.

Proposition 3.2 also establishes that the Cross rule does not provide a validation of the replicator dynamic as an evolutionary dynamic, as argued in Borgers and Sarin (1997) or Borgers et. al. (2004). In the population game, the replicator dynamic is simply the expected change in the mixed strategy of a particular player. Instead, under the Cross rule, the population state evolves according to the replicator continuity equation. Nevertheless, as we show in subsection 3.1 below, the change in the population share of a pure action is closely related to the replicator dynamic.

3.1 Homogeneous populations

The simplest case of (17) occurs when the population of agents is *homogeneous*, in the sense that they all begin by using the same mixed strategy $x_0 \in \Delta$. That is, the initial population distribution is $P_0 = \delta_{x_0}$. In this case, the solution of (19) is $P_t = \delta_{x_t}$, where x_t is the solution trajectory of the classical replicator dynamics with initial condition x_0.

To see this, it suffices to substitute this proposed solution into (17) and note that $\langle P_t \rangle = x_t$, to obtain:

$$\frac{d}{dt}\left[\phi\left(x_t\right)\right] = \nabla\phi(x_t) \cdot \left[R(x_t)x_t\right].$$

The left-hand side is $\nabla\phi(x_t) \cdot \dot{x}_t$, which is equal to the right-hand side since x_t is a solution of the replicator dynamics, $\dot{x} = R(x)x$.

This shows that the classical replicator dynamics describes the evolutionary (continuity) dynamics of a homogeneous population, in which all agents use a common mixed strategy. However, from other derivations of these dynamics, they can also be construed as mixed populations of pure-strategy players, who use an updating rule such as imitation (see the Introduction), rather than reinforcement. The use of the mixed strategy interpretation is adapted specifically to reinforcement learning.

3.2 Mean replicator dynamics

As an example of (17), it is instructive to derive a more explicit form of the dynamics for the mean $\mu_t = \langle P_t \rangle$. Taking $\phi(x) = x_i$ in (17), we have:

$$\begin{aligned}
\frac{d}{dt}\mu_i(t) &= \int_\Delta R_i(x)\mu(t)P_t(dx) \\
&= \int_\Delta x_i\left(e_i - x\right) \cdot U\mu(t)P_t(dx) \\
&= \left(e_i \cdot U\mu(t)\right)\int_\Delta x_i P_t(dx) - \int_\Delta x_i x \cdot U\mu(t)P_t(dx) \\
&= \left(e_i \cdot U\mu(t)\right)\mu_i(t) - \left(\mu(t) \cdot U\mu(t)\right)\mu_i(t) - \mathrm{Cov}_t\left(x_i, x\right) \cdot U\mu(t) \\
&= R_i\left(\mu(t)\right)\mu(t) - \mathrm{Cov}_t\left(x_i, x\right) \cdot U\mu(t),
\end{aligned}$$

where $\mathrm{Cov}(x,x)$ is the covariance matrix

$$\mathrm{Cov}(x,x)_{ij} = \mathrm{Cov}(x_i,x_j) = \int_\Delta (x_i-\mu_i)(x_j-\mu_j)P(dx). \tag{18}$$

That is, the continuity replicator dynamics for means can be written in the form

$$\frac{d\mu}{dt} = R(\mu)\mu - \mathrm{Cov}(x,x)U\mu. \tag{19}$$

Equation (19) makes clear that the continuity replicator dynamics of means differs from the classical replicator dynamics applied to the population mean through a covariance term, which cannot be reduced to a function of the mean.[14]

3.3 Rest points

If x^* is a rest point of the symmetric replicator operator, i.e. $R(x^*)x^*=0$, and we take $P_0=\delta_{x^*}$,[15] the mass-point distribution at x^*, then the initial mean is $\langle P_0\rangle = x^*$, and hence from (17), $\frac{d}{dt}\langle\phi \mid P_t\rangle|_{t=0}=0$. Thus, δ_{x^*} defines a rest point of the replicator continuity dynamics (17). In particular, this is the case if x^* is a Nash equilibrium of the underlying game. In this section we give a general characterization of rest points of the symmetric continuity equation (17).

Let $\mu\in\Delta$ be fixed, and define

$$\Delta(\mu) = \{x\in\Delta \mid R(x)\mu = 0\}, \tag{20}$$

Define the *generalised support of* $\mu\in\Delta$ by

$$\mathcal{S}_U(\mu) = \{S\subseteq \mathbf{n} \mid e_i\cdot U\mu = \pi_S \text{ for each } i\in S\}. \tag{21}$$

Here, π_S denotes the common value of $e_i\cdot U\mu$ for $i\in S$. Note that $\{i\}\in\mathcal{S}(\mu)$ for each $i\in\mathbf{n}$. Now observe that

$$x\in\Delta(\mu) \iff \mathrm{supp}(x)\in\mathcal{S}_U(\mu), \tag{22}$$

With this notation in place, we can now characterize rest points of the replicator continuity dynamics.

Proposition 3.3 *Let P_0 be a probability distribution on Δ, with $\langle P_0\rangle=\mu\in\Delta$. Then $P_t=P_0$ is*

[14]We note that Hines (1980) derives an equation for mean dynamics from the mean-payoff adjusted replicator dynamic (see footnote 13). His equation has the form

$$\dot\mu = \frac{1}{\mu\cdot U\mu}\mathrm{Cov}\,(x,x)\,U\mu.$$

[15]The mass-point, or Dirac measure at $x^*\in\Delta$ is defined by: $\langle\phi\mid\delta_{x^*}\rangle=\phi(x^*)$ for any differentiable function ϕ on Δ. By convention, this distribution is represented by the Dirac probability 'density': $\delta_{x^*}(dx)=\delta(x-x^*)dx$. We sometimes adopt this convention.

a stationary solution of the symmetric continuity replicator dynamics (17) if and only if

$$supp(P_0) \subseteq \Delta(\mu). \tag{23}$$

Proof. A. The stationarity condition for the symmetric replicator continuity equation is

$$\frac{d}{dt}\langle \phi \mid P_t \rangle\bigg|_{t=0} = \int_\Delta \nabla\phi(x) \cdot [R(x)\mu]\, P_0(dx) = 0,$$

for any differentiable test function ϕ. This condition holds if and only if $R(x)\mu = 0$ for all $x \in \text{supp}(P_0)$[16]. From (22), these conditions are equivalent to (23). ■

We give two specific examples below.

3.3.1 Example

Suppose given $\mu \in \Delta$. Then $R(e_i)\mu = 0$ for $1 \le i \le n$, and hence $\{e_1, \ldots, e_n\} \subset \Delta(\mu)$. It follows from proposition 3.3 that any initial probability distribution P_0 with $\text{supp}(P_0) = \{e_1, \ldots, e_n\}$ and $\langle P_0 \rangle = \mu$ is a rest point of the symmetric replicator continuity dynamics (17). That is, any probability distribution of the form

$$P_0 = \sum_{i=1}^{n} \mu_i \delta_{e_i}, \tag{24}$$

is a rest point of (17).

3.3.2 Example

We show that *any* initial distribution P_0 whose mean x^* a rest point of the replicator dynamics, is a stationary solution of the continuity replicator dynamics (17). In particular, this is the case if x^* is a Nash equilibrium of the game of the underlying game.

Suppose that $\mu = x^* \in \Delta$ is a rest point of the replicator dynamics, so that $R(x^*)x^* = 0$. Then by definition of $\mathcal{S}(\mu)$ as in (21), it follows that $S \in \mathcal{S}(\mu)$ for any $S \subseteq \text{supp}(\mu)$. In particular, if x^* has full support, then (22) implies that $\Delta(\mu) = \Delta$. In this case (23) is automatically satisfied, so proposition 3.3 implies that a probability distribution P_0 is a rest point only if it's mean is μ.

If x^* does not have full support, then P_0 cannot have mean $\mu = x^*$ unless P_0 is supported on the lowest dimensional face of Δ containing x^*. Since x^* has full support on this face, the above argument shows that P_0 is a rest point if $\langle P_0 \rangle = x^*$.

[16] Recall that the support of a Borel probability measure P is the smallest closed set C for which $P(C) = 1$.

4 Solution of the General Continuity Equation: Liouville's Formula

Our approach to solving the non-linear continuity equations we have constructed is to begin by solving a different, but related problem. Thus, instead of confronting the non-linearities directly, we first consider a linear continuity equation, but one defined by an explicitly time-dependent vector field. We will later show how a solution of the non-linear continuity equations of interest can be constructed from explicit solutions of linear continuity equations of this type.

4.1 Liouville's Formula

Let $X = X(x,t) \in I\!R^n$ be a (possibly time-dependent) smooth vector field defined for x in a neighbourhood of the state space $\Omega \subset I\!R^n$, where Ω is a compact, connected domain with non-empty interior and piecewise smooth boundary. We assume that Ω is invariant under the flow determined by $X(x,t)$. Let P_t be a probability measure on Ω satisfying the linear weak continuity equation

$$\frac{d}{dt}\langle \phi \mid P_t\rangle = \int_\Omega \nabla\phi(x)\cdot X(x,t)P_t(dx), \tag{25}$$

for all smooth test functions $\phi(x)$, and for given initial measure P_0. The solution to this initial-value problem may be described as follows.

We first introduce some notation to describe the solution trajectories to the (non-autonomous) differential equations defined by X,

$$\frac{dx}{dt} = X(x,t). \tag{26}$$

Let $x_{t_0,t}(x), t \in I\!R$, denote the solution trajectory to (26) that passes through the point $x \in \Omega$ at time t_0. Thus, the trajectory that passes through x at time t starts at the point $x_{t,0}(x)$ when $t = 0$.[17] After time $s \geq 0$, this trajectory has reached the point $x_{t,s}(x) = x_{0,s}\left(x_{t,0}(x)\right)$. In particular, $x_{t,t}(x) = x_{0,t}\left(x_{t,0}(x)\right) = x$, and by definition $x_{t,0}\left(x_{0,t}(x)\right) = x$.

We can now write down the solution to the initial value problem (25). For a Borel set B,

$$P_t(B) = P_0\left(x_{t,0}(B)\right). \tag{27}$$

This is Liouville's formula for measures. A proof is given in Appendix A.2.

In the case in which P_0 is absolutely continuous, so that $P_0(dx) = p_0(x)dx$ for an initial density function $p_0(x)$, then it is also shown in Appendix A.2 that the solution (27) is described by a more classical form of Liouville's formula which determines the density function $p_t(x)$ associated to P_t:

$$p_t(x) = p_0\left(x_{t,0}(x)\right)\exp\left\{-\int_0^t \left[\nabla\cdot X\right]\left(x_{t,s}(x), s\right) ds\right\}. \tag{28}$$

[17]Note that the situation for a non-autonomous vector field is more complicated than for the more familiar autonomous case. This is because the explicit time dependence of $X(x,t)$ imposes an absolute, rather than a relative, time-scale on the dynamics. In particular, the initial time $t = 0$ is exogenously determined.

4.2 Expected Values

Liouville's formula (27) allows us to calculate expected values of associated variables in terms of the initial measure P_0 and solutions of the characteristic system (26). Thus, for a smooth function $\phi(x)$, define its expected value with respect to the probability density P_t satisfying (25) by:

$$\langle \phi \mid P_t \rangle = \int_\Omega \phi(x) P_t(dx). \tag{29}$$

Then we have:

Proposition 4.1 *The expected value $\langle \phi \mid P_t \rangle$ may be expressed in the form:*

$$\langle \phi \mid P_t \rangle = \int_\Omega \phi\left(x_{0,t}(x)\right) P_0(dx). \tag{30}$$

A proof is given in Appendix A.2.

As an example of the use of (30), the following Corollary shows that the trajectories of the underlying characteristic dynamics (26) may be recovered as solutions of the continuity equation (25) for initial conditions which are mass points.

5 Application of Liouville's Formula to Replicator Continuity Equation

In this section, we use Liouville's formula (27) to lay the foundations for a solution to the nonlinear replicator continuity equation (17). To do this, we replace the matching scenario described in section 2.1 by the following simpler scenario.

Suppose given a specified mixed strategy history $y(t) \in \Delta$, $t \geq 0$. We assume there is a player called 'Nature' (or the 'Environment') who uses the mixed strategy $y(t)$ at time t when playing against an opponent chosen from the given population. In effect, in the matching scenario of section 2.1, one of the two chosen players is replaced by 'Nature'.

We associate a time-dependent replicator vector field to this scenario: $X(x,t) = R(x)y(t)$. This defines a continuity equation of the form (25). The associated characteristic ODE system (26) we call the *pseudo replicator dynamic* associated to the history $y(t)$. This takes the form of the explicitly time-dependent dynamical system

$$\dot{x}_i = R_i(x)y(t) = x_i(e_i - x) \cdot Uy(t), \tag{31}$$

whose solutions specify the time-development of the population players' mixed strategies in response to to plays against Nature. To solve the associated continuity equation (25), we begin by solving the characteristic system (31). We can then find the solution of any associated initial value problem of the form (25) by means of Liouville's formula (27).

5.1 Solution of the pseudo Replicator Dynamics

Write $c(t) = Uy(t) \in I\!R^n$, a time-dependent vector-payoff stream to row players. Then the pseudo-replicator equations (31) can be written as:

$$\frac{dx_i}{dt} = x_i\,(e_i - x)\cdot c(t), \qquad\qquad c(t) = Uy(t). \tag{32}$$

Write

$$C(t) = \int_0^t c(s)ds. \tag{33}$$

Then we can express the solutions of (32) as follows.

Proposition 5.1 *The solution trajectory of the pseudo-replicator dynamics (32) passing through $x \in \Delta$ at time $t = t_0$ is:*

$$x_{t_0,t}(x)_i = \frac{x_i e^{C_i(t)-C_i(t_0)}}{x\cdot e^{C(t)-C(t_0)}}. \tag{34}$$

In particular:

$$x_{0,t}(x)_i = \frac{x_i e^{C_i(t)}}{x\cdot e^{C(t)}}, \qquad \mathit{and} \qquad x_{t,0}(x)_i = \frac{x_i e^{-C_i(t)}}{x\cdot e^{-C(t)}}. \tag{35}$$

Proof. With $x_{t_0,t}(x)$ given by (34), a direct calculation gives

$$\frac{d}{dt}\left[x_{t_0,t}(x)_i\right] = x_{t_0,t}(x)_i\left\{e_i - x_{t_0,t}(x)\right\}\cdot c(t),$$

which shows that $x_{t_0,t}(x)$ is a solution of (32). It also follows from (34) that $x_{t_0,t_0}(x) = x$, as required. ■

5.2 Solution of the pseudo Replicator Continuity Equation

In the case in which P_0 is absolutely continuous, we may use Liouville's formula (28), together with Proposition 5.1, to compute the solution to the replicator continuity equation (25) associated with a pseudo-replicator vector field of the form (32). This is given in the following proposition, proved in Appendix A.3.

Proposition 5.2 *The solution of the initial value problem (25) with initial density $p_0(x)$, associated to the characteristic vector field (32) is:*

$$p_t(x) = p_0\left(\frac{xe^{-C(t)}}{x\cdot e^{-C(t)}}\right)\left(\frac{1}{x\cdot e^{-C(t)}}\right)^n \exp\left\{-e\cdot C(t)\right\}, \tag{36}$$

where $C(t) \in I\!R^n$ is given by (33) and $e \in \mathbb{R}^n$ is the vector all of whose entries are 1.

More generally, we may obtain the expected value of a continuous function $\phi(x)$ from (30) and (35):

$$\langle \phi \mid P_t \rangle = \int_\Delta \phi\left(\frac{xe^{C(t)}}{x \cdot e^{C(t)}}\right) P_0(dx). \tag{37}$$

We now present an immediate implication of (37) giving conditions under which a row-player pure strategy is eventually eliminated.

Proposition 5.3 *Suppose there exists an i such that $[C_i(t) - C_j(t)] \to \infty$ as $t \to \infty$ for some $j \neq i$, and the i-th face, $\partial\Delta^{(i)} = \{x \in \Delta : x_i = 0\}$, has P_0-measure zero. Then $\langle P_t \rangle_j \to 0$ as $t \to \infty$.*

Proof. For $x \in \Delta \setminus \partial\Delta^{(i)}$, we have $x_i > 0$. Thus:

$$\frac{x_j e^{C_j(t)}}{x \cdot e^{C(t)}} = \frac{x_j e^{C_j(t)}}{x_i e^{C_i(t)} + \sum_{k \neq i} x_k e^{C_k(t)}} = \frac{x_j e^{-[C_i(t) - C_j(t)]}}{x_i + \sum_{k \neq i} x_k e^{-[C_i(t) - C_k(t)]}} \to 0 \qquad \text{as } t \to \infty,$$

since the denominator is never zero. Hence, from (37),

$$\langle P_t \rangle_j = \int_{\Delta \setminus \partial\Delta^{(i)}} \left(\frac{x_j e^{C_j(t)}}{x \cdot e^{C(t)}}\right) P_0(dx) \;\to\; 0 \qquad \text{as } t \to \infty. \quad \blacksquare$$

6 Distributional Replicator Dynamics

In this section we show how a solution to the replicator continuity equation (17) associated with a 2-player, n-strategy symmetric game having $n \times n$ payoff matrix U. In terms of the theory of section 5, this is the continuity equation associated to the time-dependent mixed strategy history given by $y(t) = \langle P_t \rangle$. This identifies 'Nature' as an (average) population player. That is, $c(t) = U\langle P_t \rangle$. Thus, from (32) and (33) we have

$$\frac{dC(t)}{dt} = c(t) = U\langle P_t \rangle, \tag{38}$$

and using (37) with $\phi = \iota$, we therefore obtain a system of n differential equations in the variables $C_1, \ldots, C_n$:

$$\frac{dC_i}{dt} = \sum_{j=1}^{n} u_{ij} \int_\Delta \left(\frac{x_j e^{C_j}}{x \cdot e^C}\right) P_0(dx), \qquad C_i(0) = 0, \quad 1 \leq i \leq n. \tag{39}$$

We call equations (39) the symmetric *distributional replicator dynamics* (DRD) associated with the initial measure P_0. The solutions of these equations with the given initial conditions define trajectories $C(t)$, in terms of which the continuity dynamics can be completely specified as in (37), or (36) in the absolutely continuous case.

Note that at most $n-1$ of equations (39) are independent.[18] For example, setting $A_i = C_i - C_n$,

[18] Because of the constraints $\sum_i \langle P_t \rangle_i = 1$.

equations (39) can be reduced to

$$\frac{dA_i}{dt} = \sum_{j=1}^{n} (u_{ij} - u_{nj}) \int_{\Omega} \left(\frac{x_j e^{A_j}}{x \cdot e^A} \right) P_0(dx), \qquad A_i(0) = 0, \quad 1 \le i \le n-1, \tag{40}$$

where $\Omega \subset I\!R^{n-1}$ is the projection of Δ onto $I\!R^{n-1}$ obtained by setting $x_n = 1 - \sum_{i=1}^{n-1} x_i$. Of course $A_n = 0$. Note that the formulae (36) and (37) can be expressed in terms of the A_i's.

6.1 Alternative forms of DRD

For our purposes, it is most useful to express the DRD (39) in a modified formulation. First write equations (39) in the vector form

$$\frac{dC}{dt} = UF\left(e^C \mid P_0\right), \qquad C(0) = 0, \tag{41}$$

where where $F(\cdot \mid P_0) : I\!R^n_+ \to \Delta$ is the function

$$F_i(\xi \mid P_0) = \int_{\Delta} \left(\frac{\xi_i x_i}{\xi \cdot x} \right) P_0(dx) \qquad 0 \le i \le n. \tag{42}$$

Clearly $0 \le F_i \le 1$ and $e \cdot F = 1$, and hence $F \in \Delta$. Further, F is homogeneous of degree 0 in ξ; i.e. $F(\alpha\xi \mid P_0) = F(\xi \mid P_0)$ for any non-zero scaler α. Also $F(e \mid P_0) = \langle P_0 \rangle$. More generally, it follows from (37) that if $C(t)$ is the solution trajectory of (41) with $C(0) = 0$, then

$$F(e^{C(t)} \mid P_0) = \langle P_t \rangle. \tag{43}$$

Additional key properties of the function F are proved in Appendix A.4.1.

Now define a new variable $\xi = e^C/(e \cdot e^C) \in \Delta$. Then a straightforward calculation shows that the distributional replicator equation (41) can be written in the form

$$\frac{d\xi}{dt} = R(\xi)F(\xi \mid P_0), \qquad \xi(0) = \frac{1}{n}e, \tag{44}$$

where $R(\xi)$ is the replicator operator defined in (16).

6.1.1 The classical replicator dynamics as DRD

We can reconstruct the trajectory of the classical replicator dynamics with given initial condition $x_0 \in \Delta$ from a solution of (44) as follows. Take $P_0 = \delta_{x_0}$. Then from (42), $F(e/n \mid P_0) = \langle P_0 \rangle = x_0$. Thus, if ξ_t is the solution trajectory of (44) with initial condition $\xi_0 = e/n$, then from (43), $F(\xi_t \mid P_0) = \langle P_t \rangle$, where P_t is the solution of the replicator continuity equation with initial condition δ_{x_0}. As noted in section 3.1, this solution is $P_t = \delta_{x_t}$, where x_t is the solution trajectory of the classical replicator dynamics with initial condition x_0. In particular, $\langle P_t \rangle = x_t$, and hence

$$x_t = F(\xi_t \mid \delta_{x_0}) = \frac{x_0 \xi_t}{x_0 \cdot \xi_t}. \tag{45}$$

As discussed in section 3.1, this situation describes a homogeneous population in which all agents use the same (evolving) mixed strategy.

In the following sections, we study the distributional replicator dynamics to explore properties of solutions of the replicator continuity equation (17) for various classes of games.

7 Application to classes of games

7.1 Negative definite games

Example 3.3.2 establishes that any probability measure whose mean is a Nash equilibrium is a stationary solution of the replicator continuity equation (17). In this section, we study the convergence of trajectories of this dynamic to such stationary points in the class of negative definite and negative semi-definite games. These results are of interest since this class of games encompass a wide variety of well known games. For example, games with an interior ESS are negative definite games whereas two player zero-sum games are negative semi-definite games. The stability properties of Nash equilibria in such games have been established under a wide range of evolutionary dynamics.[19] Our interest is in seeing whether such stability results can be extended to our dynamical formulation.

A symmetric game with $n \times n$ payoff matrix U is said to be negative semi-definite on $\mathbb{R}^n_0$ if

$$z \cdot Uz \leq 0 \qquad \text{for all } z \in \mathbb{R}^n_0,$$

and is negative definite if this inequality is strict when $z \neq 0$. An attractive feature of negative definite games is that they have a unique Nash equilibrium (Sandholm, 2009; Theorem 3.3.16). For a negative semi-definite game, the set of Nash equilibria is convex. Zero-sum games are the most prominent examples of negative semi-definite games.

We show that in negative definite games, the mean social state always converges towards the unique Nash equilibrium under the replicator continuity dynamics. The following theorem is proved in Appendix A.4.2.

Theorem 7.1 *Consider a symmetric $n \times n$ game with payoff matrix U, and suppose that U is negative definite on $\mathbb{R}^n_0$, with unique Nash equilibrium $x^* \in \Delta$. Let P_0 be a probability distribution on Δ for which $\partial\Delta$ has zero P_0-measure, and let P_t be the solution of the continuity replicator dynamics with initial condition P_0. Then the mean population state $\langle P_t \rangle \to x^*$ as $t \to \infty$. In particular, the Nash equilibrium is globally asymptotically stable under the mean replicator dynamics.*

[19]See Sandholm (2009) for a discussion of these results. Sandholm (2009) refers to negative semi- definite games as stable games and negative definite games as strictly stable games.

If the game U is negative semi-definite, then the convex set of Nash equilibria is Lyapunov stable.

7.2 Positive definite games

A symmetric game with $n \times n$ payoff matrix U is *positive definite* if $z \cdot Uz > 0$ for all non-zero $z \in \mathbb{R}^n_0$. Such a game is generic if U is invertible. In contrast to negative-definite games, a (generic) positive-definite game can have many equilibria. However, if the game is generic and admits an interior Nash equilibrium $x^* \in \text{int}\ \Delta$, then x^* is unique and is given by

$$x^* = \frac{U^{-1}e}{e \cdot U^{-1}e}. \tag{46}$$

In this section we show that if x^* exits, then it is necessarily *totally unstable* under the continuity replicator dynamics. That is, every trajectory of means beginning arbitrarily close to x^* eventually moves away from x^*. More precisely, in Appendix A.4.4 we prove the following.

Theorem 7.2 *Let U be the $n \times n$ payoff matrix of a generic, positive-definite symmetric game which admits an interior Nash equilibrium $x^* \in$ int Δ. Let P_0 be a probability measure on Δ for which $\partial\Delta$ has zero P_0-measure, and let P_t be the solution of the continuity replicator dynamics with initial condition P_0. Suppose that $\langle P_0 \rangle \neq x^*$. Then there exists a neighbourhood N_0 of x^* in int Δ such that the forward trajectory of means $\langle P_t \rangle$ eventually leaves N_0. That is, if there exists a $t_0 \geq 0$ such that $\langle P_{t_0} \rangle \in N_0$, then there exists a $t_1 > t_0$ such that $\langle P_{t_1} \rangle \notin N_0$.*

7.3 Doubly symmetric games and mean payoff

A symmetric population game is called doubly symmetric if the payoff matrix U itself is symmetric, so that $u_{ij} = u_{ji}$ (e.g. a coordination game with positive diagonal elements and zero off diagonal elements). Define the *mean payoff* with respect to the distribution P to be $\bar{w}(\mu) = \mu \cdot U\mu$, where $\mu = \langle P \rangle$ is the population mean. We first show that $\bar{w}(\mu)$ increases along non-equilibrium trajectories.[20] We then use this result to establish convergence of the population mean to the set of Nash equilibria. More precisely, in Appendix A.4.5 we prove the following.

Theorem 7.3 *Let U be the payoff matrix of a symmetric game, and suppose that U is a symmetric matrix. Let P_0 be a probability measure on Δ for which $\partial\Delta$ has zero P_0-measure. Then $\bar{w}\left(\langle P_t \rangle\right)$ increases along non-equilibrium trajectories of the continuity replicator dynamics, and the mean $\langle P_t \rangle$ converges to a level set (with respect to $\bar{w}$) of Nash equilibria.*

[20] The biological interpretation of this result is the well known Fundamental Theorem of Natural Selection in classical population genetics, in which the entries of U are genotype fitnesses. We also note that the alternative definition of mean payoff given by $\bar{w} = \langle x \cdot Ux \rangle$ need not increase along non-equlibrium trajectories.

8 Limiting distributions

Theorem 7.1 shows that, at least for negative definite games, the asymptotics of the mean are essentially independent of the initial distribution. However, this does not mean that the asymptotic distribution itself is independent of the initial distribution. As is evident from the characterization of rest points in section 3.3, there are many such stationary distributions having mean x^*, and in particular P_t need not converge to a mass-point distribution at x^*. This is very important. Although on average, the equilibrium population plays the Nash equilibrium, at the individual level the population can be very heterogeneous.

We can characterize such limiting probability distributions when they exist as follows. Suppose P_0 is a probability distribution on Δ for which $\partial\Delta$ has zero P_0-measure, and let $\{P_t\}_{t\geq 0}$ be the solution of the weak replicator continuity dynamics (17). We are interested in determining a limiting probability distribution P_∞ of this solution as $t \to \infty$. The appropriate notion of convergence here is *weak convergence* of probability measures[21]. Thus, we say that $\{P_t\}_{t\geq 0}$ converges weakly to P_∞, written $P_t \xrightarrow{w} P_\infty$ as $t \to \infty$ if

$$\langle \phi \mid P_t \rangle \longrightarrow \langle \phi \mid P_\infty \rangle, \tag{47}$$

for every continuous function $\phi : \Delta \to \mathbb{R}$.

To determine a limiting density, consider the trajectory $\xi(t)$ of the distributional replicator dynamics (44), with initial condition $\xi_0 = (1/n)e$. Then from (43) we have that $\langle P_t \rangle = F(\xi(t) \mid P_0)$[22]. Suppose that $\xi(t) \to \xi^* \in \Delta$ as $t \to \infty$, and let $x^* = F(\xi^* \mid P_0) \in \Delta$. By Lemma A.2, x^* is uniquely determined by ξ^*, and conversely. Clearly ξ^* is a rest point of the dynamics (44). Thus, by Lemma A.3, x^* is rest point of the replicator dynamics, and if a limiting distribution P_∞ of $\{P_t\}_{t\geq 0}$ exists, then P_∞ has mean x^* and hence is a stationary distribution of the replicator continuity equation (17) – see Example 3.3.2.

Under the above assumptions, we can be more specific concerning the limiting distribution, not just its mean. For $x \in \operatorname{int}\Delta$, we have

$$\begin{aligned} \Xi_t(x)_i \quad &:= \quad \frac{e^{C_i(t)} x_i}{e^{C(t)} \cdot x} = \frac{\xi_i(t) x_i}{\xi(t) \cdot x} \\ &\longrightarrow \quad \frac{\xi_i^* x_i}{\xi^* \cdot x} := \Xi^*(x)_i \qquad \text{as } t \to \infty. \end{aligned} \tag{48}$$

Thus, $\Xi^*(x)$ is defined on $\operatorname{int}\Delta$, and $\operatorname{supp}(\Xi^*(x)) = \operatorname{supp}(\xi^*)$. In particular, $\Xi^*(x)$ is defined P_0-a.e. on Δ.

Now apply the characterization (37) of the solution of the weak continuity replicator dynamics (17). Thus, if $\phi : \Delta \to \mathbb{R}$ is a continuous function, then $|\phi(x)|$ is uniformly bounded on Δ, and

[21] See, for example, Parthasarathy (1967), chapter II, section 6.

[22] Recall that $\xi = e^C/(e^C \cdot e)$ and $F(e^C \mid P_0) = F(\xi \mid P_0)$.

hence the Lebesgue dominated convergence theorem[23] implies that

$$\langle \phi \mid P_t \rangle = \int_\Delta \phi\left(\Xi_t(x)\right) P_0(dx) \longrightarrow \langle \phi \mid P_\infty \rangle = \int_\Delta \phi\left(\Xi^*(x)\right) P_0(dx) \qquad \text{as } t \to \infty. \tag{49}$$

This defines the limiting probability distribution P_∞.[24] In fact, if $B \subset \Delta$ is a Borel set, we have

$$P_\infty(B) = \int_\Delta \chi_B\left(\Xi^*(x)\right) P_0(dx), \tag{50}$$

where χ_B is the characteristic function of B.[25]

The above discussion is summarized in the following proposition.

Proposition 8.1 *Let P_0 be a probability measure on Δ for which $\partial\Delta$ has zero P_0-measure. Let ξ_t be the solution trajectory of the distributional replicator dynamics (44) with initial condition $\xi_0 = (1/n)e$. Suppose that $\xi_t \to \xi^*$ as $t \to \infty$, and let $x^* = F(\xi^* \mid P_0)$. Suppose supp$(\xi^*) = S \subseteq \mathbf{n}$, and let $\Delta_S = \{x \in \Delta \mid x_j = 0$ for $j \notin S\}$ be the face defined by S. Let $\{P_t\}_{t\geq 0}$ be the solution of the weak continuity dynamics (17) with initial condition P_0, and let P_∞ be the probability distribution given by (48) and (50). Then $P_t \xrightarrow{w} P_\infty$ as $t \to \infty$. Further, P_∞ is supported on Δ_S, and $\langle P_\infty \rangle = x^*$.*

Note in particular, that if $S = \{i\}$ is a pure strategy, then $\xi^* = x^* = e_i$, and hence $P_\infty = \delta_{e_i}$ is a mass-point distribution on the pure strategy i.

9 Simple examples

9.1 Generic 2×2 Symmetric Games

Consider a generic 2×2 symmetric game with invertible payoff matrix U. We consider the generic situation in which the payoff differences $u_{11} - u_{21}$ and $u_{22} - u_{12}$ are non-zero and have the same sign. We define two parameters λ and x^* by

$$\lambda = (u_{11} - u_{21}) + (u_{22} - u_{12}), \qquad \text{and} \qquad x^* = \frac{u_{22} - u_{12}}{(u_{11} - u_{21}) + (u_{22} - u_{12})}. \tag{51}$$

[23] Dunford and Schwartz (1964), Corollary 16, p151. Specifically, for any sequence $t_n \uparrow \infty$, define $\phi_n(x) = \phi\left(\Xi_{t_n}(x)\right)$. Then ϕ_n is defined P_0-a.e. on Δ and is continuous on int Δ for each n, and by (48) and the continuity of ϕ, $\phi_n(x) \to \phi^*(x) = \phi\left(\Xi^*(x)\right)$ P_0-a.e. Further, $\{|\phi_n|\}$ is uniformly bounded P_0-a.e. on Δ. The Lebesque theorem therefore implies that ϕ^* is P_0-integrable, and $\langle \phi_n \mid P_0 \rangle \to \langle \phi^* \mid P_0 \rangle$. This yields the statement (49).

[24] The existence of P_∞ satisfying (44) follows from a standard theorem which represents linear functionals on the space of continuous functions as integrals with respect to a unique measure: for example, Parthasarathy (1967), Theorem 5.8.

[25] More precisely, let $\{E_n\}_{n\geq 0}$ be a sequence of open neighbourhoods of B in Δ satisfying $B \subset \ldots \subset E_n \subset E_{n-1} \ldots \subset E_0$ and $B = \cap_n E_n$. Let $\phi_n : \Delta \to [0,1]$ be a continuous function satisfying $\phi_n(x) = 1$ for $x \in B$ and $\phi_n(x) = 0$ for $x \in \Delta \setminus E_n$. Then, using (49) we have

$$P_\infty(B) = \inf_n P_\infty(E_n) = \lim_{n\to\infty} \int_\Delta \phi_n\left(\Xi^*(x)\right) P_0(dx).$$

Then $z \cdot Uz = \frac{1}{2}\lambda|z|^2$ for $z \in I\!R^2_0$, and $(x^*, 1-x^*) \in \operatorname{int}\Delta$ is the unique interior Nash equilibrium.

The classical replicator dynamics for such a game can be expressed in terms of the parameters λ and x^*. Thus, for $x \in [0,1]$ there is a single independent replicator dynamic equation

$$\dot{x} = \lambda x(1-x)(x-x^*). \tag{52}$$

This may be compared to the continuity replicator dynamics for means given by equation (19), which in this case reduces to the single equation

$$\dot{\mu} = \lambda \{\mu(1-\mu) - V\} (\mu - x^*), \tag{53}$$

where V is the variance:

$$V_t = \int_0^1 (x-\mu_t)^2 P_t(dx). \tag{54}$$

As observed in section 3.3, Example 3.3.1, any distribution of the form $P_0 = (1-\alpha)\delta_0 + \alpha\delta_1$ with $\alpha \in [0,1]$ defines a rest point of the continuity equation (17), with mean $\mu = \alpha$ and variance $V = \alpha(1-\alpha)$. In particular δ_0 and δ_1 are rest points. However, in contrast to the classical case (52), the sense in which these are (if they are) locally stable rest points of the continuity dynamics is not immediately clear.

To address this question, we consider the distributional replicator dynamics in the form (40) which, for the 2×2 case, reduces to the single equation

$$\dot{A} = \lambda \left\{-x^* + F(e^A \mid P_0)\right\}, \qquad F(\xi \mid P_0) = \int_0^1 \frac{x\xi}{1-x+x\xi} P_0(dx), \tag{55}$$

with initial condition $A(0) = 0$. Note that, if P_0 has no mass points at $x = 0$ or $x = 1$, then $F(1 \mid P_0) = \langle P_0 \rangle$, and that $F(\xi \mid P_0)$ is monotonically increasing in ξ.

9.1.1 The case $\lambda < 0$: convergence to equilibrium distribution

In this case both payoff differences $u_{11} - u_{21}$ and $u_{22} - u_{12}$ are negative, and U is a negative definite game. The unique interior Nash equilibrium x^* is globally asymptotically stable for $x \in (0,1)$ under the classical replicator dynamics (52). For the continuity replicator dynamics, Theorem 7.1 and Proposition 8.1 are applicable, and determine a unique limiting distribution P_∞ with $\langle P_\infty \rangle = x^*$.

For example, suppose that P_0 is represented by a probability density function on $[0,1]$, $P_0(dx) = p_0(x)dx$. Then, using the variable $A = C_1 - C_2$, the equilibrium equation reduces to

$$F_1(e^A \mid P_0) = \int_0^1 \frac{xe^A}{1-x+xe^A} p_0(x)dx = x^*.$$

Since $F_1(e^A \mid P_0)$ is monotonically increasing in A, with $F_1(e^A \mid P_0) \to 0$ as $A \to -\infty$ and $F_1(e^A \mid P_0) \to 1$ as $A \to \infty$, this equation has a unique solution $A^* \in I\!R$.

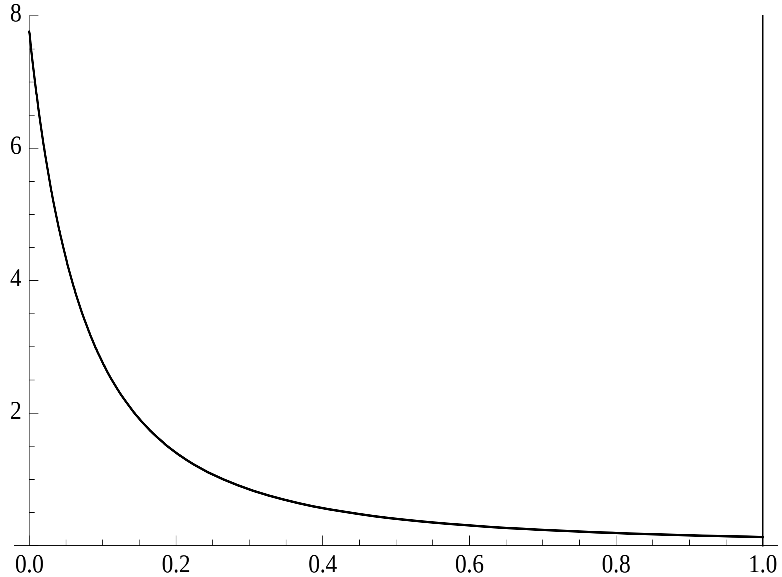

Figure 1: The limiting density $p_\infty(x)$ defined by equation (56) for the uniform distribution $p_0(x) = 1$. In this example x^* in (51) is $x^* = 0.2$, and A^* defined by (57) is $A^* = -2.0491$. This density has mean x^* and variance 0.04875.

We can now use the formula (36) to obtain the limiting probability density:

$$p_\infty(x) = p_0\left(\frac{xe^{-A^*}}{1-x+xe^{-A^*}}\right)\frac{e^{-A^*}}{(1-x+xe^{-A^*})^2}. \tag{56}$$

For example, for the uniform distribution, $p_0(x) = 1$, A^* is the solution of

$$\int_0^1 \frac{xe^A}{1-x+xe^A}dx = \frac{e^A\left(e^A - A - 1\right)}{\left(e^A-1\right)^2} = x^*. \tag{57}$$

The resulting limiting probability density (56) is illustrated in Figure 1.

9.1.2 The case $\lambda > 0$: convergence to pure strategy equilibria

In this case both payoff differences $u_{11} - u_{21}$ and $u_{22} - u_{12}$ are positive. For the replicator dynamic (52), the equilibria $x = 0$ and $x = 1$ are both locally asymptotically stable, with basins of attraction $0 \le x < x^*$ and $x^* < x \le 1$, respectively. For the distributional dynamic (55), the following lemma relates the asymptotic behaviour of $A(t)$ to the initial density function.

Lemma 9.1 *Suppose $\lambda > 0$ and $x^* \in (0,1)$, and that the initial distribution P_0 has no mass point at $x = 0$ or $x = 1$. Let $\mu_0 = \langle P_0 \rangle$ be the associated mean mixed strategy.*

1. *If $\mu_0 < x^*$, then $A(t)$ is monotonically decreasing in t, and $A(t) \to -\infty$ as $t \to \infty$.*

2. *If $\mu_0 > x^*$, then $A(t)$ is monotonically increasing in t, and $A(t) \to \infty$ as $t \to \infty$.*

Proof. From (55), we have $\dot{A}(0) = \lambda(-x^* + \mu_0)$. Since $\lambda > 0$, $\dot{A}(0) > 0$ if $\mu_0 > x^*$ and $\dot{A}(0) < 0$ if $\mu_0 < x^*$. Moreover, the monotonicity properties of $F\left(e^A \mid P_0\right)$ imply that the initial conditions are self-reinforcing as t increases. Hence, if $\mu_0 > x^*$, then $\dot{A}(t) > 0$, and if $\mu_0 < x^*$, then $\dot{A}(t) < 0$, for all $t \geq 0$. ■

We now use Proposition 5.3 and Lemma 9.1 to derive the following proposition.

Proposition 9.2 *Consider a generic 2×2 symmetric game with $\lambda > 0$ and $0 < x^* < 1$. Let $\mu_t = \langle P_t \rangle$ be the mean with respect to the solution measure P_t. If P_0 has no mass point at $x = 0$ or $x = 1$, then $P_t \to \delta_1$ if $\mu_0 > x^*$, and hence $\mu_t \to 1$, and $P_t \to \delta_0$ if $\mu_0 < x^*$, and hence $\mu_t \to 0$ as $t \to \infty$.*

Earlier, we interpreted the mean $\mu_t = \langle P_t \rangle$ as the aggregate social state generated by P_t. Proposition 9.2 implies in the type of 2×2 symmetric games we are considering, and for suitable initial distributions, there is no difference in the long-run aggregate social state under the replicator continuity equation and the classical replicator dynamic. In the long run either all agents play action 1 or all play action 2. Nevertheless, the time-course trajectories of the aggregate state under the two dynamics generally differ. In principle, therefore, it would be possible to distinguish whether agents are playing pure or mixed strategies by observing the solution trajectories.

9.2 Population heterogeneity: a 3×3 example

As shown in section 9.1.2, for generic 2×2 symmetric games there is no difference in the long-run aggregate social state under the replicator continuity equation and the classical replicator dynamic. This coincident asymptotic behavior is not, however, a general result, and does not hold for $n \times n$ symmetric games with $n > 2$. In this section we give an example for $n = 3$.

This issue is closely related to population heterogeneity. As noted in section 3.1, solutions of the classical replicator dynamics represent the evolutionary continuity dynamics of a homogeneous population, in which all agents use the same mixed strategy, say $P_0 = \delta_{x_0}$. We compare this with a simple heterogeneous population which initially contains two subpopulations, using mixed strategies $a_1, a_2 \in \operatorname{int} \Delta$. In this case, $P_0 = \alpha_1 \delta_{a_1} + \alpha_2 \delta_{a_2}$, where $\alpha_1, \alpha_2 > 0$ and $\alpha_1 + \alpha_2 = 1$. Thus, α_i is the proportion of the population using mixed strategy a_i, $i = 1, 2$. We assume that the two populations initially have the same mean: $\mu_0 = \alpha_1 a_1 + \alpha_2 a_2 = x_0$. Thus, regarded as mixed populations of pure strategy players, the two populations are initially indistinguishable.

Consider a symmetric 3×3 game with diagonal payoff matrix $U = \operatorname{diag}\{\lambda_1, \lambda_2, \lambda_3\}$. The classical replicator dynamics are

$$\dot{x}_1 = x_1 \left\{\lambda_1 x_1 (1 - x_1) - \lambda_2 x_2^2 - \lambda_3 x_3^2\right\}, \tag{58}$$

$$\dot{x}_2 = x_2 \left\{-\lambda_1 x_1^2 + \lambda_2 x_2 (1 - x_2) - \lambda_3 x_3^2\right\}, \tag{59}$$

with $x_3 = 1 - x_1 - x_2$. If $\lambda_1, \lambda_2, \lambda_3$ are positive, then the pure strategy equilibria e_1, e_2, e_3 are all asymptotically stable, and there is an interior equilibrium x^* with $x_i^* \propto \lambda_i^{-1}$, which is unstable.

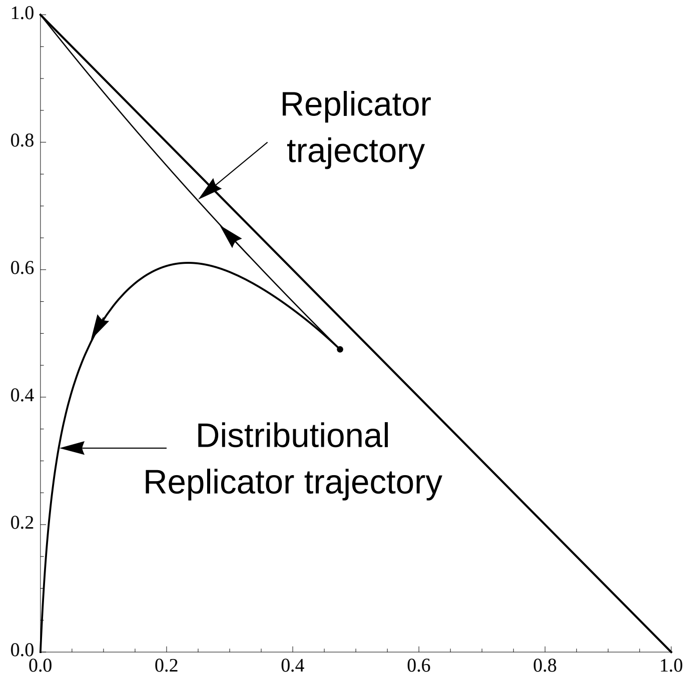

Figure 2: Trajectories of the replicator dynamics (58)-(59) (thin curve) and the mean (62) associated with the distributional replicator dynamics (60)-(61). Initial conditions for both trajectories are the same: $(x_1, x_2) = (\mu_{0,1}, \mu_{0,2})$. The parameters are: $(\lambda_1, \lambda_2, \lambda_3) = (1, 2, 15)$; $(\alpha_1, \alpha_2) = (\frac{1}{2}, \frac{1}{2})$; $(a_{11}, a_{12}, a_{13}) = (0.9, 0.05, 0.05)$; $(a_{21}, a_{22}, a_{23}) = (0.05, 0.9, 0.05)$.

The associated distributional replicator dynamics (40) are

$$\frac{dA_1}{dt} = \sum_{k=1}^{2} \alpha_k \frac{\lambda_1 a_{k1} e^{A_1} - \lambda_3 a_{k3}}{a_{k1} e^{A_1} + a_{k2} e^{A_2} + a_{k3}}, \tag{60}$$

$$\frac{dA_2}{dt} = \sum_{k=1}^{2} \alpha_k \frac{\lambda_2 a_{k2} e^{A_2} - \lambda_3 a_{k3}}{a_{k1} e^{A_1} + a_{k2} e^{A_2} + a_{k3}}, \tag{61}$$

with initial conditions $A_1(0) = A_2(0) = 0$. If $(A_1(t), A_2(t))$ is the solution trajectory of these equations, then from (43) the associated trajectory of the mean $\mu_t = \langle P_t \rangle$ is

$$\mu_{t,i} = \sum_{k=1}^{2} \alpha_k \frac{a_{ki} e^{A_i(t)}}{a_{k1} e^{A_1(t)} + a_{k2} e^{A_2(t)} + a_{k3}}, \qquad i = 1, 2. \tag{62}$$

We compare this trajectory with the trajectory of the replicator dynamics (58)-(59) with initial condition $x_0 = \mu_0$, and show that parameters can be chosen so that these two trajectories converge to different pure-strategy equilibria. An example is shown in Figure 2.

10 Application: A Price Dispersion game

We apply the results of Section 7 to analyze a model of pricing in which the unique equilibrium is a dispersed price equilibrium. This is a mixed strategy equilibrium in which different sellers charge different prices from consumers. We analyze this model in our evolutionary framework, allowing sellers to use mixed strategies over prices, and to use reinforcement learning to update their pricing strategies. The result we look for is whether this equilibrium is stable or not under the replicator continuity dynamic.[26]

The model we consider a simplified case of the finite dimensional Burdett and Judd (1983) price dispersion model analyzed in Lahkar (2010). There exists a large population of sellers, each selling the same homogeneous product to a large population of consumers. A strategy for a seller is to quote one of three prices p_0, p_1, p_2 with $0 < p_0 < p_1 < p_2$. In this simplified case, we exogenously restrict consumer behavior to two types. The first type of consumer picks a price quotation at random and buys the product at that price. The second type uses a more discerning strategy, paying a small cost to compare two random price quotations and then buying the product at the lower of the two prices, or choosing either one with equal probability if they coincide. We denote the proportion of the first type of consumer by y_1 and that of the second type by y_2 with $0 < y_1 < 1$ and $y_1 + y_2 = 1$.

An evolutionary analysis of this price dispersion model is credible for two reasons. First, there is a large number of sellers in such a model so that it can be analyzed as a population game. Second, sellers can be expected to behave in a myopic fashion since in the presence of a large number of competitors, it would be unrealistic to assume that they would possess the level of rationality and knowledge required to coordinate on the exact mixed equilibrium prediction. In particular, information about prices charged by competitors (even successful ones) may not be publicly available.

We first show that the game we have constructed is a positive definite game. We then identify conditions under which the model has a unique interior mixed strategy equilibrium. This is the dispersed price equilibrium. It then becomes a simple matter to apply Theorem 7.2 to argue that the dispersed price equilibrium is unstable under reinforcement learning.

If we denote by x_i the proportion of sellers charging price p_i, $i \in \{0, 1, 2\}$, then the expected payoff obtained by a seller charging price p_i is

$$\pi_i(x) = \frac{1}{2} p_i \left(y_1 + 2y_2 \left(\frac{x_i}{2} + \sum_{j>i} x_j \right) \right). \qquad (63)$$

[26]Price dispersion is a well documented fact. See Hopkins (2006) for a review of the evidence on price dispersion. Numerous theoretical models (for example; Varian, 1980; Burdett and Judd, 1983) explain this as a mixed strategy equilibrium. But empirical and experimental results in Lach (2002) and Cason, Friedman and Wagener (2005) respectively suggest dispersed price equilibria are unstable. This suggests we need to find some other explanation for observed price dispersion. Lahkar (2010) argues theoretically that price dispersion may exist as an evolutionary limit cycle.

Formally this is equivalent to a symmetric 2-player 3×3 normal form game between sellers with the following payoff matrix:

$$U = \begin{pmatrix} \frac{1}{2}p_2 & \frac{1}{2}p_2 y_1 & \frac{1}{2}p_2 y_1 \\ \frac{1}{2}p_1 y_1 + p_1 y_2 & \frac{1}{2}p_1 & \frac{1}{2}p_1 y_1 \\ \frac{1}{2}p_0 y_1 + p_0 y_2 & \frac{1}{2}p_0 y_1 + p_0 y_2 & \frac{1}{2}p_0 \end{pmatrix}. \tag{64}$$

Here, the rows correspond to bids of p_2, p_1, p_0 from top to bottom, and the columns to p_2, p_1, p_0 from left to right.

We now show that the game (64) is a positive definite game. For this purpose, we denote a typical element in the subspace $\mathbb{R}^3_0$ by $z = \{z_0, z_1, z_2\}$.

Lemma 10.1 *The game with payoff matrix U is a positive definite game.*

Proof. In order to establish positive definiteness, we need to show $z \cdot Uz > 0$, for all $z \in \mathbb{R}^3_0 \setminus \{0\}$. Writing $z_2 = -z_0 - z_1$ and $y_2 = 1 - y_1$, we obtain

$$\begin{aligned} z \cdot Uz &= (y_1 - 1)\Big(p_0 z_0^2 - p_2 (z_0 + z_1)^2 + p_1 z_1 (2z_0 + z_1)\Big) \\ &= (1 - y_1)\Big(z_0^2 (p_1 - p_0) + (z_0 + z_1)^2 (p_2 - p_1)\Big) > 0, \end{aligned}$$

since $0 < y_1 < 1$ and $p_0 < p_1 < p_2$. ■

Now suppose that the game (64) has a cyclic best response structure. That is, p_2 is a best-response to p_0, p_1 is a best response to p_2, and p_0 is a best-response to p_1. This is the case if

$$p_2 y_1 > p_0, \tag{65}$$

$$p_1(y_1 + 2y_2) > p_2, \tag{66}$$

$$p_0(y_1 + 2y_2) > \max\{p_1, p_2 y_1\}. \tag{67}$$

We also assume that

$$p_1 \geq \frac{1}{2}(p_0 + p_2). \tag{68}$$

For example, if $(p_0, p_1, p_2) = (1, \frac{5}{4}, \frac{3}{2})$, then y_1 must lie in the range $\frac{2}{3} < y_1 < \frac{3}{4}$, and hence $\frac{1}{4} < y_2 < \frac{1}{3}$.

Under these conditions, the game has a unique mixed strategy equilibrium in which p_i is played with probability x_i^*, where $x^* = (x_0^*, x_1^*, x_2^*)$ is given by

$$x^* = \frac{U^{-1}e}{e \cdot U^{-1}e}. \tag{69}$$

For example, for $(p_0, p_1, p_2) = (1, \frac{5}{4}, \frac{3}{2})$ and $y_1 = \frac{1}{2}(\frac{2}{3} + \frac{3}{4}) = \frac{17}{24}$, we obtain

$$(x_0^*, x_1^*, x_2^*) = \frac{1}{91}\,(19, 29, 43)\,.$$

Observe that $x_0^* < x_1^* < x_2^*$, so that the highest price is offered with the highest frequency, and the lowest with the lowest frequency.

We now analyze the stability properties of this equilibrium under reinforcement learning. A naive approach would be to apply reinforcement learning directly to the payoff matrix in (64), obtain the replicator continuity equation and then apply Theorem 7.2. There is, however, a difficulty with this approach. The payoffs corresponding to each strategy profile in (64) are *expected* payoffs rather than *realized* payoffs. For example, under the strategy profile (p_1, p_2) offered by two sellers to a consumer, the realized payoff for the firm charging p_1 is either p_1 or 0 depending upon whether the sale materializes or not, whereas the expected payoff to this seller is $\frac{1}{2}p_1 y_1 + p_1 y_2$. Given that reinforcement learning is determined by realized payoffs to individuals, we cannot straightforwardly use (64) to derive the replicator continuity equation. Instead, we need to consider realized payoffs.

The realized payoff to a seller depends upon the type and action of the consumer with whom he is dealing. A random consumer may adopt one of the following strategies with their respective probabilities:

$U1$: Choose not to compare bids and award the bid to player 1:	probability $\frac{1}{2}y_1$
$U2$: Choose not to compare bids and award the bid to player 2:	probability $\frac{1}{2}y_1$
$I1$: Choose to compare bids and award bid to player 1 if there is a tie:	probability $\frac{1}{2}y_2$
$I2$: Choose to compare bids and award bid to player 2 if there is a tie:	probability $\frac{1}{2}y_2$

A choice of one of these pure strategies by the consumer determines a well-defined payoff matrix for the symmetric game between the two sellers. The four realized payoff matrices (for seller player 1) corresponding to these consumer choices are:

$$U^{U1} = \begin{pmatrix} p_2 & p_2 & p_2 \\ p_1 & p_1 & p_1 \\ p_0 & p_0 & p_0 \end{pmatrix}, \quad U^{U2} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$
$$U^{I1} = \begin{pmatrix} p_2 & 0 & 0 \\ p_1 & p_1 & 0 \\ p_0 & p_0 & p_0 \end{pmatrix}, \quad U^{I2} = \begin{pmatrix} 0 & 0 & 0 \\ p_1 & 0 & 0 \\ p_0 & p_0 & 0 \end{pmatrix}. \tag{70}$$

As in (64), the rows in each of these matrices correspond to player 1 bids of p_2, p_1, p_0 from top to bottom, and the columns to player 2 bids of p_2, p_1, p_0 from left to right. Given the possible behaviors of consumers, each of these payoff matrices is generated according to the probability distribution $\{z_{U1}, z_{U2}, z_{I1}, z_{I2}\} = \{\frac{1}{2}y_1, \frac{1}{2}y_1, \frac{1}{2}y_2, \frac{1}{2}y_2\}$, where z_α is the probability of the realization of the payoff matrix U^α.

We derive the continuity equation for this game with indeterminate payoffs. In doing so, we need to account for the probabilistic nature of the realization of the payoff matrices. This is easily done by incorporating the probability distribution z in the updating equation of the probability measure P. If the consumer uses strategy α and the sellers play pure strategies (i, j), then the updating rule (5) for player 1 is generalized to

$$x' = x + \eta f_{ij}^{\alpha}(x). \tag{71}$$

We now proceed exactly as in section 2 to obtain the (continuous time) weak form of the continuity equation

$$\frac{d}{dt}\langle \phi \mid P_t \rangle = \int_{x \in \Delta} \nabla \phi(x) \cdot [\mathcal{F}(x)\langle P_t \rangle]\, P_t(dx), \tag{72}$$

where now the operator $\mathcal{F}(x)$ of (11) is generalized to

$$\mathcal{F}_{ij}(x) = \sum_{\alpha} \sum_{r \in \mathbf{n}} x_r f_{rj,i}^{\alpha}(x) z_{\alpha}. \tag{73}$$

This has the same form as the continuity equation (10).

In particular, the Börgers and Sarin (1997) updating rule for an individual seller involved in a game of type α, has the form

$$\begin{aligned} x'_r &= \delta_{ir} u_{ij}^{\alpha} \eta + (1 - u_{ij}^{\alpha} \eta) x_r, & (74) \\ f_{ij,r}^{\alpha}(x) &= (\delta_{ir} - x_r) u_{ij}^{\alpha}. & (75) \end{aligned}$$

Under this rule,[27] we have from (73)

$$\begin{aligned} [\mathcal{F}(x)y]_r &= \sum_j \mathcal{F}_{rj}(x) y_j \\ &= \sum_j \sum_{i,\alpha} x_i f_{ij,r}^{\alpha}(x) y_j z_{\alpha} \\ &= \sum_{i,j,\alpha} (\delta_{ir} - x_r) u_{ij}^{\alpha} x_i y_j z_{\alpha} \\ &= x_r \left(\sum_{j,\alpha} u_{rj}^{\alpha} y_j z_{\alpha} - \sum_{i,j,\alpha} u_{ij}^{\alpha} x_i y_j z_{\alpha} \right) \\ &= x_r \left([Uy]_r - x \cdot Uy \right) \qquad \text{where } U = \sum_{\alpha} U^{\alpha} z_{\alpha} \\ &= R_r(x) y. \end{aligned} \tag{76}$$

[27]We are using the payoff matrices in (70) to define the updating equations in (74). This, however, does not mean we are assuming a random matching structure in this game; i.e. we are not assuming that a seller is aware of the strategy being used by his consumer or of the identity of his rival in case the consumer is comparing two bids. All that a player needs to know to apply (74) is his own realized payoff which is his quoted price if it results in a sale or zero if there is no sale. The probabilities with which these payoffs occur are accounted for in the continuity equation (72).

Applying (76) to (72), we obtain the replicator continuity equation

$$\frac{d}{dt}\langle \phi \mid P_t \rangle = \int_{x\in\Delta} \nabla\phi(x) \cdot [R(x)\langle P_t\rangle]\, P_t(dx). \tag{77}$$

Note that despite the uncertainty in the realization of any particular payoff matrix in (70), the continuity dynamic we obtain in (77) is exactly the same as that obtained in (17) for the case with a deterministic payoff matrix. Hence, we can readily apply the general results obtained in the previous sections to analyze the stability properties of the equilibrium in our price dispersion model.

Proposition 10.2 *Consider the game given by payoff function (63), or equivalently by the expected payoff matrix U given by (64). Suppose the game has a unique interior equilibrium given by (69). Then this dispersed price equilibrium is totally unstable under the replicator continuity dynamic.*

Proof. First, we note that the expected payoff matrix U in (64) is equal to $\sum_\alpha U^\alpha z^\alpha$. It therefore follows from (76) that (77) is also the replicator continuity equation associated with the payoff matrix U. Hence, we obtain the desired result from Theorem 7.2 if we show that U is the payoff matrix of a positive definite game. But this follows from Lemma 10.1. ■

Now consider the three pure strategies in this game. That is, a seller is primed with a fixed price that he asks if chosen to bid. If all sellers use such a pure strategy, then the population distribution is a mass point at that pure strategy. These pure strategy mass point distributions are rest points of the continuity replicator dynamics (Example 3.3.1). We show in Appendix A.5 that if the cyclic best-response conditions (65)-(67) hold, then these are the only boundary rest points, and they are all unstable.

It now follows that all equilibria are unstable, and the solution trajectory ξ_t converges onto a heteroclinic cycle at the boundary of the simplex. This is illustrated in Fig 3 for an initial probability distribution which is a mixture of three types of seller

$$P_0 = \alpha_1\delta_{a_1} + \alpha_2\delta_{a_2} + \alpha_3\delta_{a_3}, \tag{78}$$

where $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ and $a_i = (a_{i0}, a_{i1}, a_{i2}) \in \operatorname{int}\Delta$ for $i = 1, 2, 3$. Note that a type-i seller uses the mixed strategy a_i over prices (p_0, p_1, p_2). The proportion of sellers of type i in the population is α_i. Also note that it follows from (37) that the distributional trajectory $\{P_t\}$ in this example is given by

$$\langle \phi \mid P_t \rangle = \sum_{i=1}^{3} \alpha_i \phi\left(\frac{\xi_t a_i}{\xi_t \cdot a_i}\right), \tag{79}$$

for any continuous function ϕ on Δ, where ξ_t is the trajectory of the distributional replicator dynamics illustrated in Fig 3.

Results of a similar nature on the instability of dispersed price equilibria have been obtained in Hopkins and Seymour (2002) (using the replicator dynamic) and Lahkar (2010) (using the logit

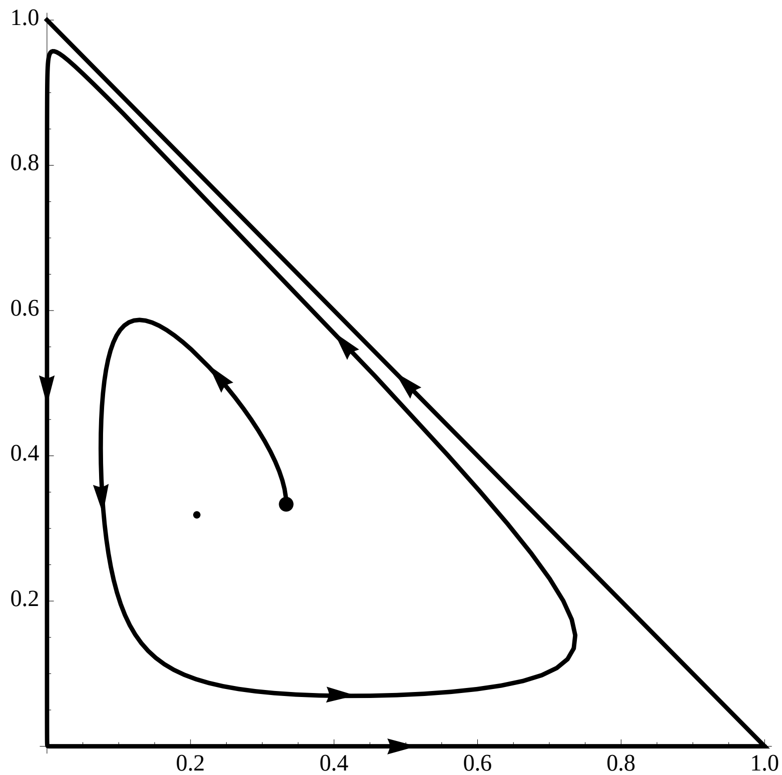

Figure 3: The trajectory ξ_t of the distributional replicator dynamic (44) for the price-dispersion game with expected payoff matrix (64) and initial probability distribution (78). The initial condition (heavy dot) is $\xi_0 = (1/3)e$. The unstable dispersed price equilibrium x^* is indicated by the small dot. The game parameters are: $(p_0, p_1, p_2) = \frac{1}{4}(4, 5, 6), (y_1, y_2) = \frac{1}{24}(17, 7)$ and $x^* = \frac{1}{91}(19, 29, 43)$. The seller distribution parameters are: $(\alpha_1, \alpha_2, \alpha_3) = \frac{1}{30}(13, 13, 4)$, and $a_1 = (0.1, 0.45, 0.45), a_2 = (0.45, 0.1, 0.45), a_3 = (0.45, 0.45, 0.1)$.

dynamic, Fudenberg and Levine, 1998). However, the microfoundations we use in deriving our result are very different from those used in the other papers. In motivating our work on the replicator continuity dynamic, we have expressed our reservations about the feasibility of the revision protocols that generate the replicator or the logit dynamic. Even in the simplest imitative revision protocol that generates the replicator dynamic, a seller would need to observe the price quotation observed by some rival seller. On the other hand, under reinforcement learning as we have applied it here, a seller need be guided only by his own personal experience without even requiring to observe the price being charged by the seller with whom he is currently matched. The informational requirements of our strategy revision procedure are therefore far less onerous. In many situations such a procedure may be more realistic since agents are more likely to be guided by their personal histories than by any information about the wider social state, even when information about the social state is available.

11 Discussion and Conclusion

The motivation behind this paper was to provide rigorous, learning-based foundations for evolutionary game theory that allow agents in large populations to practice a wider range of behaviors based upon their individual histories and experience. Specifically, we have focused on reinforcement learning, as formally developed by Cross (1973) and Börgers and Sarin (1997), and applied these learning procedures in the context of population games. A byproduct of this approach is that it allows experience based, in contrast to the more usual observation based, strategy revision models into evolutionary game theory, permitting agents to exhibit adaptive behaviors even when they may not possess any knowledge of wider social characteristics. Further, as shown in Börgers et. al. (2004), the Cross learning rule is the archetypical representative of a wider class of learning rules that satisfy conditions that these authors call *absolute expediency* and *monotonicity*. These conditions provide criteria to judge individual behavioral norms in evolutionary models on the basis of first principles. The theory developed here, therefore, can be seen as a first step towards the systematic extension of the framework of Börgers et. al. (2004) to large population games.

In our model, players from a large population of agents, randomly matched in each time period to play a two player symmetric game, apply the Cross learning rule to update their mixed strategies. This leads to the replicator continuity equation (Section 3) that traces the evolution of the probability measure over the set of mixed strategies, the population state in our model. This is a particular example of the general class of continuity dynamics (Section 2) that are generated when we allow agents to play mixed strategies instead of confining them to pure strategies as in classical evolutionary models. The replicator continuity equation cannot be solved explicitly (any more than can the classical replicator dynamic). But we have proposed a general solution method using Liouville's formula and an associated finite-dimensional, autonomous ODE system that we call the *distributional replicator dynamics*, which can be applied to any finite normal form game.

We have shown in Section 3.3 that the replicator continuity dynamics admits a large class of stationary solutions; in particular, any probability distribution whose mean over the space of mixed strategies is a Nash equilibrium (Example 3.3.2). This implies that equilibrium populations can be very heterogeneous, in that different players can play very different mixed strategies. However, the population is a 'mixture' of mixed strategy players, rather than of pure strategy players, as in the classical case. This constitutes a much richer behavioral structure than is usually considered. In particular, 'rationality' exists only at the aggregate mean level, with individual agents possibly exhibiting inconsistent choices even at equilibrium.

We have shown that such a rich equilibrium can arise for negative-definite symmetric games, and that the unique Nash equilibrium for such a game is globally attracting for the population mean. However, although the equilibrium population mean is fixed, the equilibrium *distribution* depends on the initial population distribution over agents' behavioral dispositions (mixed strategies), and thus is a function of 'history'. We have also shown that the population mean converges globally to a set of Nash equilibria in doubly symmetric games. In contrast, the unique interior equilibrium for a positive-definite game is always totally unstable (i.e. unstable in every direction). Such results on convergence and non-convergence in negative and positive definite games and doubly symmetric games are, of course, standard in both learning and evolutionary game theory. It is, however, significant that such results are obtained even when we have integrated the two approaches here. We have obtained these results by requiring agents to know only their own personal history in previous rounds of the game. This greatly expands the scope of evolutionary game theory from its traditional focus on revision procedures that are functions of wider social characteristics. This raises the interesting possibility that other well known results from the learning literature may also be obtained in the context of population games. For example, it would be interesting to examine whether, under the Hart and Mas-Colell (2000) regret matching rule, regrets would be eliminated for all agents in the population as they are in learning in finite player games

By allowing agents to employ different mixed strategies, we have also been able to analyze the effect of heterogeneity of agent behavior in the population. In the classical approach to evolutionary game theory, agent behavior is homogeneous since all agents play the same mixed strategy at any given time (equivalently, agents play only pure strategies so that there is a given mixture of pure strategies in the population). We have focused on situations in which results from our distributional theory differ markedly from corresponding results for the standard replicator dynamics. In particular, for 2×2 games with alternative pure strategy ESS, we have shown that convergence results for the mean state under the continuity replicator dynamics follow those of the classical replicator dynamics with equivalent initial conditions. However, this need not be true for $n \times n$ games with $n > 2$, and we have constructed a 3×3 example in which the replicator continuity equations lead to very different predictions about the observed long-run social state from those of the classical replicator dynamic, even though the two dynamics have the same (mean) initial condition (see Figure 3). This example illustrates the impact of initial agent heterogeneity on the long run social state. For the classical replicator dynamics, such initial heterogeneity takes the form of a specified population mixture of pure strategies — or, alternatively, a single mixed strategy used by all agents. In contrast, in the continuity case, the initial condition can represent a mixture of mixed strategies, incorporating different subpopulations using different mixed strategies. This shows that local stability properties of (in these cases, pure) ESS equilibria can be quite different in the distributional context.

Finally, we have analyzed a simple model of price dispersion and concluded that the dispersed price equilibria in this model is unstable under the replicator continuity equation. This follows from our theorem on the instability of the interior equilibrium in positive definite games. However, we have noted that the microfoundations of this result are quite different from similar results obtained by Hopkins and Seymour (2002) and Lahkar (2010). Our result shows that the instability of dispersed equilibria holds even in this new scenario where a seller cannot observe the behaviour of rivals and needs to rely on his personal experience of different pricing strategies.

There are many directions in which our general continuity equation approach to population games could be taken. For example, the analysis could be extended to consider a larger class of reinforcement learning rules, such as those discussed in Börgers et al (2004). It should also be possible to analyze mixed strategy evolution in other types of player-matching schemes than the simple pairwise-matching scheme discussed here. In this paper, a player interacts with a potentially different partner in each round of the game. However, the theory has a straightforward extension to the case in which some fixed proportion of agents are matched in each round. Alternatively, one may fix the population into matched pairs of players at the beginning, and allow these pairs to interact repeatedly using some learning protocol. The change in the distribution of mixed strategies in the populations can then be studied using a continuity equation. Or one can consider a more realistic scenario of a combination of the two matching schemes — where players play with a fixed partner for some number of periods and then change partners. Such problems can form a substantial research agenda for the future.

A Appendix

A.1 The weak form of continuity equations

We work with probability measures defined on the Borel sets in Δ. Let P_t be a probability measure at time $t \geq 0$ for the population. As discussed in section 2, the updating equation (9) is

$$\langle \phi \mid P_{t+\tau} \rangle = \sum_{i,j \in \mathbf{n}} \int_{x \in \Delta} \phi(x + \eta f_{ij}(x)) x_i P_t\,(dx)\,\langle P_t \rangle_j,$$

for any real-valued differentiable function ϕ on Δ, and $\eta = \eta(\tau)$ satisfies conditions (i), (ii) and (iii) given in section 2.1. Now Taylor expand the $\phi(\cdot)$ term up to terms of order η to obtain

$$\langle \phi \mid P_{t+\tau} \rangle = \sum_{i,j \in \mathbf{n}} \int_{x \in \Delta} \{\phi(x) + \eta \nabla \phi(x) \cdot f_{ij}(x)\}\, x_i P_t\,(dx)\,\langle P_t \rangle_j + \mathcal{O}\left[\eta^2\right].$$

Noting that $\sum_{i,j} x_i \langle P_t \rangle_j = 1$, and using (3), this can be written in the form:

$$
\begin{aligned}
&\frac{1}{\tau}\left\{\int_{x\in\Delta}\phi(x)P_{t+\tau}(dx) - \int_{x\in\Delta}\phi(x)P_t(dx)\right\} \\
&\qquad = \frac{\eta(\tau)}{\tau}\sum_{i,j\in\mathbf{n}}\int_{x\in\Delta}\nabla\phi(x)\cdot f_{ij}(x)x_i P_t(dx)\,\langle P_t\rangle_j + \mathcal{O}\left[\left(\frac{\eta(\tau)}{\tau}\right)\eta(\tau)\right] \\
&\qquad = \frac{\eta(\tau)}{\tau}\int_{x\in\Delta}\nabla\phi(x)\cdot\left\{\sum_{i,j\in\mathbf{n}} x_i f_{ij}(x)\langle P_t\rangle_j\right\}P_t(dx) + \mathcal{O}\left[\left(\frac{\eta(\tau)}{\tau}\right)\eta(\tau)\right] \\
&\qquad = \frac{\eta(\tau)}{\tau}\int_{x\in\Delta}\nabla\phi(x)\cdot\left[\mathcal{F}(x)\langle P_t\rangle\right]P_t(dx) + \mathcal{O}\left[\left(\frac{\eta(\tau)}{\tau}\right)\eta(\tau)\right]
\end{aligned}
$$

where $\mathcal{F}(x)$ is given by (11). Since $\eta(\tau) \to 0$, and $\eta(\tau)/\tau \to \eta'(0) = 1$ as $\tau \to 0$ (see section 2.1), taking the limit as $\tau \to 0$ now gives:

$$\frac{d}{dt}\langle \phi \mid P_t \rangle = \int_{x\in\Delta}\nabla\phi(x)\cdot\left[\mathcal{F}(x)\langle P_t\rangle\right]P_t(dx)\,. \tag{80}$$

This is the *weak form of the continuity equation* for Borel probability measures (10), which exists provided the integral on the right exists for all $t \geq 0$. This is the case if, for example, the forward state change vectors, $f_{ij}(x)$, are continuous in x, since then $\mathcal{F}(x)$, given by (11), is also continuous, and hence bounded on Δ. Since $\nabla\phi(x)$ is continuous, and hence bounded, and P_t is a probability measure, it follows that the integral always exists. This shows that $\langle \phi \mid P_t \rangle$ is differentiable in t, with time-derivative given by (80).

A.2 Proof of Liouville's formula

We are required to solve a weak continuity equation of the form

$$\frac{d}{dt}\langle \phi \mid P_t \rangle = \int_{x\in\Omega}\nabla\phi(x)\cdot X(x,t)P_t(dx)\,, \tag{81}$$

where $X(x,t)$ is a smooth, time-dependent vector field on a compact, regular domain $\Omega \subset I\!R^n$, a domain which is invariant under the flow defined by X, and we are given an initial probability measure P_0 at time $t = 0$. Let $x_{t_0,t}(x)$ be the solutions trajectory of the characteristic system $\dot{x} = X(x,t)$, that passes through x at time t_0.

Consider the generalized function γ_t defined by:

$$\gamma_t(\phi) = \int_\Omega \phi\left(x_{0,t}(x)\right)P_0(dx)\,. \tag{82}$$

Then

$$\begin{aligned}\frac{d\gamma_t}{dt}(\phi) &= \frac{d}{dt}\int_\Omega \phi\left(x_{0,t}(x)\right) P_0\left(dx\right) \\ &= \int_\Omega \nabla\phi(x_{0,t}(x)) \cdot \dot{x}_{0,t}(x) P_0\left(dx\right) \\ &= \int_\Omega \nabla\phi(x_{0,t}(x)) \cdot X\left(x_{0,t}(x), t\right) P_0\left(dx\right).\end{aligned}$$

Now apply the smooth change of variables $\xi = x_{0,t}(x)$, which has inverse $x = x_{t,0}(\xi)$, since Ω is invariant under the flow defined by X. Then:

$$\frac{d\gamma_t}{dt}(\phi) = \int_\Omega \nabla\phi(\xi) \cdot X\left(\xi, t\right) P_t\left(d\xi\right), \tag{83}$$

where P_t is the measure defined by[28]

$$P_t(B) = P_0\left(x_{t,0}(B)\right). \tag{84}$$

Applying the same change of variable to (82), we also have:

$$\gamma_t(\phi) = \int_\Omega \phi(\xi) P_t(d\xi) = \langle \phi \mid P_t \rangle,$$

and hence from (83)

$$\frac{d}{dt}\langle \phi \mid P_t \rangle = \int_\Omega \nabla\phi(\xi) \cdot X(\xi, t) P_t(d\xi).$$

This shows that P_t given by (84) is the solution of the weak form of the continuity equation (81) with the given initial measure P_0. Equation (84) is a measure-theoretic form of Liouville's formula. We also obtain expected values of smooth test functions:

$$\langle \phi \mid P_t \rangle = \int_\Omega \phi(\xi) P_t(d\xi) = \int_\Omega \phi\left(x_{0,t}(x)\right) P_0\left(dx\right). \tag{85}$$

This yields the formula (30), and hence proves Proposition 4.1.

A.2.1 Absolute continuity

Now suppose that P_0 is absolutely continuous with respect to Lebesgue measure. That is, there is a Lebesgue-integrable density function $p_0(x)$ such that $P_0(dx) = p_0(x)dx$. Then it follows from (84) that $P_t(d\xi) = P_0\left(x_{t,0}(d\xi)\right) = p_0\left(x_{t,0}(\xi)\right) dx_{t,0}(\xi)$. We also have $dx = dx_{t,0}(\xi) = |J_t(x;\xi)| d\xi$, where $J_t(x;\xi)$ is the Jacobian matrix:

$$J_t(x;\xi) = \det\left(\frac{\partial x_i}{\partial \xi_j}\right).$$

[28]See Dunford and Schwartz (1964), Lemma 8, p 182, for the 'change of variable' result used here.

To compute this Jacobian, consider the generalized Jacobian

$$J_{t,s}(\xi) = \det\left(\tfrac{\partial x_{t,s}(\xi)_i}{\partial \xi_j}\right).$$

Then $J_t(x;\xi) = J_{t,0}(\xi)$, and $J_{t,t}(\xi) = 1$. Next, observe that, by definition of the trajectories $x_{t,s}(\xi)$, we have

$$\frac{d}{ds}\left[\frac{\partial x_{t,s}(\xi)_i}{\partial \xi_j}\right] = \frac{\partial}{\partial \xi_j}\left[\frac{dx_{t,s}(\xi)_i}{ds}\right] = \frac{\partial}{\partial \xi_j}\left[X_i\left(x_{t,s}(\xi), s\right)\right] = \sum_{k=1}^{n} \frac{\partial X_i}{\partial x_k}\left(x_{t,s}(\xi), s\right)\frac{\partial x_{t,s}(\xi)_k}{\partial \xi_j}. \tag{86}$$

Let $J^{(i)}_{t,s}(\xi)$ be the determinant of the matrix obtained from $J_{t,s}(\xi)$ by taking the time derivatives with respect to s of the entries in the i-th row, as in (86), but leaving the other rows unchanged. Let $\left[J_{t,s}(\xi)\right]_{i,j}$ be the ij-th minor of $J_{t,s}(\xi)$.[29] Then:

$$\begin{aligned}
\frac{dJ_{t,s}(\xi)}{ds} &= \sum_{i=1}^{n} J^{(i)}_{t,s}(\xi) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}(-1)^{i+j}\frac{d}{ds}\left[\frac{\partial x_{t,s}(\xi)_i}{\partial \xi_j}\right]\left[J_{t,s}(\xi)\right]_{i,j} && \text{expanding } J^{(i)}_{t,s}(\xi) \text{ by the } i\text{-th row} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}(-1)^{i+j}\frac{\partial X_i}{\partial x_k}\left(x_{t,s}(\xi), s\right)\frac{\partial x_{t,s}(\xi)_k}{\partial \xi_j}\left[J_{t,s}(\xi)\right]_{i,j} && \text{using (86)} \\
&= \sum_{i=1}^{n}\sum_{k=1}^{n}(-1)^{i+k}\frac{\partial X_i}{\partial x_k}\left(x_{t,s}(\xi), s\right)\left\{\sum_{j=1}^{n}(-1)^{k+j}\frac{\partial x_{t,s}(\xi)_k}{\partial \xi_j}\left[J_{t,s}(\xi)\right]_{i,j}\right\} \\
&= \sum_{i=1}^{n}\sum_{k=1}^{n}(-1)^{i+k}\frac{\partial X_i}{\partial x_k}\left(x_{t,s}(\xi), s\right)\delta_{ik}J_{t,s}(\xi).
\end{aligned}$$

The last equality holds because, for $k \neq i$, the expression in {} is the determinant of an $n \times n$ matrix whose i-th and k-th rows are identical, and hence is zero. We therefore have:

$$\frac{dJ_{t,s}(\xi)}{ds} = J_{t,s}(\xi)\sum_{i=1}^{n}\frac{\partial X_i}{\partial x_i}\left(x_{t,s}(\xi), s\right) = J_{t,s}(\xi)\left[\nabla \cdot X\right]\left(x_{t,s}(\xi), s\right).$$

Integrating this from $s = 0$ to $s = t$ and recalling that $J_{t,t}(\xi) = 1$ and $J_{t,0}(\xi) = J_t(x;\xi)$, gives:

$$|J_t(x;\xi)| = \exp\left\{-\int_0^t \left[\nabla \cdot X\right]\left(x_{t,s}(\xi), s\right) ds\right\}.$$

It now follows that P_t is absolutely continuous with respect to Lebesgue measure, with associated

[29] That is, the determinant of the $(n-1)\times(n-1)$-matrix obtained from $J_{t,s}(\xi)$ by deleting the i-th row and the j-th column.

density function $p_t(\xi) = p_0(x_{t,0}(\xi))\,|J_t(x;\xi)|$. That is:

$$p_t(\xi) = p_0(x_{t,0}(\xi)) \exp\left\{-\int_0^t [\nabla \cdot X](x_{t,s}(\xi), s)\, ds\right\}.$$

This yields the probability-density function form of Liouville's formula (28).

A.3 Proof of Proposition 5.2

For the pseudo-replicator vector field $X(x,t) = R(x)y(t)$ on the simplex $\Delta \subset I\!R^n$ defined in (31), we have $\sum_{i=1}^{n_1} x_i = 1$ and $\sum_{i=1}^{n} X_i = 0$. Hence, the independent components are x_i and X_i for $1 \leq i \leq n-1$. We therefore take the state space to be the projection of Δ into $I\!R^{n-1}$ defined by:

$$\Omega_1 = \left\{(x_1, \ldots, x_{n-1}) \in I\!R^{n-1} : 0 \leq x_i \leq \sum_{i=1}^{n-1} x_i \leq 1\right\}. \tag{87}$$

Then, if $(x_1, \ldots, x_{n-1}) \in \Omega$, the associated point $x \in \Delta$ is $x = (x_1, \ldots, x_{n-1}, x_n)$ with $x_n = 1 - \sum_{i=1}^{n-1} x_i$. Generally x denotes a point in Δ, but relevant operations often involve only the independent components, i.e. the associated point in Ω.

Let $L_{ij}(x) = x_i(\delta_{ij} - x_j)$. Then, from (32) we can write the divergence of X on Ω as:

$$\nabla \cdot X(x,t) = \sum_{i=1}^{n-1}\left\{\frac{\partial}{\partial x_i} - \frac{\partial}{\partial x_n}\right\} X_i(x,t) = \sum_{i=1}^{n-1}\sum_{j=1}^{n}\left\{\frac{\partial}{\partial x_i} - \frac{\partial}{\partial x_n}\right\} L_{ij}(x)c_j(t),$$

for $x \in \Delta_1$. Also, for $1 \leq i, j \leq n-1$:

$$\begin{aligned} \frac{\partial}{\partial x_i}[L_{ij}(x)] &= (1 - x_i)\delta_{ij} - x_j, & \frac{\partial}{\partial x_i}[L_{in}(x)] &= -x_n, \\ \frac{\partial}{\partial x_n}[L_{ij}(x)] &= 0, & \frac{\partial}{\partial x_n}[L_{in}(x)] &= -x_i. \end{aligned}$$

Hence,

$$\begin{aligned} \nabla \cdot X(x,t) &= \sum_{i,j=1}^{n-1}\{(1 - x_i)\delta_{ij} - x_j\}\, c_j(t) + \sum_{i=1}^{n-1}(x_i - x_n)c_n(t) \\ &= \sum_{i=1}^{n} c_i(t) - n\sum_{i=1}^{n} x_i c_i(t) \\ &= \{e - nx\} \cdot c(t). \end{aligned}$$

It now follows that, if $x_{t,s}(x)$ are the solution trajectories of the pseudo-replicator equations

(34), then we obtain

$$[\nabla \cdot X]\left(x_{t,s}(x), s\right) = \{e - n x_{t,s}(x)\} \cdot c(s) = e \cdot c(s) - n \sum_{i=1}^{n} \frac{x_i c_i(s) e^{C_i(s,t)}}{x \cdot e^{C(s,t)}},$$

where $C(s,t) = C(s) - C(t)$. Thus

$$\begin{aligned}
\int_0^t [\nabla \cdot X]\left(x_{t,s}(x), s\right) ds &= e \cdot \int_0^t c(s) ds - n \sum_{i=1}^{n} \int_0^t \frac{x_i e^{C_i(s,t)}}{x \cdot e^{C(s,t)}} c_i(s) ds \\
&= e \cdot C(t) - n \int_0^t \frac{d}{ds}\left[\ln\left(x \cdot e^{C(s,t)}\right)\right] ds \\
&= e \cdot C(t) + n \ln\left[x \cdot e^{-C(t)}\right],
\end{aligned}$$

because $C(t,t) = 0$, $C(0,t) = -C(t)$ and $e \cdot x = 1$. We therefore have:

$$\exp\left\{-\int_0^t [\nabla \cdot X]\left(x_{t,s}(x), s\right) ds\right\} = \left(\frac{1}{x \cdot e^{-C(t)}}\right)^n \exp\left\{-e \cdot C(t)\right\}.$$

Substituting in Liouville's formula (28), it now follows that the solution of the weak continuity equation for density functions associated to a pseudo-replicator vector field (32) is given by (36). ■

A.4 Proofs for Section 7

A.4.1 Properties of F

We begin by proving some key properties of the function F defined in (42) .

First observe that the function $F(\xi \mid P_0)$ is well-defined for all $\xi \in \Delta$ if $\partial\Delta$ has P_0-measure zero, because the integral in (42) can be taken over $\operatorname{int}\Delta$, and $\xi \cdot x$ is never zero for any $\xi \in \Delta$ and $x \in \operatorname{int}\Delta$. In particular, this condition rules out initial distributions of the form (24). The formulation (44) therefore extends the distributional replicator dynamics to a dynamical system having compact phase space, namely the simplex Δ.[30]

In order to state the key properties of the function $F(\cdot \mid P_0) : \Delta \to \Delta$ that we shall need, we require some notation.

Let $S \subseteq \mathbf{n}$ be a (possibly empty) set of pure strategies. Define a subspace $I\!R^n_S = \{x \in I\!R^n \mid x_j = 0 \text{ for } j \notin S\}$, and let $e_S = \sum_{i \in S} e_i \in I\!R^n_S$.[31] We note that $I\!R^n$ may be decomposed as

$$I\!R^n = I\!R^n_S \oplus I\!R^n_{S'}, \tag{88}$$

[30] We note that the form (44) defines an example of a larger class of dynamics: *Positive Definite Adaptive* (PDA) dynamics with non-linear payoff function $\Pi(\xi) = UF(\xi \mid P_0)$. A general discussion of this class of dynamics is given in Hopkins and Seymour (2002, section 3).

[31] We take $I\!R^n_\emptyset = \{0\}$ and $e_\emptyset = 0$.

where $S' = \mathbf{n} \setminus S$ is the complement of S in $\mathbf{n}$. That is, $I\!R^n_{S'} = \{x \in I\!R^n \mid x_i = 0 \text{ for } i \in S\}$. In addition, $I\!R^n_S$ may be decomposed as

$$I\!R^n_S = [e_S] \oplus I\!R^n_{S0}, \tag{89}$$

where $[e_S]$ is the 1-dimensional subspace generated by e_S, and $I\!R^n_{S0} = \{x \in I\!R^n_S \mid \sum_i x_i = 0\}$.

Recall that an $n \times n$ real matrix A is said to be *positive (resp. negative) definite* on a subspace $\Sigma \subseteq I\!R^n$ if $z \cdot Az > 0$ (resp. < 0) for all non-zero $z \in \Sigma$. We can now state the key properties of $F(\cdot \mid P_0)$ that we need.

Lemma A.1 *Suppose $\partial\Delta$ has zero P_0-measure. Then the derivative $\tilde{D}F(\xi \mid P_0) = \Big(\partial F_i(\xi \mid P_0)/\partial \ln \xi_j\Big)$ is symmetric. If $S = supp(\xi)$, then the j-th row and j-th column of $\tilde{D}F(\xi)$ are zero for $j \notin S$, and $e_S \cdot \tilde{D}F(\xi) = \tilde{D}F(\xi)e_S = 0$. Further, $\tilde{D}F(\xi)$ is positive-definite on $I\!R^n_{S0}$.*

Proof. Let $\xi = e^\zeta$, and $\tilde{D}F(\xi) = \left(\frac{\partial F_i(\xi)}{\partial \zeta_j}\right)$. A formal calculation from (42) shows that

$$\frac{\partial F_i(\xi)}{\partial \zeta_j} = \int_\Delta \left(\frac{\xi_i x_i}{\xi \cdot x}\right) \left\{\delta_{ij} - \left(\frac{\xi_j x_j}{\xi \cdot x}\right)\right\} P_0(dx). \tag{90}$$

Clearly $\tilde{D}F$ is symmetric. Also, if $S = \text{supp}(\xi)$, then $F_j(\xi) = 0$ and the j-th row and j-th column of $\tilde{D}F$ defined by (90) are zero for $j \notin S$. However, since $\text{int}\,\Delta$ has positive P_0-measure, $\tilde{D}F$ has positive diagonal entries and negative off-diagonal entries for row and column indices $i, j \in S$. Further, from (90), we have $\tilde{D}Fe_S = e_S \cdot \tilde{D}F = 0$. Hence, $\tilde{D}F$ maps $I\!R^n$ onto $I\!R^n_{S0}$. A straightforward calculation from (90) now shows that, for $z \in I\!R^n_{S0}$,

$$z \cdot \tilde{D}F(\xi)z = \sum_{i \in S} F_i(\xi)(z_i - \bar{z})^2, \qquad \text{where } \bar{z} = z \cdot F(\xi),$$

and hence $\tilde{D}F(\xi)$ is positive-definite on $I\!R^n_{S0}$. ■

Lemma A.2 *Suppose $\partial\Delta$ has zero P_0-measure. Then $F(\cdot \mid P_0) : \Delta \to \Delta$ is a homeomorphism, and a diffeomorphism on int Δ. In particular, if $F(\xi \mid P_0) = u$, then $supp(\xi) = supp(u)$.*

Proof. For $\xi \in \Delta$, consider the dynamical system

$$\dot{\xi} = u - F(\xi \mid P_0) \in I\!R^n_0, \tag{91}$$

because $u, F \in \Delta$. Also,

$$\dot{\xi}_i\Big|_{\xi_i=0} = u_i - F_i(\xi \mid P_0)|_{\xi_i=0} = u_i \geq 0,$$

from which it follows that Δ is forward-invariant under the flow of the system (91). It now follows from standard results that Δ contains at least one equilibrium $\xi^* = \xi^*(u)$ of (91).[32] Further, it is clear from the definition of F in (42) that $\text{supp}(\xi^*) = \text{supp}(u)$.

[32]See, for example, Spanier (1966), Theorem 12, p197.

It remains to show that ξ^* is unique. Suppose that u, and hence ξ^* has full support. For $\xi \in I\!R^n_+$, let $\zeta = \ln \xi$, and consider the potential function

$$K(\zeta \mid P_0) = -u \cdot \zeta + \int_\Delta \ln(e^\zeta \cdot x) P_0(dx).$$

Then $\nabla K = -u + F$, and hence $\nabla K(\zeta^*) = 0$, where $\zeta^* = \ln \xi^*$. Further

$$[\nabla^2 K]_{ij} = \frac{\partial^2 K}{\partial \zeta_i \partial \zeta_j} = \frac{\partial F_i}{\partial \zeta_j}. \tag{92}$$

That is $\nabla^2 K = \tilde{D}F$, which is positive definite on $I\!R^n_0$ by Lemma A.1. Hence, ζ^* is the unique global minimum of K subject to the constraint $e \cdot e^\zeta = e \cdot \xi = 1$. Since (91) can be written as $\dot{\zeta} = -e^{-\zeta} \nabla K(\zeta)$, it follows that any equilibria must satisfy $\nabla K(\zeta) = 0$, and hence $\xi^* = e^{\zeta^*}$ is the unique equilibrium satisfying the constraint $\xi^* \in \Delta$.

Now suppose that u does not have full support. If $u = e_i$, then it is clear from the definition (42) that $\xi^* = e_i$ is the unique solution of $F = e_i$. So, we may suppose that the support of u contains at least two elements. If $S = \text{supp}(u)$, then (92) defines an $|S| \times |S|$ symmetric matrix, $\nabla^2 K_S$, by taking $i, j \in S$. The argument of lemma A.1 shows that this matrix is positive definite on $I\!R^n_{S0}$ (which has dimension at least 1), and hence $\xi^*_S = e^{\zeta^*_S} \in \Delta_S \subset I\!R^n_S$ is the unique solution of $F = u$.

Now observe that, for $\xi \in \text{int}\,\Delta$, the derivative $DF(\xi) = \big(\partial F_i / \partial \xi_j\big)$ satisfies: $DF(\xi) = \tilde{D}F(\xi) W(\xi)$, where $W(\xi) = \text{diag}\{\xi_1^{-1}, \ldots, \xi_n^{-1}\}$. Thus, by Lemma A.1, $DF(\xi) : T_\xi \Delta = I\!R^n_0 \to I\!R^n_0 = T_\xi \Delta$ is an isomorphism for all $\xi \in \text{int}\,\Delta$. That $F(\xi)$ is a diffeomorphism on $\text{int}\,\Delta$ now follows from the Inverse Function Theorem. Finally, since $F(\xi) : \Delta \to \Delta$ is a continuous bijection with continuous inverse on $\text{int}\,\Delta$, which is dense in Δ, it follows that the unique inverse $F^{-1}(x)$ is continuous, and hence that F is a homeomorphism on Δ. ■

Finally, we note the following observation concerning rest points.

Lemma A.3 *Suppose $\partial\Delta$ has zero P_0-measure. Let $\xi^* \in \Delta$ be a rest point of the distributional replicator dynamics (44), and let $x^* = F(\xi^* \mid P_0)$. Then x^* is a rest point of the relicator dynamics. Hence, any probability distribution P_∞ on Δ with mean x^* is a stationary distribution of the replicator continuity equation (17).*

Proof. If ξ^* is a rest point of (44) with support $S \subseteq \mathbf{n}$, then $R_i(\xi^*) F(\xi^* \mid P_0) = R_i(\xi^*) x^* = \xi_i \{e_i - \xi^*\} \cdot U x^* = 0$ for each i. In particular, $e_i \cdot U x^* = \xi^* \cdot U x^* = \pi^*$ (a constant) for each $i \in S$. Since $\text{supp}(x^*) = S$ by Lemma A.2, it follows that $e_i \cdot U x^* = x^* \cdot U x^* = \pi^*$ for each $i \in S$. Hence, $R_i(x^*) x^* = 0$. That is, x^* is a rest point of the replicator dynamics $\dot{x} = R(x) x$. That any probability distribution with mean x^* is a stationary distribution for the continuity equation (17) follows from the characterization rest points given in Example 3.3.2. ■

A.4.2 Proof of theorem 7.1

Let $\xi^* \in \Delta$ be the unique point satisfying $F(\xi^* \mid P_0) = x^*$ (Lemma A.2). For fixed P_0, we show the global dynamic stability of the equilibrium ξ^* under the dynamics (44) using the Lyapunov function

$$K(\xi \mid P_0) = -x^* \cdot \ln \xi + \int_\Delta \ln(\xi \cdot x)\, P_0(dx). \tag{93}$$

This is well-defined on the subset $\mathcal{S}(x^*) = \{\xi \in \Delta \mid \text{supp}(x^*) \subseteq \text{supp}(\xi)\}$[33]. One checks that x^* is a global minimum of K on Δ (cf. proof of lemma A.2). Then, for $\xi \neq \xi^*$,

$$\begin{aligned} \frac{dK}{dt} &= \sum_{i=1}^{n} \frac{1}{\xi_i} \{-x_i^* + F_i(\xi \mid P_0)\} \dot{\xi}_i \\ &= \sum_{i=1}^{n} \{-x_i^* + F_i(\xi \mid P_0)\} (e_i^1 - \xi) \cdot UF(\xi) \qquad \text{using (44)} \\ &= -(x^* - F(\xi)) \cdot UF(\xi) \\ &= (x^* - F(\xi)) \cdot U(x^* - F(\xi)) - (x^* - F(\xi)) \cdot Ux^*. \end{aligned}$$

The second term is non-negative since x^* is a Nash equilibrium, and the first term is negative if U is negative-definite, since $x^* - F(\xi) \in I\!R_0^n$. Clearly $\dot{K}(\xi^*) = 0$. Thus, $K(\xi) - K(\xi^*)$ is a global Lyapunov function, and it follows that any trajectory beginning at $\xi_0 \in \mathcal{S}(x^*)$ converges asymptotically to ξ^*. In particular, the trajectory beginning at the initial condition $\xi_0 = (1/n)e$. But, by (43) and construction of the distributional replicator dynamics (44), this trajectory satisfies $F(\xi_t \mid P_0) = \langle P_t \rangle$ for $t \geq 0$, and the result therefore follows.

If U is negative semi-definite, then $\dot{K}(\xi) \leq 0$, and the Lypunov stability of any Nash equilibrium x^* follows. ■

A.4.3 Mean entropy

A standard proof of stability of evolutionarily stable equilibria for the classical replicator dynamics uses the entropy function (e.g. Hofbauer and Sigmund, 1998, Chapter 7). For a Nash equilibrium of a symmetric game, x^*, this is defined on the subset $\{x \in \Delta \mid S(x^*) \subseteq S(x)\}$ by

$$L(x) = -\sum_i x_i^* \ln(x_i). \tag{94}$$

In the distributional case, consider the mean entropy: $\bar{L}_t = \langle L \mid P_t \rangle$. Then using (37), we obtain

$$\begin{aligned} \bar{L}_t &= -\sum_i x_i^* \int_\Delta \ln\left(\frac{\xi_i(t) x_i}{\xi(t) \cdot x}\right) P_0(dx) \\ &= \langle L \mid P_0 \rangle - x^* \cdot \ln(\xi_t) + \int_\Delta \ln(\xi_t \cdot x) P_0(dx). \end{aligned}$$

[33] As usual, we take $0 \ln 0 = 0$.

This provides a relationship between mean entropy and the Lyapunov function (93) used in the proof of theorem 7.1, namely:

$$\bar{L}_t - \bar{L}_0 = K(\xi_t \mid P_0) - K(\xi_0 \mid P_0), \tag{95}$$

where ξ_t is the trajectory of the distributional replicator dynamics (44) with initial condition $\xi_0 = (1/n)e$. In particular, it follows from the proof of Theorem 7.1 that if U is negative definite on $I\!R_0^n$, with unique equilibrium x^*, then mean entropy decreases along this trajectory.

A.4.4 Proof of theorem 7.2

Let $\xi^* \in \operatorname{int}\Delta$ be the unique equilibrium of the distributional replicator dynamics (44) satisfying $F(\xi^* \mid P_0) = x^*$ (Lemma A.2). We first show that ξ^* is a source node for the dynamics (44). Let $\pi^* = x^* \cdot Ux^*$ be the equilibrium payoff. We begin by considering the distributional replicator dynamics in the form (41). Introduce a new set of variables by setting $\Gamma(t) = C(t) - (\pi^* t)e$. Then the distributional replicator dynamics (41) can be written as

$$\frac{d\Gamma}{dt} = U\left\{-x^* + F(e^{\Gamma} \mid P_0)\right\}. \tag{96}$$

Note that $\xi = e^{C}/(e^{C}\cdot e) = e^{\Gamma}/(e^{\Gamma}\cdot e)$. It follows that $\Gamma^* = \ln \xi^*$ is an equilibrium of (96). In fact, (96) admits a 1-dimensional affine subspace of equilibria $\mathcal{E} = \{\Gamma^* + \alpha e \mid \alpha \in I\!R\}$[34]. We consider the stability of this equilibrium set.

The Jacobian matrix J for the system (96) is

$$J = U(D_\Gamma F), \qquad D_\Gamma F = \left(\frac{\partial F_i}{\partial \Gamma_j}\right), \tag{97}$$

where

$$\frac{\partial F_i}{\partial \Gamma_j} = \int_\Delta \frac{\partial}{\partial \Gamma_j}\left(\frac{x_i e^{\Gamma_i}}{x\cdot e^{\Gamma}}\right) P_0(dx) = \int_\Delta \left(\frac{x_i e^{\Gamma_i}}{x\cdot e^{\Gamma}}\right)\left\{\delta_{ij} - \left(\frac{x_j e^{\Gamma_j}}{x\cdot e^{\Gamma}}\right)\right\} P_0(dx).$$

Again, setting $\xi = e^{\Gamma}/(e^{\Gamma}\cdot e)$, we obtain $D_\Gamma F = \tilde{D}F(\xi \mid P_0)$, as defined in Lemma A.1[35]. Thus, since $\tilde{D}F(\xi^* \mid P_0)$ is symmetric and positive definite on $I\!R_0^n$ (by Lemma A.1), and U is positive definite on $I\!R_0^n$, it follows that $J(\xi^*)e = 0$ and the eigenvalues of $J(\xi^*)$ restricted to $I\!R_0^n$ all have positive real part[36]. This shows that the set $\mathcal{E}$ of equilibria of (96) is unstable.

It now follows that there exists a neighbourhood $\tilde{M}_0$ of $\mathcal{E}$ in $I\!R^n$ such that any trajectory Γ_t of (96) with initial condition $\Gamma_0 \in \tilde{M}_0 \setminus \mathcal{E}$ eventually leaves $\tilde{M}_0$.

Since $\mathcal{E}$ maps to ξ^* under the map $\Gamma \to \xi = e^{\Gamma}/(e^{\Gamma}\cdot e) \in \Delta$, it follows that this map projects $\tilde{M}_0$ onto a neighbourhood $\tilde{N}_0$ of ξ^* in $\operatorname{int}\Delta$. Hence, $\tilde{N}_0$ has the property that any trajectory ξ_t of the distributional replicator dynamics (44) with initial condition $\xi_0 \in \tilde{N}_0 \setminus \{\xi^*\}$ eventually leaves

[34] It is easy to see that Γ is an equilibrium of (96) if and only if $\Gamma = \Gamma^* + \alpha e$ for some constant α.

[35] See equation (90).

[36] See, for example, Hines (1980), pp 338-39.

$\tilde{N}_0$.

By Lemma A.2, we can now define a neighbourhood $N_0 = F(\tilde{N}_0 \mid P_0)$ of x^* in $\operatorname{int}\Delta$. From (43), $\langle P_t \rangle = F(\xi_t \mid P_0)$, where ξ_t is the trajectory of (44) with initial condition $\xi_0 = (1/n)e$. It follows that if $\langle P_{t_0} \rangle \in N_0$ for some $t_0 \geq 0$, then $\xi_{t_0} \in \tilde{N}_0$, and hence there is a $t_1 > t_0$ such that $\xi_{t_1} \notin \tilde{N}_0$. Hence, $\langle P_{t_1} \rangle = F(\xi_{t_1} \mid P_0) \notin N_0$. This proves the theorem. ■

A.4.5 Proof of theorem 7.3

Write $\mu_t = \langle P_t \rangle$. Since U is symmetric we have

$$\begin{aligned}
\frac{1}{2}\dot{\bar{w}}(\mu) &= \mu \cdot U\dot{\mu} \\
&= F(e^C \mid P_0) \cdot U\dot{F}(e^C \mid P_0) \\
&= F(e^C) \cdot UDF(e^C)\dot{C} \\
&= F(e^C) \cdot UDF(e^C)UF(e^C) \qquad \text{from (41)} \\
&= [(UF) \cdot DF\,(UF)]\,(\xi).
\end{aligned}$$

This is positive by Lemma A.1, provided $UF(\xi)$ has a non-zero component in $I\!R^n_{S0}$, where $S = \operatorname{supp}(\xi)$. From the decomposition (88) and (89), this is not the case if and only if $UF(\xi) = \pi^* e_S + v$ for some constant π^* and $v \in I\!R^n$ with $v_i = 0$ for $i \in S$, in which case $\dot{\bar{w}}(\mu) = 0$. If ξ^* is such a point, then $R(\xi^*)F(\xi^*) = 0$, and hence ξ^* is an equilibrium of the distributional dynamics (44). If $x^* = F(\xi^*)$, this implies that x^* is a rest point of the standard replicator dynamic, $R(x^*)x^* = 0$, since $\operatorname{supp}(x^*) = \operatorname{supp}(\xi^*)$.

Hence, $\dot{\bar{w}}(\mu) \geq 0$, for all μ with the equality holding only if μ is a rest point of the classical replicator dynamic. However, it is known that any local maximum of the mean payoff function of a doubly symmetric game is a Nash equilibrium (see Sandholm, 2009; Theorem 3.1.7). Hence, the mean population state converges to a level set of Nash equilibria along non-equilibrium trajectories. ■

A.5 Price dispersion: instability of pure equilibria

To prove the instability of the three pure equilibria in the price dispersion game of section 10 when the cyclic best-response conditions (65)-(67) hold, we consider the distributional replicator dynamics (44) restricted to each face of the ξ-simplex.

Case 1: $\xi_3 = 0$.

Then $\xi_1 = 1 - \xi_2$, and $F_3(\xi) = 0$, so that $F_1 = 1 - F_2$. Thus, the system reduces to the

1-dimensional system

$$\begin{aligned}\dot{\xi}_2 &= \xi_2\left(e_2 - \xi_2 e_2 - (1-\xi_2)e_1\right)\cdot U\left(F_2 e_2 + (1-F_2)e_1\right)\\ &= \xi_2(1-\xi_2)(e_2-e_1)\cdot U\left(F_2(e_2-e_1)+e_1\right)\\ &= \xi_2(1-\xi_2)\left\{F_2(e_2-e_1)\cdot U(e_2-e_1) + (e_2-e_1)\cdot Ue_1\right\}\end{aligned}$$

Thus, since U is positive definite, $\dot{\xi}_2 > 0$ for all $0 < F_2 < 1$ if and only if $(e_2 - e_1)\cdot Ue_1 \geq 0$. That is, if and only if $u_{21} - u_{11} \geq 0$. From (64) this condition is $p_1(y_1 + 2y_2) \geq p_2$. But this is always the case if (66) holds. This shows that e_1 is *unstable*.

Case 2: $\xi_2 = 0$.

Then $\xi_3 = 1 - \xi_1$, and $F_2(\xi) = 0$, so that $F_3 = 1 - F_1$. Thus, the system reduces to the 1-dimensional system

$$\begin{aligned}\dot{\xi}_1 &= \xi_1\left(e_1 - \xi_1 e_1 - (1-\xi_1)e_3\right)\cdot U\left(F_1 e_1 + (1-F_1)e_3\right)\\ &= \xi_1(1-\xi_1)(e_1-e_3)\cdot U\left(F_1(e_1-e_3)+e_3\right)\\ &= \xi_1(1-\xi_1)\left\{F_1(e_1-e_3)\cdot U(e_1-e_3) + (e_1-e_3)\cdot Ue_3\right\}\end{aligned}$$

Thus, since U is positive definite, $\dot{\xi}_1 > 0$ for all $0 < F_1 < 1$ if $(e_1 - e_3)\cdot Ue_3 \geq 0$. That is, if $u_{13} - u_{33} \geq 0$. From (64) this condition is $p_2 y_1 \geq p_0$. But this is always the case if (65) holds. This shows that e_3 is *unstable*.

Case 3: $\xi_1 = 0$.

Then $\xi_2 = 1 - \xi_3$, and $F_1(\xi) = 0$, so that $F_2 = 1 - F_3$. Thus, the system reduces to the 1-dimensional system

$$\begin{aligned}\dot{\xi}_3 &= \xi_3\left(e_3 - \xi_3 e_3 - (1-\xi_3)e_2\right)\cdot U\left(F_3 e_3 + (1-F_3)e_2\right)\\ &= \xi_3(1-\xi_3)(e_3-e_2)\cdot U\left(F_3(e_3-e_2)+e_2\right)\\ &= \xi_3(1-\xi_3)\left\{F_3(e_3-e_2)\cdot U(e_3-e_2) + (e_3-e_2)\cdot Ue_2\right\}\end{aligned}$$

Thus, since U is positive definite, $\dot{\xi}_3 > 0$ for all $0 < F_3 < 1$ if $(e_3 - e_2)\cdot Ue_2 \geq 0$. That is, if $u_{32} - u_{22} \geq 0$. From (64) this condition is $p_0(y_1 + 2y_2) \geq p_1$. But this is always the case if (67) holds. In this case, e_2 is *unstable*.

References

[1] Björnerstedt, J., Weibull, J.W., 1996. Nash equilibrium and evolution by imitation. In: Arrow, K.J., et al. (Eds.), *The Rational Foundations of Economic Behavior*. St. Martins, New York, pp. 155181.

[2] Börgers, T., Sarin, R., 1997. Learning Through Reinforcement and Replicator Dynamics. *Journal of Economic Theory*, 77, 1-14.

[3] Börgers, T., Sarin, R., 2000. Naive reinforcement learning with endogenous aspirations. *International Economic Review*, 41, 921-950.

[4] Börgers, T., Morales, A., Sarin, R., 2004. Expedient and Monotone Learning Rules. *Econometrica*, 72, 383-405.

[5] Burdett, K., Judd, K., 1983. Equilibrium Price Dispersion. *Econometrica* 51, 955-969.

[6] Cason, T., Friedman, D., Wagener, F., 2005. The Dynamics of Price Dispersion, or Edgeworth Variations. *Journal of Economic Dynamics & Control* 29, 801-822.

[7] Cross, J., 1973. A Stochastic Model of Learning Behavior. *Quarterly Journal of Economic Theory.* 87, 239-266.

[8] Dunford, N., Schwartz, J.T., 1964. *Linear Operators. Part I: General Theory*, 2nd ed. Interscience Publishers, Inc. New York.

[9] Erev, I, Roth, A.E., 1998. Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *Am. Econ. Rev.* 88(4), 848-881.

[10] Fudenberg, D., Levine, D., 1998. *Theory of Learning in Games*. MIT Press.

[11] Hart, S., Mas-Colell, A., 2000. A Simple Adaptive Procedure leading to Correlated Equilibrium. *Econometrica* 68, 1127-1150.

[12] Hartman, P., 1964. *Ordinary Differential Equations.* John Wiley & Sons. New York, London, Sydney.

[13] Hines, W.G.S., 1980. Strategy stability in complex populations. *Journal of Applied Probability* 17, 600-610

[14] Hofbauer, J., Sandholm, W., 2005. Evolution in Games with Randomly Disturbed Payoffs. *Journal of Economic Theory* 132, 47-69.

[15] Hofbauer, J., Sigmund, K., 1988. *Theory of Evolution and Dynamical Systems.* Cambridge: Cambridge University Press.

[16] Hofbauer, J., Sigmund, K., 1998. *Evolutionary Games and Population Dynamics.* Cambridge: Cambridge University Press.

[17] Hopkins, E., 1998. Two Competing Models of How People Learn in Games. *Econometrica* 70, 2141-2166.

[18] Hopkins, E., Seymour, R., 2002. The Stability of Price Dispersion under Seller and Consumer Learning. *International Economic Review* 43, 1157-1190.

[19] Lahkar, R., 2009. The Dynamic Instability of Dispersed Price Equilibria. Unpublished Manuscript, IFMR, Chennai. http://ifmr.ac.in/facultyportal/ratullahkar/wp-content/uploads/2009/08/price-dispersion-draft1.pdf

[20] Lach, S., 2002. Existence and Persistence of Price Dispersion: An Empirical Analysis. *Review of Economics and Statistics* 84, 433–444.

[21] Margenau, H., Murphy, GM., 1962. *The Mathematics of Physics and Chemistry*, 2nd edition. Van Nostrand. Princeton, NJ.

[22] Maynard Smith, J. 1982. *Evolution and the Theory of Games.* Cambridge University Press, Cambridge.

[23] Ramsza, M., Seymour, R.M., 2009. Fictitious Play in an Evolutionary Environment. *Games and Economic Behavior*. Published online: doi:10.1016/j.geb.2009.05.003

[24] Samuelson, L., 1997. *Evolutionary Games and Equilibrium Selection.* MIT Press, Cambridge Mass.

[25] Sandholm, W.H., 2009. Population Games and Evolutionary Dynamics. University of Wisconsin-Madison. To be published by MIT Press. http://www.ssc.wisc.edu/ whs/book/pged.pdf

[26] Schlag, K., 1998. Why imitate, and if so, how? A Boundedly Rational Approach to Multi-armed Bandits. *Journal of Economic Theory* 78, 130-156.

[27] Spanier, E.H., 1966. *Algebraic Topology.* McGraw-Hill Book Company, New York

[28] Taylor, P.D., Jonker, L., 1978. Evolutionarily stable strategies and game dynamics. *Math. Biosci.* 40, 145156.

[29] Weibull, J. W., 1995. *Evolutionary Game Theory.* MIT Press, Cambridge.

[30] Young, H. P., 2005. *Strategic Learning and Its Limits.* Arne Ryde Memorial Lectures Series. Oxford University Press, USA.

[31] Zeeman, E.C., 1981. Dynamics of the evolution of animal conflicts; *Journal of Theoretical Biology*; 89, 249-270.