Published in final edited form as:

J Theor Biol. 2012 May 21; 301: 161–173. doi:10.1016/j.jtbi.2012.02.021.

Evolutionary game dynamics in populations with different learners

Krishnendu Chatterjee^{a,1}, Damien Zufferey^{a,1}, and Martin A. Nowak^b

^aIST Austria (Institute of Science and Technology Austria)

^bProgram of Evolutionary Dynamics, Harvard University

Abstract

We study evolutionary game theory in a setting where individuals learn from each other. We extend the traditional approach by assuming that a population contains individuals with different learning abilities. In particular we explore the situation where individuals have different search spaces, when attempting to learn the strategies of others. The search space of an individual specifies the set of strategies learnable by that individual. The search space is genetically given and does not change under social evolutionary dynamics. We introduce a general framework and study a specific example in the context of direct reciprocity. For this example we obtain the counter intuitive result that cooperation can only evolve for intermediate benefit-to-cost ratios, while small and large benefit-to-cost ratios favor defection. Our paper is a step toward making a connection between computational learning theory and evolutionary game dynamics.

Keywords

Evolutionary game theory; Direct reciprocity (Prisoners dilemma); Learning theory

1. Introduction

Traditional evolutionary game theory considers a well-mixed population of individuals that use a finite number of strategies (Maynard Smith & Price, 1973; Maynard Smith, 1982; Hofbauer & Sigmund, 1988; Fudenberg & Tirole, 1991; Weibull, 1995; Skyrms, 1996; Samuelson, 1997; Hofbauer & Sigmund, 1998; Gintis, 2000; Cressman, 2003; Hofbauer & Sigmund, 2003; Nowak & Sigmund, 2004; Sigmund, 2010). The result of the interaction of any two strategies is specified by a payoff matrix. Payoff determines the fitness. Deterministic evolutionary dynamics are described by the replicator equation (Taylor & Jonker, 1978; Hofbauer et al., 1979; Zeeman, 1980), which can be interpreted in terms of genetic or cultural evolution. In the latter context, individuals imitate or learn the strategies of others. Learners might preferentially choose teachers that have a high payoff. In the standard replicator equation the outcome of the learning process is independent of the learner; all learners can adopt the strategy of any teacher.

© 2012 Elsevier Ltd. All rights reserved.

¹Joint first authors

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Here we assume that the outcome of a learning interaction also depends on the learner and that learners may differ in the set of strategies that they can learn (in their search space). In other words the population is *heterogeneous* containing individuals with different learning abilities. It is possible that the strategy used by a teacher is not available in the search space of the learner. In this case the learner may approximate the teacher's strategy with another (possibly similar) strategy. We assume that each individual has a fixed search space, which specifies the set of strategies learnable by that individual.

Any learning task that involves 'inductive inference' (as opposed to simple memorization) requires a limited search space (Gold, 1967; Vapnik, 1998; Valiant, 1984; Osherson et al., 1986). The learner receives examples of behavior and has to infer the rules that generate these behaviors. In the context of language learning the examples are given by the linguistic forms received by the child learner. The rules are the grammar that is used by the native speaker (the 'teacher'). The limited search space is described by 'universal grammar' (Chomsky, 1965, 1981). Universal grammar is a consequence of the architecture of the human brain, which is genetically determined. Our approach in this paper is inspired by mathematical models of language evolution (Nowak et al., 2001, 2002; Mitchener & Nowak, 2003; Komarova, 2004; Komarova & Niyogi, 2004; Niyogi, 2006; Mitchener, 2007; Komarova & Levin, 2010).

Similar principles apply in situations where individuals attempt to learn strategies for playing repeated games by observing other individuals. The learner only sees examples of behaviors and has to infer the rule that is used by the teacher. Thus the learning task requires inductive inference. After seeing enough examples, the learner should have a chance to infer the correct strategy. Such a learning task only succeeds if there is a limited search space. The search space itself cannot be learned. It is given before learning starts. It is the set of all strategies learnable by this individual. The search space cannot be unlimited. It can either contain a finite number of strategies or an infinite number with a prior distribution determining which strategies are more likely than others.

Any human learner trying to learn strategies for repeated games by observing others must have a limited search space, which is ultimately genetically determined. It is possible that modern humans have the same search space (as they are believed to have the same universal grammar), but sometime in human evolution there must have been modifications of the search space. In this paper we study cultural evolution in a population where individuals differ in their learning abilities (search spaces). We do not model genetic evolution in this paper. Therefore we assume that the population consists of fixed fractions of learners that have different search spaces. We explore how heterogeneity in learn-ability affects cultural evolutionary dynamics.

We are interested in the fundamental principles of evolutionary dynamics. Therefore we study very simple cases. In Section 4 below, for example, we contrast two types of learners for the repeated Prisoner's dilemma. One type can only learn the unconditional strategies, always cooperate (ALLC) and always defect (ALLD), while the other type can learn ALLC, ALLD and tit-for-tat (TFT). Many other examples can be imagined and should be studied in due course. The concept also gives rise to an empirical research program, where human subjects could be evaluated according to their learning behavior when observing (successful) strategies in repeated games.

There is a large literature on learning in evolutionary games (Fudenberg & Levine, 1995; Riechmann, 2001; Samuelson, 1997; Kandori et al., 1993; Panait et al., 2008; Sigmund et al., 2010; Benaïm et al., 2009; Sandholm, 2010). (Helbing, 1992) explores stochastic equations of social learning which generalize the replicator equation. (Schlag, 1998) studies

learning in the context of games against nature (multi-arm bandit learning), while (Hofbauer & Schlag, 2000) analyze learning in two-player zero-sum games. The two later papers present learning rules that lead to Nash-equilibria in games with one player against nature or two-player zero-sum games. Both these works focus on describing learning rules to optimize payoff in a homogeneous population, whereas in our setting the learning rule is simple and fixed and the population is heterogeneous. Mistakes in the learning process can be seen as cultural mutation events that help to explore the search space (Traulsen et al., 2009). Our paper adds to this literature on learning in games the idea that cultural evolutionary dynamics can occur in heterogeneous populations, where individuals differ in their learning abilities.

2. The basic model

In this section we introduce our basic model. We provide a stepwise generalization starting from the standard replicator equation:

$$\dot{x}_k = x_k(f_k - \bar{f}) \quad k=1 \dots n \quad (1)$$

Here x_k denotes the frequency of strategy k . There are n strategies. The total payoff (fitness) of strategy k is given by $f_k = \sum_j a_{kj}x_j$, where a_{kj} are the entries of the payoff matrix, A . The quantity a_{kj} denotes the payoff that strategy k receives when interacting with strategy j . The average fitness of the population is given by $\bar{f} = \sum_k x_k f_k$. Note that $\sum_k x_k = 1$. The standard replicator equation represents a situation where (i) the outcome of the learning process depends only on the teacher (and not on the learner), (ii) the learning process occurs without mistakes, and (iii) all individuals can learn all strategies.

If we add the possibility that learning occurs with mistakes, then we obtain the replicator-mutator equation:

$$\dot{x}_k = \sum_{j=1}^n x_j f_j Q_{jk} - \bar{f} x_k \quad k=1 \dots n \quad (2)$$

Again teachers are selected proportional to their abundance and fitness, $x_j f_j$. The entries of the mutation matrix, Q_{jk} , denote the probability the learner will adopt strategy k when learning from a teacher who uses strategy j . Q is a stochastic matrix: all entries are probabilities and the rows sum to one. If Q is the identity matrix then we obtain the replicator equation as a special case. In general, equation (2) describes learning with mistakes, but again the learning process only depends on the teacher and all individuals can learn all strategies. For the replicator-mutator equation (2) we require that all payoffs are non-negative, $a_{ij} \geq 0$, otherwise inconsistencies can arise (see Appendix for a discussion of this requirement).

We now generalize this approach to allow the situation where the learning process also depends on the strategy that is currently used by the learner. Thus, there can be different types of individuals with different learning abilities and different search spaces. We introduce a three-dimensional transition matrix, T , whose entries T_{ijk} specify the probability that a learner with strategy i learning from a teacher with strategy j will adopt strategy k .

Hence for all i, j we have $\sum_{k=1}^n T_{ijk} = 1$. The transition matrix can be used to encode a heterogeneous population of learners, where some individuals can only learn a subset of all the strategies.

We obtain the following equation:

$$\dot{x}_k = \sum_{i=1}^n \sum_{j=1}^n x_i x_j f_j T_{ijk} - \bar{f} x_k \quad k=1 \dots n \quad (3)$$

The underlying idea is that a learner with strategy i picks a teacher with strategy j proportional to $x_j f_j$ (average payoff times frequency); the result of this interaction is that the learner adopts strategy k . As before x_i denotes the frequency of strategy i . There are n strategies, and we have $\sum_{i=1}^n x_i = 1$. The average payoff (fitness) of strategy i is $f_i = \sum_{j=1}^n a_{ij} x_j$ where a_{ij} is the payoff that strategy i receives from strategy j . We require that all payoffs are non-negative, $a_{ij} \geq 0$. The average fitness of the population is $\bar{f} = \sum_{i=1}^n f_i x_i$.

The two novel features of equation (3) are as follows: (i) the learning process depends on the learner as well as on the teacher, and (ii) the population can contain individuals with different learning abilities. Observe that equation (3) reverts back to the standard replicator equation if $T_{ijk} = 1$ for $j = k$ and $T_{ijk} = 0$ otherwise. Also observe that equation (3) subsumes equation (2) for the special case $T_{ijk} = Q_{jk}$ for all i .

3. Examples

In this section we present two simple examples which illustrate how equation (3) works.

Example 1

Consider a population where individuals adopt one of two strategies, A or B . The payoff matrix is given as follows:

$$\begin{matrix} & \begin{matrix} A & B \end{matrix} \\ \begin{matrix} A \\ B \end{matrix} & \begin{pmatrix} a & b \\ c & d \end{pmatrix} \end{matrix}$$

Here a, b, c, d are non-negative numbers. The payoff matrix has the following meaning: (i) if A plays A , then both get a , (ii) if B plays B , then both get d , and (iii) if A plays B , A gets b and B gets c .

Let us consider the following learning events: (i) if learner and teacher have the same strategy then nothing changes; (ii) if a learner with strategy A learns from a teacher with strategy B , then the learner adopts B with probability p and A with probability $1 - p$; (iii) if a learner with strategy B learns from a teacher with strategy A , then the learner adopts A with probability q and B with probability $1 - q$. Therefore the strategies A and B can imply different propensities to change.

We show the $2 \times 2 \times 2$ transition matrix, T , in two parts, $T_{-,A,-}$ and $T_{-,B,-}$, by fixing the teacher A and B , respectively:

$$T_{-,A,-} = \begin{matrix} & \begin{matrix} A & B \end{matrix} \\ \begin{matrix} A \\ B \end{matrix} & \begin{pmatrix} 1 & 0 \\ q & 1-q \end{pmatrix} \end{matrix} \quad T_{-,B,-} = \begin{matrix} & \begin{matrix} A & B \end{matrix} \\ \begin{matrix} A \\ B \end{matrix} & \begin{pmatrix} 1-p & p \\ 0 & 1 \end{pmatrix} \end{matrix}$$

Let x denote the frequency of individuals with strategy A . Consequently $1-x$ denotes the frequency of individuals with strategy B . Equation (3) leads to

$$\dot{x} = x(1-x)(aqx - bqx - cpx + dp + bq - dp)$$

Hence, the probabilities p and q simply multiply the entries of the payoff matrix. The system is equivalent to a standard replicator equation for the modified payoff matrix

$$\begin{matrix} A \\ B \end{matrix} \begin{pmatrix} A & B \\ aq & bq \\ cp & dp \end{pmatrix}$$

The learning asymmetry has the simple effect of transforming the payoff matrix. A is a strict Nash equilibrium if $aq > cp$. B is a strict Nash equilibrium if $bq < dp$.

Example 2

Again we consider a game with two strategies A and B , but this time only a fraction, η , of the population can choose between A and B while the remaining fraction, $1 - \eta$, always plays B . We model this situation by considering a population with three strategies, A , B , and B' , and the payoff matrix:

$$\begin{matrix} A \\ B \\ B' \end{matrix} \begin{pmatrix} A & B & B' \\ a & b & b \\ c & d & d \\ c & d & d \end{pmatrix}$$

The learning events are as follows. A learner who uses A or B will adopt A if the teacher uses A , but it will adopt B if the teacher uses B or B' . A learner who uses B' will always stay with B' . The 3×3 transition matrix, T , can be written as follows:

$$T_{-A,-} = \begin{matrix} A \\ B \\ B' \end{matrix} \begin{pmatrix} A & B & B' \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad T_{-B,-} = \begin{matrix} A \\ B \\ B' \end{matrix} \begin{pmatrix} A & B & B' \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad T_{-B',-} = \begin{matrix} A \\ B \\ B' \end{matrix} \begin{pmatrix} A & B & B' \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Let x_A and x_B denote the frequencies of type A and B individuals, respectively. The effect of the presence of type B' individuals can be interpreted as changing the payoff of the players, and adding a frequency dependent mutation for type B players. Equation (3) leads to the following:

$$\begin{aligned} \dot{x}_A &= (x_A + x_B)x_A(ax_A + bx_B + \underbrace{b(1-\eta)}_{\text{payoff change}}) - \bar{f}x_A \\ \dot{x}_B &= (x_A + x_B)x_B(cx_A + dx_B + \underbrace{d(1-\eta)}_{\text{payoff change}}) + \underbrace{(x_A + x_B)(1-\eta)(cx_A + dx_B + d(1-\eta))}_{\text{frequency dependent mutation}} - \bar{f}x_B \end{aligned}$$

For the sake of analysis, we now consider the above equation with only one variable x_A , and we have $x_B = \eta - x_A$. Let x denote the frequency of individuals with strategy A, where $0 \leq x \leq \eta \leq 1$. Then equation (3) gives us

$$\dot{x} = x((\eta - x)(ax + b(1 - x)) - (1 - x)(cx + d(1 - x)))$$

This can be written as

$$\dot{x} = xF(x)$$

where

$$F(x) = x^2(b - a + c - d) + x(a\eta - (1 + \eta)b - c + 2d) + b\eta - d$$

For $\eta = 1$ we obtain a standard replicator equation, where in addition to a possible interior equilibrium there are always equilibrium points at $x = 0$ and $x = 1$. For $0 < \eta < 1$ we observe that $x = 0$ is still an equilibrium but $x = \eta$ is typically not an equilibrium. Note that $x = \eta$ is an equilibrium only if $c = d = 0$; in this case $F(\eta) = 0$. But if $c\eta + d(1 - \eta) > 0$ then $F(\eta) < 0$. (Remember that a, b, c, d must be non-negative; otherwise inconsistencies can arise. One such inconsistency would be $F(\eta) > 0$.)

If $F(\eta) = 0$ then there are two cases: (i) if $a = b = 0$ then the process is neutral; every point x in $[0, \eta]$ is an equilibrium; (ii) if a or b are positive then the only stable equilibrium is $x = \eta$.

We now analyze the situation for $F(\eta) < 0$. Again there are two cases

- *Case 1:* $\eta b > d$. We have $F(0) > 0$ and $F(\eta) < 0$. Since $F(x)$ is a quadratic continuous function, there exists exactly one $0 < x^* < \eta$ such that $F(x^*) = 0$. The point x^* is a stable equilibrium.
- *Case 2:* $\eta b < d$. We have $F(0) < 0$ and $F(\eta) < 0$. There are three possibilities: (i) there is no solution $F(x) = 0$ in the range $[0, \eta]$; the system converges to $x = 0$; (ii) there are two solutions, $0 < x_1^* < x_2^* < \eta$, of $F(x) = 0$; solution x_1^* is an unstable equilibrium; solution x_2^* is a stable equilibrium; depending on the initial condition the system converges either to $x = 0$ or to $x = x_2^*$; and (iii) there is exactly one solution, x^* , of $F(x) = 0$; the line $x = 0$ is the tangent to $F(x) = 0$; here x^* is a saddle point; the system converges to x^* if the initial condition, x_0 , is greater than x^* ; the system converges to 0 if $x_0 < x^*$.

An interesting numerical example is the following: $a = 9, b = 4, c = 5, d = 3$. In the standard game (for $\eta = 1$) it is clear that A is a strict Nash equilibrium and dominates B. But for $\eta = 0.7$ the presence of B' players, who never change their strategy, topples the Nash equilibrium; now the selection dynamics favor B over A, and all-B is the only stable equilibrium.

4. Direct reciprocity

Let us now study an example of direct reciprocity in the context of the repeated Prisoner's Dilemma (Rapaport & Chammah, 1965; Trivers, 1971; Axelrod, 1984; Fudenberg &

Maskin, 1986; May, 1987; Milinski, 1987; Boyd, 1989; Fudenberg & Maskin, 1990; Nowak & Sigmund, 1989, 1992, 1993a; Hauert, 2010). An interaction between two players consists of multiple rounds. In each round, both players can either cooperate or defect.

We consider three strategies: always cooperate (ALLC), always defect (ALLD) and tit-for-tat (TFT). We assume there are two types of individuals in the population. Type A individuals can only learn two strategies, ALLC and ALLD. Type B individuals can learn all three strategies. Therefore, the population is heterogeneous in terms of learning abilities. Type A individuals can learn strategies that require no memory (unconditional strategies), while type B individuals can learn strategies with a one-bit memory (reactive strategies).

The learning event can be seen as query based. The learner queries how the teacher reacts to cooperation and defection. If a teacher uses ALLC or ALLD, then both type A and B individuals will learn the teacher's strategy. But if a teacher uses TFT, then only type B learners adopt TFT, while type A learners adopt ALLC or ALLD with equal probability. A type A learner observes that the teacher uses both cooperation and defection, and since the strategy space for this learner is only ALLC or ALLD it chooses one of those strategies with equal probability.

The transition matrix T is shown below. T_{ijk} represents the probability that an individual with strategy i learning from an individual with strategy j adopts strategy k . To distinguish between A and B players, we label ALLC and ALLD as C and D when played by type A individuals. The indexes of T are (1) C, (2) D, (3) ALLC, (4) ALLD, and (5) TFT. We show sections of the transition matrix, $T_{-,X,-}$, where the teacher uses strategy X.

$$\begin{aligned}
 T_{-,C,-} &= T_{-,ALLC,-} = \begin{matrix} & \begin{matrix} C & D & ALLC & ALLD & TFT \end{matrix} \\ \begin{matrix} C \\ D \\ ALLC \\ ALLD \\ TFT \end{matrix} & \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \end{matrix} \\
 T_{-,D,-} &= T_{-,ALLD,-} = \begin{matrix} & \begin{matrix} C & D & ALLC & ALLD & TFT \end{matrix} \\ \begin{matrix} C \\ D \\ ALLC \\ ALLD \\ TFT \end{matrix} & \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \end{matrix} \\
 T_{-,TFT,-} &= \begin{matrix} & \begin{matrix} C & D & ALLC & ALLD & TFT \end{matrix} \\ \begin{matrix} C \\ D \\ ALLC \\ ALLD \\ TFT \end{matrix} & \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{matrix}
 \end{aligned}$$

The frequencies of the two types are constant: η denotes the frequency of B and $1 - \eta$ the frequency of A. The parameter η is given and does not change under cultural evolutionary dynamics. Let x_1 and x_2 denote the frequencies of A individuals that use ALLC and ALLD, respectively. Let x_3 , x_4 and x_5 denote the frequencies of B individuals that use ALLC, ALLD and TFT, respectively. The state space of our system is a subset of the simplex, S_5 , and is given by $x_1 + x_2 = 1 - \eta$ and $x_3 + x_4 + x_5 = \eta$.

We consider a simplified Prisoner's Dilemma where cooperation means paying a cost, c , for the other player to receive a benefit, b . Defection implies no cost and induces no benefit. If $b > c > 0$ then mutual cooperation is better than mutual defection and the game is a Prisoner's Dilemma. We assume there is a fixed probability to play another round of the game; the expected number of rounds is given by m . Taking the average payoff per round we obtain the payoff matrix

$$\begin{array}{c}
 C \\
 D \\
 ALLC \\
 ALLD \\
 TFT
 \end{array}
 \begin{pmatrix}
 C & D & ALLC & ALLD & TFT \\
 b-c & -c & b-c & -c & b-c \\
 b & 0 & b & 0 & b/m \\
 b-c & -c & b-c & -c & b-c \\
 b & 0 & b & 0 & b/m \\
 b-c & -c/m & b-c & -c/m & b-c
 \end{pmatrix}$$

We add a constant $\gamma \geq c$ to all entries to ensure that all payoffs are non-negative. For our numerical examples and simulations we set $\gamma = 1$, $c = 1$ and $m = 100$, while we vary η and b .

Let us first discuss the case $\eta = 1$. This means there are no type A individuals, and hence $x_C = x_D = 0$. Our system simplifies to a standard replicator equation on the simplex S_3 given by $x_{ALLC} + x_{ALLD} + x_{TFT} = 1$. There is no interior equilibrium, because the payoff of TFT individuals exceeds that of ALLC individuals by the amount $c x_{ALLD}$. For $x_{ALLD} > 0$ the payoff of TFT strictly exceeds the payoff of ALLC, and therefore both $\dot{x}_{ALLC}$ and $\dot{x}_{TFT}$ cannot simultaneously be 0. Thus there cannot be an equilibrium with $x_{ALLC} > 0$, $x_{ALLD} > 0$ and $x_{TFT} > 0$. We have the following equilibria on the boundary of the simplex: (i) the ALLD-corner, given by $x_{ALLD} = 1$, $x_{ALLC} = x_{TFT} = 0$, is a stable equilibrium, (ii) every point on the edge from ALLC to TFT is an equilibrium; these points are given by $x_{ALLD} = 0$, $x_{ALLC} = x$ and $x_{TFT} = 1 - x$ where $0 \leq x \leq 1$; there is a threshold value $x^* = 1 - [c/(b(1-1/m))]$; if $x < x^*$ the equilibrium is stable; if $x > x^*$ the equilibrium is unstable, and (iii) there is one bistable equilibrium on the edge from ALLD to TFT, the point is given by $x_{ALLC} = 0$, $x_{TFT} = 1 - [c/((b-c)(m-1))]$, $x_{ALLD} = 1 - x_{TFT}$. Some evolutionary trajectories converge to the ALLD-corner, while others converge to the stable points on the ALLC to TFT edge.

We now discuss the evolutionary dynamics of our system as observed in computer simulation. We show results for different values of b/c and η . Recall that a fraction η of individuals can learn ALLC, ALLD, and TFT, while the remaining $1 - \eta$ fraction of individuals can only learn ALLC and ALLD.

Figure 1 shows the evolutionary dynamics where we fix $b/c = 1.5$ and vary η . The case $\eta = 1$ behaves as discussed above. For $\eta = 0.95$ we find there is a stable interior equilibrium. Intuitively, the coexistence between the strategies is possible because the presence of type A learners favors ALLD over TFT for two reasons: (i) the presence of unconditional cooperators of type A increases the payoff of ALLD and (ii) the presence of unconditional defectors of type A leads some type B learners to adopt ALLD.

For $\eta = 0.9$ and $\eta = 0.85$, there is no interior equilibrium and the population converges to ALLD. For any given b/c value, there is a threshold of η below which there are no interior equilibria and ALLD is the only stable equilibrium. For example, for $b/c = 1.5$ there exist interior equilibria as long as $\eta > 0.93$. For smaller η values there are no interior equilibria. Thus, as η decreases the system favors defection. This makes sense: cooperation can only prevail if the fraction of individuals, which can learn a conditional strategy, is high enough. Note also that interior equilibria always arise in pairs: one is stable and the other is unstable.

Figure 2 shows the evolutionary dynamics for $b/c = 5$. For $\eta = 0.95$ and $\eta = 0.9$ we find a stable limit cycle in the interior of the phase space. For lower values of η the limit cycle moves closer to ALLD; this means the time average of ALLD increases. For $\eta = 0.85$ the population converges to ALLD. Again we find that increasing the proportion of type A learners makes it harder for cooperation to be maintained in the population, which makes sense. The occurrence of a limit cycle, whose amplitude grows as η declines, is of interest.

Next we study the system for fixed $\eta = 0.9$ while varying b/c (Figure 3). For low b/c the only stable equilibrium is the one where all players defect (Figure 3(a)). As we increase b/c a stable interior equilibrium appears (Figure 3(b)). (In Appendix we give an analytic approximation for this threshold.) Depending on the initial condition the population either converges to the interior equilibrium or to ALLD. At the interior equilibrium A learners adopt a mixture between ALLC and ALLD, while B learners adopt a mixture between ALLC, ALLD, and TFT. For larger b/c ratios the interior equilibrium is approached in damped oscillations (Figure 3(c)). Further increase in b/c results in the interior equilibrium losing its stability: now the limit cycle appears (Figure 3(d)). As we increase b/c the size of the limit cycle increases (Figure 3(d) to Figure 3(h)); the oscillations have larger amplitude. At a critical b/c ratio the limit cycle collides with the boundary of the state space and disappears. Now the only stable equilibrium is again ALLD (Figure 3(i)). Therefore, our system has the interesting property that cooperation only emerges for intermediate benefit-to-cost ratios. For large or small benefit-to-cost ratios there is no possibility for the evolution of cooperation.

Figure 5 shows the particular parameter region where the limit cycle goes closer and closer to the ALLD attractor region. As b/c increases the population spends longer and longer time periods very close to the state where everyone defects. For $b/c > 6.23$ the population is absorbed into the ALLD state.

Simulation results for a wide range of parameters are available at <http://pub.ist.ac.at/~zufferey/cooperation/index.html>.

In summary, we observe that the presence of type A learners can lead to interior equilibria and limit cycles. For given values of η (less than one) there is an interesting nonmonotonic behavior: for small and for large b/c ratios there is no evolution of cooperation, but for intermediate b/c ratios evolution of cooperation is possible. One reason for this non-monotonicity is that type A learners promote the success of ALLD relative to TFT. When type A individuals learn from TFT they can either become unconditional cooperators or defectors. Unconditional cooperators increase the payoff of ALLD. Unconditional defectors induce some B learners to adopt ALLD (thereby generating a ‘mutational flux’ in the B subpopulation). These effects lead to the oscillatory behavior. As b/c increases the amplitude of the oscillations increases until the limit cycle collides with the boundary. At this point the only stable equilibrium is one where everyone defects.

There is a minimum frequency of TFT players needed to prevent the population from being attracted to the state where everyone defects. As the amplitude of the oscillations increases the frequency of TFT drops below this threshold and cooperation becomes extinct.

5. Adding mutation

The formalism that is described by eq (3) can also include mutation, which means errors in learning. The mutation probabilities can be directly incorporated into the transition matrix, T , because the entry T_{ijk} denotes the probability that an i individual learning from a j individual adopts strategy k .

Another possibility is to consider a separate mutation matrix, M . Then we can write the transition matrix, T , as a linear combination of an error-free transition matrix, T^0 , and the mutation matrix. Suppose the entry, M_{ik} , of the mutation matrix specifies the probability that an individual with strategy type i mutates to become an individual with strategy type k . We can write:

$$T_{-j-} = (1 - \mu)T_{-j-}^0 + \mu M$$

The parameter μ denotes the probability that a mutation event occurs instead of an (error-free) learning event.

For the example that we have discussed in the previous section it is natural to consider the following mutation matrix:

$$M = \begin{matrix} & \begin{matrix} C & D & ALLC & ALLD & TFT \end{matrix} \\ \begin{matrix} C \\ D \\ ALLC \\ ALLD \\ TFT \end{matrix} & \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1/3 & 1/3 & 1/3 \end{pmatrix} \end{matrix}$$

This mutation matrix specifies that ‘mutation’ of an *A* individual leads to *C* and *D* with equal probability. If a *B* individual mutates, then the strategies, ALLC, ALLD, and TFT, are adopted with equal probability.

In Figure 6 we show the effect of mutation on the evolutionary dynamics. We observe that mutation pulls the population toward the center of the simplex. Most interestingly, mutation preserves certain fraction of TFT strategies, and hence prevents the population being attracted to ALLD strategies. Thus for high benefit-to-cost ratios mutation acts to stabilize cooperation (McNamara et al., 2004; McNamara & Leimar, 2010).

6. Conclusion

The goal of this research is to promote the connection between learning theory (Valiant, 1984; Vapnik, 1998; Niyogi, 2006) and evolutionary game theory. Learning the strategy of other players in a repeated game is a problem of inductive inference: the learner observes examples of behavior and has to infer the rules (the grammar, the strategy) that generate such behavior. In order to solve such a problem the learner must have a limited search space (a limited set of possible strategies). The learning process between a student and a teacher can be described as the student picking a strategy from the search space and evaluating if this strategy is consistent with the behavior of the teacher.

In this paper we have proposed a simple framework to study deterministic evolutionary game dynamics in conjunction with learning. Eq (3) is a generalization of the standard replicator equation. It includes the possibility that the outcome of a learning process not only depends on the strategy of the teacher, but also on the strategy that is currently used by the learner. Eq (3) can also describe situations where a population consists of different types of individuals that differ in their search spaces. In this paper we considered the learning model where the outcome of the learning process only depends on the learner and the teacher, and is independent of the frequencies of the strategies in the population. The more involved model where the learning depends on the frequencies of other strategies in the population would be a topic for future research. (Schlag, 1993) has studied direct reciprocity with strategies played by automata. Another future research topic would be to extend such studies to heterogeneous population where memory not only incurs a cost but also limit the search space of individuals.

We have studied a particular example that arises in the context of direct reciprocity. Direct reciprocity is a mechanism for the evolution of cooperation based on conditional strategies and repeated interactions between the same two individuals. We have considered a population of heterogeneous learners. Type A individuals can only learn ALLC and ALLD. Type B individuals can learn ALLC, ALLD, and TFT. The proportion of A and B individuals does not change under cultural evolutionary dynamics (Cavalli-Sforza & Feldman, 1981; Boyd & Richerson, 1985). If the proportion of A individuals is below a critical threshold then there is a possibility for the evolution of cooperation for intermediate benefit-to-cost ratios. If the benefit-to-cost ratio is too small then the only stable equilibrium is one where everyone uses ALLD. As the benefit-to-cost ratio increases we observe at first a stable interior equilibrium and then a stable limit cycle. For larger benefit-to-cost ratios the limit cycle increases and eventually collides with the boundary of the simplex. Then again ALLD is the only stable equilibrium. Our system has the interesting (and novel) property that increasing the benefit-to-cost ratio can prevent evolution of cooperation.

We note that oscillations are a fundamental property of the evolutionary dynamics of cooperation. Cooperative states are never fully stable. The average amount of cooperation that is present in a population depends on how long cooperative states prevail and how rapidly they can be re-established after a breakdown of cooperation. For oscillations in evolution of cooperation see (Nowak & Sigmund, 1989, 1993b; Imhof et al., 2005; Tarnita et al., 2009).

Acknowledgments

We thank the anonymous reviewers for their comments that helped us to improve the paper significantly. Supports from the John Templeton Foundation and the NSF/NIH joint program in mathematical biology (NIH grant R01GM078986), ERC Start grant (279307: Graph Games), FWF NFN Grant No S11407-N23 (RiSE), FWF Grant No P 23499-N23, and Microsoft faculty fellows awards are gratefully acknowledged.

References

- Axelrod, RM. The evolution of cooperation. New York NY: Basic Books; 1984. (reprinted 1989. Harmondsworth UK: Penguin.)
- Benaïm M, Hofbauer J, Hopkins Ed. Learning in games and unstable equilibria. *Journal of Economic Theory*. 2009; 144:1694–1709.
- Boyd R. Mistakes allow evolutionary stability in the repeated prisoner's dilemma game. *J theor Biol*. 1989; 136:47–56. [PubMed: 2779259]
- Boyd, R.; Richerson, PJ. *Culture and the Evolutionary Process*. University of Chicago Press; 1985.
- Cavalli-Sforza, LL.; Feldman, MW. *Cultural transmission and evolution: A quantitative approach*. Princeton NJ: Princeton University Press; 1981.
- Chomsky, N. *Aspects of the Theory of Syntax*. MIT Press; 1965.
- Chomsky, N. Principles and parameters in syntactic theor. In: Hornsetin, N.; Lightfoot, D., editors. *Explanation in Linguistics*. London: Longman; 1981. p. 123-146.
- Cressman, R. *Evolutionary dynamics and extensive form games*. Cambridge, MA: MIT Press; 2003.
- Fudenberg, D.; Levine, KD. *The theory of learning in games*. Cambridge, MA: MIT Press; 1995.
- Fudenberg D, Maskin E. The folk theorem in repeated games with discounting or with incomplete information. *Econometrica*. 1986; 50:533–554.
- Fudenberg D, Maskin E. Evolution and cooperation in noisy repeated games. *Am Econ Rev*. 1990; 80:274–279.
- Fudenberg, D.; Tirole, J. *Game theory*. Cambridge, MA: MIT Press; 1991.
- Gintis, H. *Game theory evolving*. Princeton, NJ: Princeton University Press; 2000.
- Gold EM. Language identification in the limit. *Inform Control*. 1967; 10:447–474.

- Hauert, Ch. Replicator Dynamics of Reward & Reputation in Public Goods Games. *J theor Biol.* 2010; 267:22–28. [PubMed: 20708018]
- Helbing D. Interrelations between stochastic equations for systems with pair interactions. *Physica A: Statistical Mechanics and its Applications.* 1992; 181:29–52.
- Hofbauer J, Schlag KH. Sophisticated imitation in cyclic games. *Journal of Evolutionary Economics.* 2000; 10:523–543.
- Hofbauer J, Schuster P, Sigmund K. A note on evolutionary stable strategies and game dynamics. *J Theor Biol.* 1979; 81:609–612. [PubMed: 537389]
- Hofbauer, J.; Sigmund, K. *The theory of evolution and dynamical systems.* Cambridge, UK: Cambridge University Press; 1988.
- Hofbauer, J.; Sigmund, K. *Evolutionary games and population dynamics.* Cambridge, UK: Cambridge University Press; 1998.
- Hofbauer J, Sigmund K. Evolutionary game dynamics. *B Am Math Soc.* 2003; 40:479–519.
- Imhof LA, Fudenberg D, Nowak MA. Evolutionary cycles of cooperation and defection. *Proc Natl Acad Sci USA.* 2005; 102:10797–10800. [PubMed: 16043717]
- Kandori M, Mailath GJ, Rob R. Learning, mutation and long run equilibria in games. *Econometrica.* 1993; 61:29–56.
- Komarova NL. Replicator-mutator equation, universality property and population dynamics of learning. *J Theor Biol.* 2004; 230(2):227–239. [PubMed: 15302554]
- Komarova NL, Niyogi P. Optimizing the mutual intelligibility of linguistic agents in a shared world. *J Artificial Intelligence.* 2004; 54(1–2):1–42.
- Komarova NL, Levin SA. Eavesdropping and language dynamics. *J Theor Biol.* 2010; 264(1):104–18. [PubMed: 20093123]
- May RM. More evolution of cooperation. *Nature.* 1987; 327:15–17.
- Maynard Smith, J. *Evolution and the theory of games.* Cambridge, UK: Cambridge University Press; 1982.
- Maynard Smith J, Price GR. The logic of animal conflict. *Nature.* 1973; 246:15–18.
- McNamara JM, Barta Z, Houston AI. Variation in behaviour promotes cooperation in the Prisoner's Dilemma game. *Nature.* 2004; 428:745–748. [PubMed: 15085131]
- McNamara JM, Leimar O. Variation and the response to variation as a basis for successful cooperation. *Phil Trans R Soc B* 2010. 2010; 365:2627–2633.
- Milinski M. Tit for tat in sticklebacks and the evolution of cooperation. *Nature.* 1987; 325:433–435. [PubMed: 3808044]
- Mitchener WG, Nowak MA. Competitive exclusion and coexistence of universal grammars. *B Math Biol.* 2003; 65:67–93.
- Mitchener WG. Game Dynamics with Learning and Evolution of Universal Grammar. *B Math Biol.* 2007; 69(3):1093–1118.
- Niyogi, P. *The Computational Nature of Language Learning and Evolution.* MIT Press; 2006.
- Nowak MA, Komarova NL, Niyogi P. Evolution of universal grammar. *Science.* 2001; 291:114–118. [PubMed: 11141560]
- Nowak MA, Komarova NL, Niyogi P. Computational and evolutionary aspects of language. *Nature.* 2002; 417:611–617. [PubMed: 12050656]
- Nowak MA, Sigmund K. Oscillations in the evolution of reciprocity. *J Theor Biol.* 1989; 137:21–26. [PubMed: 2593672]
- Nowak MA, Sigmund K. Tit for tat in heterogeneous populations. *Nature.* 1992; 355:250–253.
- Nowak MA, Sigmund K. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. *Nature.* 1993; 364:56–58. [PubMed: 8316296]
- Nowak MA, Sigmund K. Chaos and the evolution of cooperation. *Proc Natl Acad Sci USA.* 1993; 90:5091–5094. [PubMed: 8506355]
- Nowak MA, Sigmund K. Evolutionary dynamics of biological games. *Science.* 2004; 303:793–799. [PubMed: 14764867]

- Osherson, DN.; Stob, M.; Weinstein, S. Systems that learn: An introduction to learning theory for cognitive and computer scientists. Cambridge: MIT Press; 1986.
- Panait L, Tuyls K, Luke S. Theoretical Advantages of Lenient Learners: An Evolutionary Game Theoretic Perspective. *Jour of Machine Learning Research*. 2008; 9:423–457.
- Rapoport, A.; Chammah, AM. Prisoner's dilemma. Ann Arbor MI: University of Michigan Press; 1965.
- Riechmann T. Genetic algorithm learning and evolutionary games. *Journal of Economic Dynamics and Control*. 2001; 25:1019–1037.
- Samuelson, L. Evolutionary games and equilibrium selection. Cambridge, MA: MIT Press; 1997.
- Sandholm, WH. Population Games and Evolutionary Dynamics. Cambridge, MA: MIT Press;
- Schlag, KH. Discussion Paper Serie B. Vol. 243. University of Bonn; Germany: 1993. Dynamic stability in the repeated prisoners' dilemma played by finite automata.
- Schlag KH. Why Imitate, and If So, How? : A Boundedly Rational Approach to Multi-armed Bandits. *Journal of Economic Theory*. 1998; 78:130–156.
- Sigmund, K. Princeton Series in Theoretical and Computational Biology. Princeton University Press; 2010. The Calculus of Selfishness.
- Sigmund K, De Silva H, Traulsen A, Hauert C. Social learning promotes institutions for governing commons. *Nature*. 2010; 466:861–863. [PubMed: 20631710]
- Skyrms, B. Evolution of the Social Contract. Cambridge, UK: Cambridge University Press; 1996.
- Tarnita CE, Antal T, Ohtsuki H, Nowak MA. Evolutionary dynamics in set structured populations. *Proc Natl Acad Sci USA*. 2009; 106(21):8601–8604. [PubMed: 19433793]
- Taylor PD, Jonker LB. Evolutionary stable strategies and game dynamics. *Math Biosci*. 1978; 40:145–156.
- Traulsen A, Hauert C, De Silva H, Nowak MA, Sigmund K. Exploration dynamics in evolutionary games. *Proc Nat Acad Sciences*. 2009; 106:709–712.
- Trivers RL. The evolution of reciprocal altruism. *Q Rev Biol*. 1971; 46:35–57.
- Valiant LG. A theory of learnable. *Commun ACM*. 1984; 27:436–445.
- Vapnik, VN. Statistical learning theory. Hoboken NJ: John Wiley; 1998.
- Weibull, JW. Evolutionary game theory. Cambridge, MA: MIT Press; 1995.
- Zeeman, EC. Population dynamics from game theory. In: Nitecki, ZH.; Robinson, RC., editors. proceedings of an international conference on global theory of dynamical systems. Vol. 819. New York, NY: Springer; 1980.

Appendix A. Discussion of requirements of equation (2) and equation (3)

We discuss that with equation (2) we require that all payoff entries a_{ij} are non-negative, as otherwise it may lead to inconsistency. We illustrate it with an example.

Consider a population with two strategies A and B and the payoff matrix is given as follows:

$$\begin{matrix} & \begin{matrix} A & B \end{matrix} \\ \begin{matrix} A \\ B \end{matrix} & \begin{pmatrix} a & b \\ c & d \end{pmatrix} \end{matrix}$$

such that $a = b = 0$ and $c = d = -1$. Let the matrix Q be defined as follows:

$$Q = \begin{matrix} & \begin{matrix} A & B \end{matrix} \\ \begin{matrix} A \\ B \end{matrix} & \begin{pmatrix} 1-u & u \\ u & 1-u \end{pmatrix} \end{matrix}$$

where $0 < u < 1$. Let x denote the frequency of type A individuals. Then for this example equation (2) gives us

$$\dot{x} = (1-x)(x-u)$$

Note that x must lie in the interval $[0, 1]$. However, $\dot{x} < 0$ at $x = 0$, which leads to inconsistency. The reason of the inconsistency is as follows: the multiplicative terms with mutations must be non-negative, and if a_{ij} are negative, then the payoff for a strategy can be negative. Since equation (3) generalizes equation (2), we also require the same condition (i.e., $a_{ij} \geq 0$) for equation (3).

Appendix B. Details of the Repeated Prisoner's Dilemma with Learners

In this section we present the details of the results of repeated Prisoner's Dilemma with two types of learners: type A that can learn only ALLC and ALLD strategies, and type B that can learn one of ALLC, ALLD, and TFT strategies. The details of the payoff and transition matrix is presented in the main paper. In the explanation below we write β for b/c (the benefit to cost ratio).

Phase 1: low β , equilibrium is ALLD

As described above, when β is close to 1, the only stable equilibria consists of the ALLD strategy. The reason that when β is close to 1 the TFT strategy cannot invade ALLD strategies is as follows: given a population of ALLD strategies, if there is a small fraction of TFT strategies, then the payoff of TFT strategies is better. However, the individuals of type A cannot learn TFT and play ALLC or ALLD with probability 1/2. Against the ALLC strategy the payoff of playing ALLD is better than the one of TFT. Hence, in Phase 1 if β is greater than 1, but close to 1, TFT individuals will do worse than ALLD individuals. The situation of Phase 1 is depicted in Figure 4(a), for $\eta = 0.9$.

Phase 2: stable interior equilibrium

When we increase β sufficiently, then interior equilibria start to exist. In other words, for β sufficiently greater than 1, defection is not the only stable equilibrium. When the interior equilibria starts to exist, they arise in pairs. However, only one of them is stable. For η close to 1 (i.e., small fraction of type A individuals), we obtain an approximate condition on β when the transition between Phase 1 and Phase 2 takes place. Our approximation gives an expression for β in terms of η where the transition takes place and the expression is given as follows:

$$\beta = \left(1 + \frac{2 \cdot (1 - \eta^2)}{2\eta + \eta^2 - 1} \right)^2 \quad (\text{B.1})$$

For $\eta = 0.9$ we get $\beta = 1.53$ from the above expression. Figure 3(a) and Figure 3(b) show that the transition takes place between $\beta = 1.5$ and $\beta = 2.0$. More precisely our experimental results show that the transition takes place at $\beta = 1.62$. Thus, our approximation shows close agreement with experimental results. When the two interior equilibria begin to coexist, one of them is stable and the other is unstable. The details of how to derive equation (B.1) and discussion about the stability of the resulting equilibria is in the following section. Figure 4(b) shows the point where a stable interior equilibrium appears. Figure 4(c) shows that as β

further increases the equilibrium is approached in damped cycles, a prelude to the next phase.

Phase 3: limit cycle

When β is further increased, the two interior equilibria still exist but neither of them are stable. In our experimental results, we started with an initial condition that is close to the equilibria and the evolutionary dynamics did not converge to the equilibria, instead it started forming a limit cycle. As the value of β is further increased the oscillations' amplitude increases. The limit cycles (or oscillations) are shown from Figure 4(d) to Figure 4(h)

Unstable equilibria

In Phase 2 and Phase 3 though there exists equilibria, some of the equilibria are unstable. In Phase 2 one of the interior equilibria is unstable and in Phase 3 both the interior equilibria are unstable. We will then argue how the figures we present already support the instability of the equilibria. In our experimental study, to understand the stability of an equilibrium, we started with initial configurations that are close to the equilibrium value, and for unstable equilibria, the evolutionary dynamics did not converge to the equilibria.

Phase 4: high β , equilibrium is ALLD

As β is increased, the size of the limit cycle continues to grow. When β is sufficiently large, the cycle eventually 'collides' with the boundary of the state space. Due to the learning dynamics this collision can occur only on the ALLC-ALLD face of the simplex, or on the ALLD-TFT face but extremely close to ALLD. Then the whole population converges to ALLD strategies. In other words, increasing β in this case is the cause of defection. This is shown in Figure 4(i). The transition from Phase 3 to Phase 4 for $\eta = 0.9$ is shown in Figure 5.

Smaller η values

We present the evolutionary dynamics of Phase 1—Phase 4 for $\eta = 0.9$, in Figure 3 and Figure 4. For a smaller value of η , e.g. 0.8, we observed that even with large value of β interior equilibria did not appear. Figure 8 shows the evolutionary dynamics for the case when $\eta = 0.8$ for $\beta = 3, 4$, and 8.

Appendix C. Mathematical details of Appendix Appendix B

In this section we present the details related to existence of interior equilibria and also stability analysis of the interior equilibria.

Appendix C.1. Analytic results for existence of interior equilibria with η close to 1 and $m = \infty$

In this section we present analytic results to approximate the threshold value of β where interior equilibria starts to exist. The approximation result is obtained for the case when η is close to 1 (i.e., the limit of η is 1) and the limit $m = \infty$. We setup the equations for the existence of equilibria. We will show an approximation that allows us to derive an expression for β , in terms of η , when η is close to 1, that ensures the transition from Phase 1 to Phase 2 (i.e., existence of stable interior equilibria).

Frequencies and payoff

Let us denote by y_1 the relative frequency of ALLD strategies of type A individuals, and x_1 and x_2 the relative frequency of ALLD and ALLC strategies of type B individuals. Furthermore, to make the distinction between A and B clearer, we label ALLC and ALLD as C and D when played by A. In other words, the absolute frequencies are as follows: (1) type A ALLD strategies: $fr_D = (1 - \eta) \cdot y_1$; (2) type A ALLC strategies: $fr_C = (1 - \eta) \cdot (1 - y_1)$; (3) type B ALLD strategies: $fr_{ALLD} = \eta \cdot x_1$; (4) type B ALLC strategies: $fr_{ALLC} = \eta \cdot x_2$; and (5) type B TFT strategies: $fr_{TFT} = \eta \cdot (1 - x_1 - x_2)$. The payoff are as follows:

$$\begin{aligned} p_D &= b \cdot (fr_C + fr_{ALLC}) + c; \\ p_C &= (b - c) \cdot (fr_{TFT} + fr_C + fr_{ALLC}) - c \cdot (fr_D + fr_{ALLD}) + c; \\ p_{ALLD} &= b \cdot (fr_C + fr_{ALLC}) + c; \\ p_{ALLC} &= (b - c) \cdot (fr_{TFT} + fr_C + fr_{ALLC}) - c \cdot (fr_D + fr_{ALLD}) + c; \\ p_{TFT} &= (b - c) \cdot (fr_{TFT} + fr_C + fr_{ALLC}) + c. \end{aligned}$$

Note that the term c is added as we add the constant $\gamma = c$ to ensure all payoff entries are non-negative. Let ℓ be the vector that denotes the probability of selecting a strategy as teacher and is proportional to the payoff, e.g., ℓ_{TFT} is proportional to p_{TFT} . Hence the new frequency vector is obtained as follows:

$$\begin{aligned} fr'_D &= (1 - \eta) \cdot (\frac{\ell_{TFT}}{2} + \ell_C + \ell_D); \\ fr'_C &= (1 - \eta) \cdot (\frac{\ell_{TFT}}{2} + \ell_C + \ell_D); \\ fr'_{ALLD} &= \eta \cdot (\ell_D + fr_{ALLD}); \\ fr'_{ALLC} &= \eta \cdot (\ell_C + fr_{ALLC}); \\ fr'_{TFT} &= \eta \cdot \ell_{TFT}. \end{aligned}$$

The equilibria condition is specified by the following five equalities that states that the new frequency and old frequency vector are the same:

$$fr'_C = fr_C; \quad fr'_D = fr_D; \quad fr'_{ALLD} = fr_{ALLD}; \quad fr'_{ALLC} = fr_{ALLC}; \quad fr'_{TFT} = fr_{TFT}.$$

From the above equations, we can eliminate all but one variable (say variable x_1) and then we obtain a degree five polynomial on x_1 .

The analysis of transition from Phase 1 to Phase 2

The degree five polynomial in x_1 , (that has β and η as parameters) is a very complicated equation. For the analysis of the transition of Phase 1 to Phase 2 we use the following observations for simplification:

1. *Elimination of y_1 .* Since η is close to one, the effect of learning from D and C is limited during the selection process. Therefore, we compute the value of y_1 at the limit $\eta \rightarrow 1$. This allows us to express y_1 only in terms of x_1, x_2, β . The resulting equation is much simpler and given by

$$y_1 = \frac{((3\beta - 1) \cdot x_1 - \beta) \cdot x_2 + (\beta - 1) \cdot x_1^2 + (3 - 2\beta) \cdot x_1 + \beta}{(2\beta - 2) \cdot x_1 \cdot x_2 + (2\beta - 2) \cdot x_1^2 + (4 - 4\beta) \cdot x_1 + 2\beta}$$

However, it is still quite long and contains β . If we plot y_1 as a function of x_1 , x_2 and β , and fit a plane of x_1 and x_2 in the region of interest ($x_2 \approx 0$), we obtain a very good approximation given as follows:

$$y_1 = \frac{1}{2} \cdot (1 + x_2 - x_2)$$

In Figure 9, we evaluate the quality of our approximation.

2. *Elimination of x_2 .* When the first interior equilibria sets in, among type B individuals the fraction of ALLC strategies must be small as compared to ALLD and TFT strategies. This is because at the starting point where TFT strategies can prevent an attraction towards ALLD strategies, the fraction of ALLC strategies must be small as ALLC strategies favor attraction to ALLD.
3. *Elimination of higher order $(1 - \eta)$ terms.* We obtain our expression for value of $1 - \eta$ which is small, and we can ignore higher order terms of $(1 - \eta)$ unless multiplied with other large numbers (such as β^2).

By fixing x_2 to a small fixed value, eliminating y_1 and higher order terms of $(1 - \eta)$ we obtain a simplified cubic equation in x_1 (we use Maxima, a computer algebra system, to obtain the equation). The simplified equation itself is quite long. One of the solution of the equation is $x_1 = 1$, and we factor out the solution of $x_1 = 1$, and consider the remaining quadratic equation. The discriminant of the simplified quadratic equation is

$$\sqrt{(\beta^2 - 2\beta + 1) \cdot \eta^4 + (4\beta^2 - 4) \cdot \eta^3 + (2\beta^2 - 4\beta + 2) \cdot \eta^2 + (4 - 4\beta^2) \cdot \eta + \beta^2 - 2\beta + 1}$$

For existence of real roots we require that the discriminant is positive. We interpret the above discriminant as a quadratic equation in β , and solve it for 0. Any value of β greater than the solution ensures that the discriminant is positive. The quadratic equation of β has two roots: one is $\beta = 1$, and the other desired one is

$$\begin{aligned} \beta &= \frac{\eta^4 - 4\eta^3 + 2\eta^2 + 4\eta + 1}{\eta^4 + 4\eta^3 + 2\eta^2 - 4\eta + 1} \\ &= \frac{(\eta^2 - 2\eta - 1)^2}{(\eta^2 + 2\eta - 1)^2} = \frac{(1 + 2\eta - \eta^2)^2}{(1 - 2\eta - \eta^2)^2} \\ &= \left(1 + \frac{2 \cdot (1 - \eta^2)}{2\eta + \eta^2 - 1}\right)^2 \end{aligned}$$

Appendix C.2. Stability analysis of the equilibria

In this section we study the stability of the equilibria, and this was done with the aid of simulation (i.e., the results are not analytical).

Equilibrium Frequencies and Instability of Equilibria: Phase 1—Phase 2

We obtained the solution of the equilibrium from the degree five polynomial (using the tool Maxima) for the Phase 1 and transition to Phase 2. For initial values of β close to 1 we observed that there are two imaginary solutions, one negative solution (out of range) and two corner solutions. Increasing the values of β we observe that the imaginary solutions become real solutions that are interior solutions. If we consider the two interior solutions we

observe the following: (1) one of the solution is close to the boundary with high value of x_1 (i.e., high fraction of ALLD strategies in type B individuals), and this equilibrium is unstable; and (2) the other solution has significantly lower value of x_1 and this equilibrium is stable. Observe that in Figure 4 we started with the initial configuration with all uniform strategies and it converged to ALLD strategies. Hence for the equilibrium that has high value of ALLD strategies, if we start with an initial configuration close to the equilibrium it always converged to the ALLD strategies. Thus it is quite clear that the equilibrium is unstable. On the other hand, we observed in Figure 4 that when we started with the all uniform initial configuration they converged to the other equilibrium. In our experimental study we started with initial configuration close to the equilibrium values and observed that they converge to the equilibrium. Thus the second interior equilibrium is stable.

Equilibrium Frequencies and Instability of Equilibria: Phase 2—Phase 3

We then considered the solution of the equilibrium obtained from the degree five polynomial for the Phase 2 and transition to Phase 3. We consider the two interior solutions and observe that as β is increased the following properties of the solutions: (1) one of the solution is close to the boundary with very high value of x_1 (i.e., large fraction of ALLD strategies), and this equilibrium is unstable, and the analysis is same as in the previous; and (2) the other solution gets closer to the center of the simplex (all strategies have nearly the same value). Observe that in Figure 4 we started with the initial configuration of all uniform strategies and observed live cycles. Hence as we increase the value of β the second equilibrium loses stability. We experimented with initial configuration close to the equilibrium and observed that live cycles were formed. Hence in Phase 3 both the interior equilibria are unstable. Figure 10 shows the interior equilibrium in the simplices and shows how they change as β increases.

The highlights of the paper titled “Evolutionary game dynamics in populations with different learners” are as follows:

1. We introduce a general framework of evolutionary game theory where the population consists of different type of learners.
2. We study an example of direct reciprocity and show that cooperation can only evolve for intermediate benefit-to-cost ratios.
3. We show that in our framework mutation acts to stabilize cooperation.
4. Our paper is a step toward making a connection between computational learning theory and evolutionary game dynamics.

Figure 1.

The figure shows the evolutionary dynamics for different values of $\eta = 1, 0.95, 0.9$ and 0.85 , and the b/c value is 1.5 . The figure shows the evolutionary trajectories with starting point of the trajectories close to the ALLC and TFT edge. The figure shows the U-shaped trajectories for $\eta = 1$, the interior equilibrium for $\eta = 0.95$, and the convergence to ALLD strategies for $\eta = 0.9$ and $\eta = 0.85$.

Figure 2.

The figure shows the evolutionary dynamics for different values of $\eta = 1, 0.95, 0.9$ and 0.85 , and the b/c value is 5. The starting point of the trajectories are close to the ALLC and TFT edge. The figure shows U-shaped trajectories for $\eta = 1$, limit cycles for $\eta = 0.95$ and $\eta = 0.9$ (the limit cycle for $\eta = 0.9$ is closer to ALLD), and the convergence to ALLD strategies for $\eta = 0.85$.

Figure 3.

The figure shows the evolutionary dynamics for different values of b/c for $\eta = 0.9$. Figure 3(a) shows the overall behavior of the population in Phase 1 for $b/c = 1.5$, before the transition to Phase 2. Though b/c is greater than 1 (but close to 1) the population still converges to all defection. Figure 3(b) and Figure 3(c) show the result of Phase 2 when there exists stable interior equilibrium. Figure 3(b) shows the result when b/c has crossed the threshold for Phase 2 to begin. Figure 3(c) shows that Phase 2 continues while b/c increases. However, the equilibrium becomes less stable and is approached in damped oscillations. Figure 3(d) to Figure 3(h) illustrate Phase 3. Figure 3(d) shows the result when Phase 3 sets in. The damped oscillations of Phase 2 turn into limit cycles (interior equilibria not stable any more). We observe that as b/c is increased the size of the limit cycle increases, i.e., the amplitude of the oscillations increases. Figure 3(i) shows Phase 4 where for large values of b/c the population converges to defection.

Figure 4.

The figure shows the evolutionary dynamics for $\eta = 0.9$ and the b/c values vary as in Figure 3, and the initial configuration is uniform distribution over the strategies.

Figure 5.

This figure shows in details the transition from Phase 3 to Phase 4, for $\eta = 0.9$. We observe that the more b/c is increased, the closer the lives cycle goes to boundary of the ALLD and TFT strategies, and it is closer to the stable defection equilibrium. When b/c is large enough, the trajectory hits the boundary close to the ALLD strategies. Once the trajectory hits the boundary very close to ALLD strategies, then TFT strategies cannot prevent attraction from ALLD strategies, and then the whole population converges to defection.

Figure 6 and Figure 7.

Figure 6 shows the results of experiments similar to Figure 3 but with a non-zero mutation rate. The mutation rate was 1%. We observe that mutation pulls the population toward the center of the simplex. Most interestingly, mutation prevents the limit cycle from colliding with the boundary of the simplex for high benefit-to-cost ratios. Thus for high benefit-to-cost ratios mutation acts to stabilize cooperation. Figure 7 shows the evolutionary dynamics similar to Figure 4 but with 1% mutation rate.

Figure 8.

In this figure we show the case of $\eta = 0.8$. In this case, the population always converges to defection even with large b/c . The shape of the curves tends to show that the population would go from Phase 1 directly to Phase 4. When b/c is large enough to admit an interior equilibrium, the trajectory is already a trajectory similar to the Phase 4 of the other populations we have seen. Figure 8(a), 8(b) and 8(c) shows the dynamics for $b/c = 3, 4$, and 8 , respectively.

Figure 9.

The figure shows that comparison of the approximation and the exact value for y_1 when $\eta = 0.9$, $x_2 = 0.02$, and $\beta = 1, 2, 3$.

Figure 10.

The figure shows the interior equilibria in the simplices and how they change as b/c is increased. Figure 10(a) and (b) show the results for $\eta = 0.9$ and $\eta = 0.95$, respectively, for different values of $\beta = b/c$.

Figure 11.

In this figure we show the simulation results where values of η are 1, 0.98, 0.95, 0.9 and 0.85, and b/c vary from 1.5 to 7.