Why Learning Doesn't Add Up: Equilibrium

Selection with a Composition of Learning Rules

Russell Golman[*]

November 1, 2010

---

[*]Department of Applied and Interdisciplinary Mathematics, University of Michigan, 2082 East Hall, 530 Church

Street, Ann Arbor, MI 48109, USA. Telephone: 734 764-7436. Fax: 734 763-0937. E-mail: rgolman@umich.edu

Abstract

In this paper, we investigate the aggregate behavior of populations of learning agents. We compare the outcomes in homogenous populations learning in accordance with imitate the best dynamics and with replicator dynamics to outcomes in populations that mix these two learning rules. New outcomes can emerge. In certain games, a linear combination of the two rules almost always attains an equilibrium that homogenous learners almost never locate. Moreover, even when almost all weight is placed on one learning rule, the outcome can differ from homogenous use of that rule. Thus, allowing even an arbitrarily small chance of using an alternative learning style can shift a population to select a different equilibrium.

KEYWORDS: Adjustment dynamics, basins of attraction, best response dynamics, equilibrium selection, evolutionary game, imitate the best dynamics, learning, replicator dynamics.

JEL classification: C72, C73

1   Introduction

Learning in games, at the population level, can be modeled by dynamic adjustment processes (Swinkels 1993, Fudenberg and Levine 1998). For example, the "proportional imitation" learning rule leads to the replicator dynamics (Schlag 1998), originally proposed by theoretical biologists to describe the growth of a haploid population (Taylor and Jonker 1978). Similarly, the best response dynamics (Gilboa and Matsui 1991) can arise as the continuous time limit of a belief-based learning rule like fictitious play or in a population of myopic best responders. The models that lead to the replicator dynamics and those that lead to the best response dynamics make mutually incompatible assumptions about how strategies are updated, offering no middle ground.

In this paper we propose combining an imitative variant of best response dynamics (Hofbauer 1995) with the replicator dynamics to form tractable composite dynamics. We find that a population is sensitive to the presence of more than one learning style in that composite learning dynamics may have basins of attraction that share little overlap with the homogeneous learning dynamics from which they're formed. We focus in particular on a linear combination of imitate-the-best and replicator dynamics, which we show to select an equilibrium not attained by homogenous use of either pure learning rule in certain matrix games. Introducing even arbitrarily small weight on an alternative learning rule may cause the resulting linear combination of rules to locate a different equilibrium than the pre-existing pure rule. A feature of

the class of games we consider is the existence of a temporary initial best response – an action that looks good at first, but that cannot survive against itself. The temptation to try such an action instead of an equilibrium action creates complex dynamics, which allow the effects of subtle differences in learning styles to build up over time.

In imitate the best dynamics (as with best response dynamics), players slowly adopt myopic best responses, ignoring the possibility that the population mixed strategy may soon change. For a matrix game, the resulting flows are piecewise linear in the direction of the best response to the current population state. In replicator dynamics, players copy others' strategies, so more successful actions spread in the population.[1] These population dynamics can be derived from revision protocols that describe how individual agents learn. We combine imitate the best and replicator dynamics by combining the underlying revision protocols, forming a rule in which either learning protocol may be tried with some positive probability.

For a specific class of coordination games, we show that introducing even a small presence of a second revision protocol into a population changes which equilibrium the population locates. This result underscores the importance of accurately determining exactly how agents learn. Our result builds off of existing comparisons of basins of attraction and equilibrium selection under the best response dynamics and replicator dynamics (Golman and Page 2010a, Golman and Page 2010b). The contribution of

---

[1]Actions initially not present in the population will never be tried with these imitative dynamics, so we always consider initial points in the interior of the strategy space.

this paper is to explore the behavior of a composition of dynamics. We find that heterogeneity across learning styles may bring about new outcomes not foreseeable from analysis of the homogenous dynamics induced by the component learning styles.

Our treatment of equilibrium selection relies on analysis of the deterministic dynamical system. The learning rules define paths through the strategy space that attain equilibria in the games we consider. The equilibrium that is selected is thus a function of the initial population mixed strategy point and the learning dynamic. Nevertheless, we obtain results that hold throughout the interior of the strategy space by focusing on cases in which the basin of attraction of a particular equilibrium approaches the entire space. Our approach does not assume random shocks that allow the population to move from one equilibrium to another, as in the stochastic stability literature (Foster and Young 1990, Young 1993, Kandori et al. 1993, Ellison 2000, Binmore and Samuelson 1999). But, our findings complement that sort of ultralong run analysis because an equilibrium whose basin of attraction approaches the entire strategy space will be uniquely stochastically stable.

The rest of the paper is organized as follows. The next section defines the revision protocols and the learning dynamics derived from them. In Section 3, we present our result comparing equilibrium selection of these rules. Section 4 concludes with a discussion of the importance of models that capture heterogeneity of learning styles.

2 The Learning Dynamics

We assume a symmetric game, with agents recurrently randomly matched from a single population. The set of actions is finite, $\mathcal{A} = \{1, \ldots, n\}$. We let x_i refer to the fraction of the society choosing action $i \in \mathcal{A}$. The population mixed strategy $\mathbf{x} = (x_1, \ldots, x_n)$ is an element of $\triangle^{n-1}$, the $(n-1)$-dimensional simplex where $x_i \geq 0$ for all i and $\sum_i x_i = 1$. Denote by $\pi_i(\mathbf{x})$ the payoff to action i. Naturally, payoffs are a function of the population mixed strategy, $\pi_i : \triangle^{n-1} \to \mathbb{R}$. Denote the vector of these payoffs by $\vec{\pi}(\mathbf{x}) = (\pi_1(\mathbf{x}), \ldots, \pi_n(\mathbf{x}))$. Let $\mathrm{BR}(\mathbf{x})$ be the set of best replies to $\mathbf{x}$,

$$\mathrm{BR}(\mathbf{x}) = \arg\max_{\mathbf{v} \in \triangle^{n-1}} \mathbf{v} \cdot \vec{\pi}(\mathbf{x}).$$

Best response dynamics can be written as

$$\dot{\mathbf{x}} \in \mathrm{BR}(\mathbf{x}) - \mathbf{x}. \tag{1}$$

The *imitate the best dynamics* are

$$\dot{x}_i = x_i \left(I(i|\mathbf{x}) - \sum_j I(j|\mathbf{x}) \, x_j \right) \tag{2}$$

where the best response indicator function $I(i|\mathbf{x}) = 1$ if $\pi_i(\mathbf{x}) = \pi_{\max}(\mathbf{x})$ and 0 otherwise. These differential equations have points of discontinuity, so a (Filippov) solution (i.e., an absolutely continuous function that coincides with the unique solution whenever the differential equations are continuous) must satisfy (2) only for almost all t (Smirnov, 2002; Aubin and Cellina, 1984; Benaim, et al., 2005). When there is a unique best response – say action b – (which, for almost any payoff matrix, there is

at almost every point in the simplex) equation (2) simplifies to $\dot{\mathbf{x}} = x_{\mathrm{b}} \left(\mathrm{BR}(\mathbf{x}) - \mathbf{x} \right)$. The additional factor of x_{b} here, as compared with equation (1), arises because the learning process underlying these dynamics relies on imitation, and best responses are only adopted at a rate proportional to how prevalent they are. In the interior of the strategy space, the trajectory of the imitate the best dynamics matches (one of the solutions of) the best response dynamics, though with different speed.[2]

The *replicator dynamics* are

$$\dot{x}_i = x_i(\pi_i(\mathbf{x}) - \bar{\pi}(\mathbf{x})) \tag{3}$$

where $\bar{\pi}(\mathbf{x}) = \mathbf{x} \cdot \vec{\pi}(\mathbf{x})$ is the average payoff.

Both imitate the best and replicator dyanamics can arise from Poisson alarm clock models (see Sandholm 2009) in which a revision protocol determines the probability that an individual agent switches to another action (as a function of population state and current payoffs) when one's clock rings. The expected motion of this stochastic process defines the learning dynamic.

The paths of the replicator and the imitate the best (and similarly the best re-

---

[2]How the imitate the best dynamics relate to the best response dynamics at the boundary of the strategy space is an open question. Clearly, though, if the best response is absent from the population, the dynamics (as defined by equation (2)) stop altogether. If this case concerns us, we might alter the best response indicator function in equation (2) to obtain the imitate-the-best-available dynamics. An additional property of the imitate the best dynamics that sets it apart from best response dynamics is that every Nash Equilibrium (even in mixed strategies) must be a steady state. Imitate the best dynamics are also related to the imitative logit dynamics as the limiting case when the logit rationality parameter approaches infinity (Weibull, 1995).

sponse) dynamics are invariant under positive affine transformations of the payoff matrix.[3] Payoff transformations do, however, affect the speed of the flow of replicator dynamics. Flows are considered identical if their paths are the same, regardless of speed, because the time parameter has no intrinsic scale.

When trying to combine imitate the best and replicator dynamics, the time parameter becomes relevant. To put both learning dynamics on the same footing, we adopt the framework of a generalized imitation revision protocol. When an agent's alarm goes off, the agent picks an opponent at random and observes this opponent's action. (The selection of an opponent to observe is entirely independent of the random matching of opponents to play against.[4]) The agent observes action j with probability x_j. The agent then switches to the opponent's strategy with probability q that depends on the opponent's payoff. If $q(\pi_j(\mathbf{x})) = \frac{\pi_j(\mathbf{x}) - \pi_{\min}(\mathbf{x})}{\pi_{\max}(\mathbf{x}) - \pi_{\min}(\mathbf{x})}$, the revision protocol generates the replicator dynamics, with a continual rescaling of time. For this *replicator protocol*, the probability of imitation increases linearly in the observed payoff. Alternatively, if $q(\pi_j(\mathbf{x})) = 1$ when $\pi_j(\mathbf{x}) = \pi_{\max}(\mathbf{x})$ and 0 otherwise, the revision protocol generates the imitate the best dynamics. We term this the *imitate the best protocol.*

One natural way to combine imitate the best and replicator dynamics is to sup-

---

[3]Transformations of payoffs that do not affect the dynamics can still determine which of multiple equilibria is payoff dominant. Thus, this traditional equilibrium selection criterion is unrelated to the dynamical equilibrium selection we presently consider.

[4]See Boylan (1992) and Gilboa and Matsui (1992) for justification of random matching in a countably infinite population and Alos-Ferrer (1999) for the case of a continuum of agents.

pose that agents use the replicator protocol with probability p and the imitate the best protocol with probability $1 - p$. [5] A *linear combination of imitate the best and replicator dynamics* emerges:

$$\dot{x}_i = x_i \left[p \frac{\pi_i(\mathbf{x}) - \bar{\pi}(\mathbf{x})}{\pi_{\max}(\mathbf{x}) - \pi_{\min}(\mathbf{x})} + (1 - p) \left(I(i|\mathbf{x}) - \sum_j I(j|\mathbf{x}) \, x_j \right) \right]. \tag{4}$$

We can recover homogeneous use of a single learning protocol by setting $p = 1$ or $p = 0$. With $p = 1$, equation (4) becomes

$$\dot{x}_i = x_i \left(\frac{\pi_i(\mathbf{x}) - \bar{\pi}(\mathbf{x})}{\pi_{\max}(\mathbf{x}) - \pi_{\min}(\mathbf{x})} \right). \tag{5}$$

The factor $\frac{1}{\pi_{\max}(\mathbf{x}) - \pi_{\min}(\mathbf{x})}$ comes from the replicator revision protocol, where it is necessary to ensure that we can interpret q as a probability, i.e., that $q \leq 1$. With the appropriate rescaling of time, it would drop out of equation (5), producing the replicator dynamics of equation (3). With $p = 0$, the linear combination dynamics (4) reduces to the imitate the best dynamics (2).

Note that the linear combination dynamics, as described by (4), is invariant under positive affine transformations of the payoff matrix, and the relative growth rate of any action is bounded between -1 and 1 (and a priori cannot be bounded any more

---

[5] Another possibility, instead of assuming that each agent sometimes uses each protocol, is that there could be some agents who always use the imitate the best protocol and other agents who always use the replicator protocol. The theoretical analysis is more complicated in this case, but by capturing heterogeneity of learning styles across the population, the model appears more realistic. This formulation is presented in the appendix.

tightly). This would not be the case for the more straightforward combination

$$\dot{x}_i = x_i \left[p \left(\pi_i(\mathbf{x}) - \bar{\pi}(\mathbf{x}) \right) + (1-p) \left(I(i|\mathbf{x}) - \sum_j I(j|\mathbf{x}) \, x_j \right) \right]. \tag{6}$$

This straightforward combination dynamic could behave quite differently from (4) because the factor $\pi_{\max}(\mathbf{x}) - \pi_{\min}(\mathbf{x})$ could get bigger or smaller over time and thus shift more or less weight to the replicator dynamics. Nevertheless, Theorem 1 (stated in the next section) would still hold with this dynamic in place of (4), given a particular renormalization of our payoff matrix.

3   Basins of Attraction

Strict Nash equilibria are asymptotically stable for the pure replicator dynamics, the imitate the best dynamics, and the linear combination dynamics.[6] However, agreement of local stability properties does not ensure similar long-run behavior. In games with multiple equilibria, we must consider basins of attraction.

Imitate the best dynamics and replicator dynamics often produce very similar basins of attraction for their equilibria. However, Golman and Page (2010a) provide a class of three-by-three coordination games in which replicator dynamics has basins of attraction arbitrarily different from those of best response dynamics (and hence from imitate the best dynamics as well). Games with such divergence between the behavior of the replicator dynamics and the imitate the best dynamics lend themselves

---

[6]See Hofbauer, et. al. (1979), Hofbauer (2000), Hofbauer and Sigmund (2003), and Hofbauer, et. al. (2009) for more stability results concerning the best response and replicator dynamics.

to the analysis of our composite dynamics. Intuition might lead one to believe that a combination of imitate the best and replicator dynamics would always share some overlap in basins of attraction with each component rule, but this is not so. The linear combination dynamics can be fundamentally different from its parts. Here, we identify a class of games in which imitate the best dynamics and replicator dynamics and the linear combination of the two all have vanishing overlap in their basins of attraction at the same time. That is, by combining an imitate the best protocol and a replicator protocol, the population can locate an equilibrium that neither rule could find on its own.

Consider a class of symmetric games with five actions and the following payoff matrix:

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & N \\ \ln N & \ln N & \ln N - N & \ln N - N & \ln N \\ 0 & 0 & 1 & 1 - \frac{1}{N} & 0 \\ 1 - \frac{1}{N} & 1 - \frac{1}{N} & 1 - \frac{1}{N} & 1 - \frac{1}{N} + \frac{1}{N^2} & 1 - \frac{1}{N} \\ -N^2 & -N^2 & -N^2 & -N^2 & -N^2 \end{pmatrix} \tag{7}$$

where the parameter N is large. Actions 2, 3, and 4 are all strict equilibrium actions. Our result holds for more general classes of games as long as the terms in the payoff matrix keep the same relative orders of magnitude (cf. Golman, 2009).

The payoffs described by (7) may appear complicated, but they permit a simple interpretation of the game. We should think of it as a coordination game, but with a couple of extra options. Action 5 is just a bad choice. So bad, in fact, that all

other actions look good in comparison to it if an agent is using the replicator learning protocol. While such strongly dominated strategies are often ignored when modeling games, it's hard to argue that they don't exist, and there is both theoretical and experimental evidence that they can affect equilibrium selection (Ochs 1995, Basov 2004). Action 1, on the other hand, looks tempting because it can exploit action 5. In fact, action 1 is almost always the initial best response, but it's only a temporary best response because its payoff shrinks as agents learn to avoid action 5. Action 2 is high risk, high reward; agents want to coordinate on it if and only if they almost never see actions 3 or 4 played. Actions 3 and 4 are safer, with moderate payoffs. Of the pair, action 3 would yield a slightly higher equilibrium payoff, whereas action 4 has the highest worst-case payoff.

In Theorem 1, we show that the imitate the best dynamics lead to everyone playing action 2, the replicator dynamics to everyone playing action 3, and the linear combination dynamics to everyone taking action 4. The imitate the best dynamics selects the action that is a best response to the temporary best response; the replicator dynamics selects the action that remains better than average as long as it spreads; and the linear combination dynamics selects something else entirely.

Theorem 1 *Consider the class of games with payoff matrix given by equation (7). As the payoff parameter N grows arbitrarily large, the imitate the best dynamics, the replicator dynamics, and the linear combination dynamics (with any positive weight on both components, i.e., any $p \in (0,1)$) from an arbitrarily large portion of the*

strategy space flow to a single strict equilibrium featuring action 2, 3, or 4 respectively. Notably, all three dynamics share with each other vanishing overlap in their basins of attraction. [7]

Proof See appendix.

For an arbitrarily large portion of the state space, all three dynamics start out heading towards the initial, temporary best response (action 1). But, there are subtle differences between the flows of each dynamic that will greatly affect which action spreads when there's hardly anybody left to exploit with action 1.

With imitate the best dynamics, the relative payoffs of the other actions are irrelevant for their decay rate; agents switch away from actions 3 and 4 just as quickly as they switch from action 5. This mitigates the downside of action 2, and with its residual high reward, it becomes the next best response. By the improvement principle for best response / imitate the best dynamics (Monderer and Sela, 1997; Sela, 2000; Viossat, 2008), if a strict equilibrium action is ever the unique best response to the population state, it remains so thereafter. Thus, action 2 eventually spreads under

---

[7]Theorem 1 would hold with (6) in place of (4) if we were to divide all the entries in our payoff matrix by N^2. Dividing our payoff matrix by N^2 would have the effect of making $\pi_{\max}(\mathbf{x}) - \pi_{\min}(\mathbf{x}) \approx 1$, and thus making (4) and (6) approximately equal. (We cannot rescale payoffs to make $\pi_{\max}(\mathbf{x}) - \pi_{\min}(\mathbf{x}) = 1$ because it varies with time.) Establishing an analogue of Theorem 1 is not a trivial corollary because this small variation in the learning dynamic could potentially have a big impact on equilibrium selection (that is the point of our result, after all), but in this case, the logic of our proof would still go through.

the imitate the best dynamics.

With replicator dynamics, the flow towards the initial best response comes at the expense of the bad choice (action 5). As the payoff to action 5 is made arbitrarily bad relative to all other payoffs, all other actions initially spread along with the initial best response. The spread of actions 3 and 4 inhibits action 2, and the success of action 1 is temporary, so we effectively have actions 3 and 4 racing to achieve critical mass in a coordination game. Although action 4 starts out in the lead because it profits more against action 1, action 3 generates for itself more positive feedback and thus has a much lower critical mass threshold (the relevant comparison in (7) is $a_{33} - a_{34} \gg a_{44} - a_{43}$). Thus, the replicator dynamics yields eventual coordination on action 3.

For the linear combination dynamics, actions 3 and 4 neither keep pace with the spread of the initial best response nor decay as fast as the bad choice. They do survive enough to still inhibit action 2. But the initial lead in payoffs that action 4 has relative to action 3, and the fact that the linear combination dynamics favors a best response more than the replicator dynamics does, means that action 4 now has the advantage. Thus, the linear combination dynamics allows action 4 to become entrenched.

Not only can the linear combination dynamics differ from its component parts, but this deviation in long-run behavior can even occur with almost all weight on one piece of the linear combination. Thus, the mere possibility of using the replicator protocol, however unlikely, is enough to completely shift the long-term behavior of a

14

population of agents using primarily the imitate the best protocol, and vice versa.

4   Discussion

In a large population of agents, we should expect heterogeneity of behavior (Hommes 2006, Kirman 2006).[8] Camerer and Ho (1999) correctly recognized that individuals may combine diverse basic learning rules. Our result demonstrates that the presence of multiple learning styles in the same population, and the specifics of how they all come into play, affects equilibrium selection. We find that for a specific class of coordination games, the outcome of a learning dynamic that is a convex combination of imitate the best dynamics and replicator dynamics is completely different than a convex combination of the outcomes produced by these individual rules. The simple assumption that the combination of two learning dynamics behaves like one of its component pieces is wrong. The explanation is that in games with many actions and multiple equilibria, the trajectory of the combination dynamics can enter parts of the strategy space that would not be approached by either of the component dynamics. Thus, the basins of attraction can be entirely reshaped when agents use multiple learning rules.

Hofbauer et al. (2009) have shown that the time-average of the replicator dynamics is a vanishingly perturbed solution of the best-reply dynamics. Our result, along

---

[8]Heterogeneity can take many forms. Ely and Sandholm (2005) consider agents who all use a common learning rule, but who have diverse preferences, i.e., heterogeneous payoff functions.

with the findings of Golman and Page (2010a; 2010b), illustrates that this does not imply that these dynamics will necessarily select the same equilibria. Even a small perturbation in the learning process can have a strong effect on long run behavior.

Appendix

The Two-Subpopulation Dynamics

Consider a partition of the population into subpopulations of best-responders and replicators, $\mathbf{y}$ and $\mathbf{z}$ respectively.[9] For all i, $x_i = y_i + z_i$. Learning styles are fixed for all time, so $\sum_i z_i = \rho$ is the fraction of agents using the replicator protocol, and $\sum_i y_i = 1 - \rho$ is the fraction using the imitate the best protocol.[10] We assume the two subpopulations are well-mixed, both for the purposes of playing the game and for finding an agent to possibly imitate. Thus, the dynamics of these two subpopulations are coupled together. For the subpopulation learning with the replicator protocol, the following dynamics emerges:

$$\dot{z}_i = \sum_j z_j z_i \left(\frac{\pi_i(\mathbf{x}) - \pi_{\min}(\mathbf{x})}{\pi_{\max}(\mathbf{x}) - \pi_{\min}(\mathbf{x})} \right) - z_i \sum_j z_j \left(\frac{\pi_j(\mathbf{x}) - \pi_{\min}(\mathbf{x})}{\pi_{\max}(\mathbf{x}) - \pi_{\min}(\mathbf{x})} \right) +$$
$$\sum_j z_j y_i \left(\frac{\pi_i(\mathbf{x}) - \pi_{\min}(\mathbf{x})}{\pi_{\max}(\mathbf{x}) - \pi_{\min}(\mathbf{x})} \right) - z_i \sum_j y_j \left(\frac{\pi_j(\mathbf{x}) - \pi_{\min}(\mathbf{x})}{\pi_{\max}(\mathbf{x}) - \pi_{\min}(\mathbf{x})} \right). \tag{8}$$

---

[9]There is reason to think that optimizers and imitators coexist in the same population (Conlisk, 1980).

[10]In a biological setting, we might want to allow the relative sizes of the subpopulations to vary with the success of the learning rule in use. The possibility of the learning rule itself evolving or being learned would be an interesting extension.

The first two terms of (8) represent learning through imitation of other agents who also use the replicator protocol while the last two terms represent learning from members of the other subpopulation. Let $\bar{\pi}_\mu(\mathbf{x})$ be the average payoff in subpopulation μ, meaning that $\mathbf{z} \cdot \vec{\pi}(\mathbf{x}) = \rho\bar{\pi}_z(\mathbf{x})$ and $\mathbf{y} \cdot \vec{\pi}(\mathbf{x}) = (1-\rho)\bar{\pi}_y(\mathbf{x})$. The dynamics (8) simplifies to

$$\dot{z}_i = \rho z_i \left(\frac{\pi_i(\mathbf{x}) - \bar{\pi}_z(\mathbf{x})}{\pi_{\max}(\mathbf{x}) - \pi_{\min}(\mathbf{x})} \right) + \rho y_i \left(\frac{\pi_i(\mathbf{x}) - \pi_{\min}(\mathbf{x})}{\pi_{\max}(\mathbf{x}) - \pi_{\min}(\mathbf{x})} \right) - (1-\rho) z_i \left(\frac{\bar{\pi}_y(\mathbf{x}) - \pi_{\min}(\mathbf{x})}{\pi_{\max}(\mathbf{x}) - \pi_{\min}(\mathbf{x})} \right).$$
$$(9)$$

For the subpopulation learning with the imitate the best protocol, the dynamics are:

$$\dot{y}_i = I(i|\mathbf{x})(1-\rho)(z_i + y_i) - y_i \sum_j I(j|\mathbf{x})(z_j + y_j). \qquad (10)$$

We refer to the system defined by (9) and (10) as the two-subpopulation dynamics.[11]

With twice the variables, the two-subpopulation dynamics is much more complicated than the linear combination dynamics, but strict Nash equilibria are still asymptotically stable, and for some games, at least, we can show that the two dynamics behave alike (Golman, 2009).

Formal Statement and Proof of Theorem 1.

Let m be the Lebesgue measure on the strategy space $\triangle^{n-1}$. Given the parameter N, let $G(N)$ be a symmetric normal form game with payoff matrix A depending on the parameter. That is, the payoff to action i against action j, a_{ij}, is a function of N.

[11]We can recover homogeneous learning dynamics, as in equation (5) or (2), from the two-subpopulation dynamics by setting $\rho = 1$ or $\rho = 0$ respectively.

Then, the expected payoff to action i given random matching with population mixed strategy $\mathbf{x} \in \triangle^{n-1}$ is $\pi_i(\mathbf{x}) = \sum_j a_{ij} x_j$. Henceforth, we omit the payoff function's argument for ease of notation, writing π_i in place of $\pi_i(\mathbf{x})$. The payoff vector can be represented succinctly with matrix multiplication, $\vec{\pi} = A\mathbf{x}$. Given a learning rule $\mathcal{R}$ and an equilibrium action i of the game $G(N)$, let $B(\mathcal{R}, i, N)$ denote the basin of attraction of $(x_i = 1, x_{-i} = 0)$. Let $\mathbf{R}$ denote the replicator dynamics, $\mathbf{B}$ the imitate the best dynamics, and $\mathbf{L(p)}$ the linear combination using the replicator protocol with probability p.

Theorem 1 *Consider the class of games with payoff matrix given by equation (7). For any $\mathcal{R} \in \{\mathbf{B}, \mathbf{R}, \mathbf{L(p)}\}$, with any $p \in (0, 1)$,*

$$\lim_{N \to \infty} m\left(B(\mathcal{R}, i, N)\right) = m\left(\triangle^4\right)$$

for some equilibrium action i, i.e., the imitate the best dynamics, the replicator dynamics, and the linear combination dynamics with any positive weight on both components each flow to a single strict equilibrium from an arbitrarily large portion of the strategy space. Still,

$$\lim_{N \to \infty} \sum_{i \in \mathcal{A}} m\left(B(\mathbf{B}, i, N) \cap B(\mathbf{R}, i, N)\right) = 0$$

and

$$\lim_{N \to \infty} \sum_{i \in \mathcal{A}} m\left(B(\mathbf{L(p)}, i, N) \cap B(\mathbf{B}, i, N)\right) = 0$$

and

$$\lim_{N \to \infty} \sum_{i \in \mathcal{A}} m\left(B(\mathbf{L(p)}, i, N) \cap B(\mathbf{R}, i, N)\right) = 0,$$

18

i.e., all three dynamics share with each other vanishing overlap in their basins of attraction. (In particular, imitate the best dynamics leads to the spread of action 2, replicator dynamics action 3, and the linear combination dynamics action 4.)

Proof Take N to be large. We first show that from an arbitrarily large portion of the strategy space, the imitate the best dynamics locates the equilibrium in which only action 2 is played. Observe that if action 2 is ever the best response, it remains so forever, by the improvement principle.

Observe that $\pi_2 \le \ln N$ as well as $\pi_3 \le 1$ and $\pi_4 \le 1$ everywhere. Consider starting points that satisfy

$$x_5(0) > \frac{1}{\ln N}, \tag{11}$$

a condition that is met in an arbitrarily large portion of the strategy space for sufficiently large N. Initially the best response must be action 1, because inequality (11) implies that $\pi_1 > \ln N$. So all x_i, $i \ne 1$, have the same relative decay rate. While $x_i \ge \frac{\ln N}{N} x_i(0)$,

$$\pi_1 \ge \ln N x_5(0) > 1.$$

If x_i is to reach this lower bound, at the same time

$$\pi_2 \ge \ln N \left(1 - x_3(0) - x_4(0)\right) > 1.$$

Thus, π_2 grows above π_3 and π_4 before π_1 falls to that level. Eventually, action 2 must become the best response, and the dynamics heads towards it thereafter.

We now show that the replicator dynamics leads to the equilibrium in which action

19

3 is always played. Under the replicator dynamics,

$$\dot{x}_3 = x_3\Big[x_3(1-x_3) + (1-\frac{1}{N})x_4(1-x_3) - (1-\frac{1}{N})x_4 - \frac{1}{N^2}x_4^2 +$$

$$(Nx_3+Nx_4-\ln N)x_2+N^2x_5-Nx_5x_1\Big]$$

$$= x_3\Big[x_3(x_1+x_2+x_5) + \frac{1}{N}x_3x_4 - \frac{1}{N^2}x_4^2 + (Nx_3+Nx_4-\ln N)x_2 + N^2x_5 - Nx_5x_1\Big]$$

and we claim this is positive. We need to restrict to starting points that satisfy

$$x_3(0) > \frac{\ln N}{N}, \tag{12}$$

another inequality that is met in an arbitrarily large portion of the strategy space for sufficiently large N. Inequality (12) implies that $\frac{1}{N}x_3x_4 - \frac{1}{N^2}x_4^2 > 0$ and $Nx_3+Nx_4-\ln N > 0$. We have $N^2x_5 - Nx_5x_1 > 0$ easily. Thus, $\dot{x}_3 > 0$ under the replicator dynamics.

Lastly, we show that the linear combination dynamics finds the equilibrium in which everyone takes action 4. Inspection of the payoff matrix reveals that π_5 is always $\pi_{\min}$. For sufficiently large N, an arbitrarily large portion of the strategy space will satisfy

$$x_4(0) > N^{1-p}\frac{\ln N}{N}. \tag{13}$$

Among other things, this means that $\pi_2 < 0$ initially. Assuming additionally that $x_5(0) > \frac{1}{N}$, as will also hold for sufficiently large N, then $\pi_1 = \pi_{\max}$ initially. While

this remains the case, the linear combination dynamics yield

$$\dot{x}_1 = x_1 \left(p \frac{N x_5 - \bar{\pi}}{N^2 + N x_5} + (1-p)(1-x_1) \right) \tag{14}$$

$$\dot{x}_3 = x_3 \left(p \frac{x_3 + (1-\frac{1}{N})x_4 - \bar{\pi}}{N^2 + N x_5} - (1-p)x_1 \right) \tag{15}$$

$$\dot{x}_4 = x_4 \left(p \frac{1 - \frac{1}{N} + \frac{1}{N^2}x_4 - \bar{\pi}}{N^2 + N x_5} - (1-p)x_1 \right) \tag{16}$$

$$\dot{x}_5 = x_5 \left(p \frac{-N^2 - \bar{\pi}}{N^2 + N x_5} - (1-p)x_1 \right). \tag{17}$$

While $x_5 > \frac{1}{N^2 - N}$ and $x_4 > \frac{\ln N}{N}$, straightforward but tedious calculation reveals that the average payoff $\bar{\pi} < 0$, and we have the following bounds:

$$\dot{x}_1 > (1-p)x_1(1-x_1) \tag{18}$$

$$\dot{x}_3 > -(1-p)x_1 x_3 \tag{19}$$

$$\dot{x}_4 > -(1-p)x_1 x_4 \tag{20}$$

$$\dot{x}_5 > -x_5. \tag{21}$$

Also note that

$$
\begin{aligned}
\frac{-N^2 - \bar{\pi}}{N^2 + N x_5} &= \frac{-N^2 - (-N^2 x_5 + N x_5 x_1 - (N x_3 + N x_4)x_2 + \ln N x_2 + \pi_3 x_3 + \pi_4 x_4)}{N^2 + N x_5} \\
&< \frac{-N^2(1 - x_5) - N x_5 x_1 + (N x_3 + N x_4)x_2}{N^2 + N x_5} \\
&= \frac{-N^2(x_1) - N x_5 x_1 + (N x_3 + N x_4)x_2 - N^2(x_2 + x_3 + x_4)}{N^2 + N x_5} \\
&< \frac{(-N^2 - N x_5)x_1}{N^2 + N x_5} = -x_1.
\end{aligned}
$$

So we can bound the left hand side of equation (17),

$$\dot{x}_5 < -x_1 x_5. \tag{22}$$

So

$$x_1(t) \; > \; \frac{1}{1 + \frac{1-x_1(0)}{x_1(0)}e^{-(1-p)t}} \tag{23}$$

$$x_3(t) \; > \; x_3(0)e^{-(1-p)\int_0^t x_1(s)ds} \tag{24}$$

$$x_4(t) \; > \; x_4(0)e^{-(1-p)\int_0^t x_1(s)ds} \tag{25}$$

$$x_5(t) \; < \; x_5(0)e^{-\int_0^t x_1(s)ds} \tag{26}$$

$$x_5(t) \; > \; x_5(0)e^{-t}. \tag{27}$$

For sufficiently large N, an arbitrarily large portion of the strategy space will satisfy $x_1(0) > \frac{1}{N-1}$. Observe that $\pi_4 > \pi_3 + \left(1 - \frac{1}{N}\right)x_1 - \frac{1}{N}x_3$. As long as the dynamics as described by equation (14) apply, $x_1(t) \geq x_1(0)$, and so $\pi_4(t) > \pi_3(t)$.

While $x_5(t) > \frac{1-\frac{1}{N}}{N}$, inequality (26) implies $e^{-\int_0^t x_1(s)ds} > \frac{1-\frac{1}{N}}{Nx_5(0)}$. So, using inequality (25),

$$x_4(t) > x_4(0)\left(\frac{1 - \frac{1}{N}}{Nx_5(0)}\right)^{1-p}. \tag{28}$$

We ensure that initially $1 - \frac{1}{N} > x_5(0)$ by taking N sufficiently large, so inequality (28) simplifies to $x_4(t) > x_4(0)\left(\frac{1}{N}\right)^{1-p}$, and then by inequality (13), we get $x_4(t) > \frac{\ln N}{N}$. So $\pi_2(t) < 0$ as long as $x_5(t) > \frac{1-\frac{1}{N}}{N}$.

Let $t_c = t : \pi_1(t) = \pi_4(t)$. Thus, dividing both sides by N and dropping a positive term from the right side, we have $x_5(t_c) > \frac{1-\frac{1}{N}}{N}$. This means $\pi_4(t) > \pi_2(t)$ and $\pi_4(t) > \pi_3(t)$ for all $t \leq t_c$. The dynamics described by equations (18) through (22) hold until t_c, at which point π_4 becomes $\pi_{\max}$.

Because $\pi_4(t) > \pi_3(t)$ for all $t \le t_c$, we can also conclude that

$$\frac{x_4(t_c)}{x_4(0)} > \frac{x_3(t_c)}{x_3(0)}. \tag{29}$$

We know that $x_5(t_c) < \frac{1}{N}$, too. Inequality (27) then implies $e^{-t_c} < \frac{1}{N x_5(0)}$, and

using inequality (23), we come to

$$x_1(t_c) > \frac{1}{1 + \frac{1-x_1(0)}{x_1(0)} \left(\frac{1}{N x_5(0)}\right)^{1-p}}. \tag{30}$$

After t_c, while π_4 remains $\pi_{\max}$, the linear combination dynamics yield

$$\begin{aligned}
\dot{x}_1 &= x_1 \left(p \frac{\pi_1 - \bar{\pi}}{N^2 + \pi_4} - (1-p)x_4\right) \\
\dot{x}_3 &= x_3 \left(p \frac{\pi_3 - \bar{\pi}}{N^2 + \pi_4} - (1-p)x_4\right) \\
\dot{x}_4 &= x_4 \left(p \frac{\pi_4 - \bar{\pi}}{N^2 + \pi_4} + (1-p)(1-x_4)\right).
\end{aligned}$$

Let

$$f(t) = p \frac{\pi_3 - \bar{\pi}}{N^2 + \pi_4} - (1-p)x_4,$$

so that $\dot{x}_3 = x_3 f(t)$. For $t \ge t_c$,

$$x_3(t) = x_3(t_c) e^{\int_{t_c}^{t} f(s)ds}. \tag{31}$$

Because $\pi_3 \le 1$ and $\pi_1 \ge 0$, we know that $\dot{x}_1 \ge x_1 \left(f(t) - p\frac{1}{N^2}\right)$. So

$$x_1(t) \ge x_1(t_c) e^{\int_{t_c}^{t} f(s)ds} \cdot e^{-\frac{p}{N^2}(t-t_c)}. \tag{32}$$

As $\pi_4 > \pi_3$, we know $\dot{x}_4 > x_4 \left(f(t) + 1 - p\right)$, and

$$x_4(t) \ge x_4(t_c) e^{\int_{t_c}^{t} f(s)ds} \cdot e^{(1-p)(t-t_c)}. \tag{33}$$

23

Recall that $\pi_4 > \pi_3$ if $(1 - \frac{1}{N})x_1 > \frac{1}{N}x_3$. Putting together equations (31) and (32), this holds as long as $(1 - \frac{1}{N})x_1(t_c)e^{-\frac{p}{N^2}(t-t_c)} > \frac{1}{N}(1 - x_1(t_c))$. Plugging in inequality (30) and arranging terms, we find that $\pi_4 > \pi_3$ if

$$e^{-(t-t_c)} > \left[\frac{1}{N-1} \frac{1 - x_1(0)}{x_1(0)} \left(\frac{1}{Nx_5(0)} \right)^{1-p} \right]^{\frac{N^2}{p}}. \tag{34}$$

Note that we also obtain $\pi_4 > \pi_3$ if $\frac{1}{N^2}x_4 > \frac{1}{N}x_3$. Putting together equations (31), (33) and (29), this holds when $\frac{1}{N^2}x_4(0)e^{(1-p)(t-t_c)} > \frac{1}{N}x_3(0)$. Arranging terms, we obtain $\pi_4 > \pi_3$ when

$$e^{-(t-t_c)} < \left(\frac{\frac{1}{N^2}x_4(0)}{\frac{1}{N}x_3(0)} \right)^{\frac{1}{1-p}}. \tag{35}$$

By combining equations (34) and (35), we can be sure $\pi_4 > \pi_3$ for all t if

$$\left(\frac{x_4(0)}{Nx_3(0)} \right)^{\frac{1}{1-p}} > \left[\frac{1}{N-1} \frac{1 - x_1(0)}{x_1(0)} \left(\frac{1}{Nx_5(0)} \right)^{1-p} \right]^{\frac{N^2}{p}}. \tag{36}$$

Inequality (36) holds for initial points in an arbitrarily large portion of the strategy space for sufficiently large N because the right hand side of the inequality decays exponentially in N.

While $\pi_4 = \pi_{\max}$, obviously x_4 is increasing, and so π_2 remains negative for all t after t_c. Thus, action 4 remains the best response forever, and the linear combination dynamics locates the equilibrium in which only action 4 is played. ∎

References

[1] Alos-Ferrer, C., (1999). Dynamical Systems with a Continuum of Randomly Matched Agents. *J. Econ. Theory* 86, 245-267.

[2] Aubin, J., Cellina, A., (1984). *Differential inclusions: set-valued maps and viability theory.* Springer.

[3] Basov, S., (2004). Equilibrium Selection in Coordination Games: Why do dominated strategies matter? *Econ. Bull.* 3, 1-3.

[4] Benaim, M., Hofbauer, J., Sorin, S., (2005). Stochastic Approximations and Differential Inclusions. *SIAM J. Control Optim.* 44, 328348.

[5] Binmore, K., Samuelson, L., (1999). Evolutionary Drift and Equilibrium Selection. *Rev. Econ. Stud.* 66, 363-394.

[6] Boylan, R., (1992). Laws of Large Numbers for Dynamical Systems with Randomly Matched Individuals. *J. Econ. Theory* 57, 473-504.

[7] Camerer, C., Ho, T.H., (1999). Experience Weighted Attraction Learning in Normal Form Games. *Econometrica* 67, 827-874.

[8] Conlisk, J., (1980). Costly Optimizers Versus Cheap Imitators. *J. Econ. Behav. Organ.* 1 (3), 275-293.

[9] Ellison, G., (2000). Basins of Attraction, Long-Run Stochastic Stability, and the Speed of Step-by-Step Evolution. *Rev. Econ. Stud.* 67 (1), 17-45.

[10] Ely, J., Sandholm, W.H., (2005). Evolution in Bayesian Games I: Theory. *Games Econ. Behav.* 53, 83-109.

[11] Foster, D., Young, H.P., (1990). Stochastic Evolutionary Game Dynamics. *Theoretical Population Biology* 38, 219-232.

[12] Fudenberg, D., Levine, D.K., (1998). *The Theory of Learning in Games.* Cambridge: MIT Press.

[13] Gilboa, I., Matsui, A., (1991). Social Stability and Equilibrium. *Econometrica* 59, 859-867.

[14] Gilboa, I., Matsui, A., (1992). A Model of Random Matching. *J. Math. Econ.* 21, 185-197.

[15] Golman, R. (2009). Essays on Population Learning Dynamics and Boundedly Rational Behavior. Ph.D. dissertation, University of Michigan.

[16] Golman, R., Page, S., (2010a). Basins of Attraction and Equilibrium Selection Under Different Learning Rules. *J. Evol. Econ.* 20, 49-75.

[17] Golman, R., Page, S., (2010b). Individual and Cultural Learning in Stag Hunt Games With Multiple Actions. *J. Econ. Behav. Organ.* 73, 359-376.

[18] Hofbauer, J., (1995). Imitation Dynamics for Games, mimeo.

[19] Hofbauer, J., (2000). From Nash and Brown to Maynard Smith: Equilibria, Dynamics and ESS. *Selection* 1, 81-88.

[20] Hofbauer, J., Schuster, P., Sigmund, K., (1979). A Note on Evolutionary Stable Strategies and Game Dynamics. *J. Theor. Biol.* 81, 609-612.

[21] Hofbauer, J., Sigmund, K., (2003). Evolutionary Game Dynamics. *Bull. Am. Math. Soc.* 40, 479-519.

[22] Hofbauer, J., Sorin, S., Viossat, Y., (2009). Time Average Replicator and Best-Reply Dynamics. *Math. Oper. Res.* forthcoming.

[23] Hommes, C., (2006). Heterogeneous Agent Models in Economics and Finance. In: Judd, K., Tesfatsion, L. (eds) Handbook of Computational Economics, vol 2, Agent-Based Computational Economics. North Holland, Amsterdam.

[24] Kandori, M., Mailath, G., Rob, R., (1993). Learning, Mutation and Long Run Equilibria in Games. *Econometrica* 61, 29-56.

[25] Kirman, A., (2006). Heterogeneity in Economics. *J. Econ. Interact. Coord.* 1, 89-117.

[26] Monderer, D., Sela, A., (1997). Fictitious Play and No-Cycling Conditions. Sfb504 Publications 97-12, University of Mannheim.

[27] Ochs, J., (1995). Coordination Problems. *Handbook of Experimental Economics.* Kagel, J., Roth, A. (eds) Princeton: Princeton University Press.

[28] Sandholm, W. H., (2009). Pairwise Comparison Dynamics and Evolutionary Foundations for Nash Equilibrium. Working paper, University of Wisconsin. http://www.ssc.wisc.edu/~whs.

[29] Schlag, K. H., (1998). Why Imitate, and If So, How? A Boundedly Rational Approach to Multi-armed Bandits. *J. Econ. Theory* 78, 130-156.

[30] Sela, A., (2000). Fictitious Play in 2×3 Games. *Games Econ. Behav.* 31, 152-162.

[31] Smirnov, G., (2002). *Introduction to the theory of differential inclusions.* Providence: AMS Bookstore.

[32] Swinkels, J. M., (1993). Adjustment Dynamics and Rational Play in Games. *Games Econ. Behav.* 5, 455-484.

[33] Taylor, P.D., Jonker, L., (1978). Evolutionarily Stable Strategies and Game Dynamics. *Math. Biosci.* 40, 145-156.

[34] Viossat, Y., (2008). Evolutionary Dynamics May Eliminate All Strategies Used in Correlated Equilibria. Mathematical Social Sciences 56, 27-43.

[35] Weibull, J., (1995). *Evolutionary Game Theory.* Cambridge, MA: The M.I.T. Press.

[36] Young, H.P., (1993). The Evolution of Conventions. *Econometrica* 61, 57-84.