

# Conformist Transmission and the Evolution of Cooperation

Jorge Peña

Institute of Applied Mathematics, University of Lausanne,  
UNIL-Dorigny, CH-1015 Lausanne, Switzerland  
jorge.pena@unil.ch

## Abstract

We study the effects of conformist transmission on the evolutionary dynamics of the Prisoner's Dilemma, the Snowdrift and the Stag Hunt games in both well-mixed and spatially structured populations. The addition of conformism introduces a transformation of the payoff matrix that favours the stability of pure equilibria and reduces the basin of attraction of risk dominant equilibria. When both conformism and local interactions are present, the system can exhibit higher levels of cooperation than those obtained in the absence of any of the two mechanisms.

## Introduction and Related Work

Evolutionary game theory (Hofbauer and Sigmund, 1998; Gintis, 2000) is the theory of evolutionary dynamics when selection is frequency-dependent, i.e. when the success of an individual is conditioned not only by the strategy she holds but also by the strategies held by other individuals in the population. Although originally developed as an application of game theory to the study of genetic evolution (Maynard Smith, 1982), evolutionary game theory has also been used to investigate cultural evolutionary processes, that is the way ideas or beliefs spread through a population of individuals capable of imitation.

In *cultural* evolutionary game-theoretic models, ideas are transmitted via biased imitation. Most of these models posit that the only important psychological bias underlying imitation is *prestige* or *payoff-based bias*, defined as the predisposition to imitate successful individuals. Under the assumption of a very large and well-mixed population, payoff-based biased transmission can be shown to generate a famous differential equation, named the *replicator dynamics* (Taylor and Jonker, 1978; Gintis, 2000). In the context of evolutionary game theory, the equilibrium points and other characteristics of the dynamics of different games are studied in order to better understand the involved evolutionary processes.

The Prisoner's Dilemma (PD), Snowdrift<sup>1</sup> (SD) and the Stag Hunt (SH) are among the most studied two-person,

symmetric games in the literature. They are used for investigating under which circumstances altruistic traits can become fixed in a population of "selfish" individuals. In social dilemmas of cooperation, individuals' behaviours are of two types: cooperative and non-cooperative. Cooperators are willing to engage in cooperative tasks, while non-cooperators (usually called *defectors*) prefer not to. The success resulting from the interaction of cooperators and defectors is given by the payoff matrix:

	$C$	$D$
$C$	$R$	$S$
$D$	$T$	$P$

where  $C$  denotes cooperators and  $D$  denotes defectors.  $R$  is the *reward* for mutual cooperation,  $P$  is the *punishment* for mutual defection,  $T$  is the *temptation* to defect and  $S$  is the *sucker's payoff*.

In all three social dilemmas, mutual cooperation is favoured over both mutual defection ( $R > P$ ) and an equal probability of unilateral cooperation and defection ( $2R > T + S$ ). The three dilemmas however differ in their ordering of payoffs. In the PD,  $T > R > P > S$ ; in SD,  $T > R > S > P$ , and in the SH,  $R > T > P > S$ .

The evolution of cooperation can be studied by looking at the stable equilibria of the replicator dynamics for each of these games. In the PD, the only stable equilibrium occurs when the population is entirely comprised of defectors. In the SD game cooperators and defectors coexist in equilibrium. In the SH there are two equilibria: when all individuals cooperate and when all individuals defect. This last equilibrium is however risk dominant, i.e. it has the largest basin of attraction.

The replicator dynamics is a rough approximation of actual cultural evolutionary dynamics as it assumes that populations are very large and well-mixed, and that payoff-based bias is the sole psychological mechanism guiding cultural transmission processes. More realistic models of cultural evolutionary processes correct at least one of these assumptions and arrive to different results than those predicted by the standard replicator dynamics.

<sup>1</sup>Also known as Hawks-Doves or Chicken.

Evolutionary graph-theoretical models (Lieberman et al., 2005; Szabo and Fath, 2007), for instance, go beyond the assumption of a large, well-mixed populations by restricting interaction and imitation to near neighbours in a graph representing spatial locality or a social network. In many cases, this graph structure has been shown to promote cooperation beyond the limits of the replicator dynamics in a well-mixed population (Nowak and May, 1992; Nowak et al., 1994; Skyrms, 2003; Santos and Pacheco, 2005).

Other researchers have augmented cultural evolution models by including other psychological biases that, together with payoff-based bias, could influence the way people imitate. In particular, *conformism* or *conformist bias* (Boyd and Richerson, 1985), which is the propensity for preferentially imitating common behaviours, has been suggested to be an important component of our social learning psychology (Asch, 1951; Coultas, 2004).<sup>2</sup> When conformist transmission is introduced in cultural evolution models, the result (in the case of large, well-mixed populations) is a *modified* replicator dynamics that can lead to different equilibrium points and different dynamics than those predicted by the standard replicator dynamics (Henrich, 2001; Skyrms, 2005). By making use of such equation, Henrich and Boyd (2001) have shown how even small amounts of conformism are able to stabilise cooperative behaviour in a public goods game if punishment is also included in the model. In related work, Skyrms (2005) has explored the effect of conformist bias in a number of symmetric two-by-two games. Analyses in that work were however restricted to some specific numerical cases and no general conclusions were formally drawn.

The aim of this paper is to study the effects of conformist transmission on the evolution of cooperation when considering two-person symmetric games such as the PD, SD and the SH. We propose an evolutionary graph-theoretical model in which cultural transmission is guided by both payoff-based and conformist biases, and study it both analytically and by means of simulation.

The paper is organised as follows. The next section gives the agent-based level specifications of the model. It is then shown how to recover the modified replicator dynamics in the limiting case of a large and well-mixed population, and the equation is studied by means of equilibrium analysis. This is followed by a simulation study of the particular case of a population organised into a regular 2D lattice. Finally, conclusions are drawn.

<sup>2</sup>From an evolutionary psychology perspective, conformist bias could have evolved because it is adaptive in the face of costly information. Boyd and Richerson (1985) and Henrich and Boyd (1998) have theoretically showed that conformist transmission is adaptive in spatially and/or temporally varying habitats since it provides a simple heuristic rule that increases the probability of acquiring locally adaptive beliefs and behaviours.

## The Model

Our model considers a population of  $n$  individuals, where the  $i$ -th individual is represented by the vertex  $v_i$  of an undirected graph  $G(V, E)$  with  $v_i \in V \forall i$ . The open neighbourhood of  $i$ ,  $N(i)$ , is the set of all individuals  $j$  such that there is an edge  $e_{ij} \in E$ . The number of neighbours of individual  $i$  is thus the degree  $k_i$  of vertex  $v_i$ . The closed neighbourhood  $N[i]$  is the set of  $i$ 's neighbours plus  $i$  itself.

Each individual is characterised by its *cultural trait* or *strategy*  $s_i \in \{A, B\}$ . Social interaction is modelled by means of a two-person, symmetric game with a payoff matrix  $M$  given by<sup>3</sup>:

$$\begin{array}{c|cc} & A & B \\ \hline A & a & b \\ B & c & d \end{array}$$

Each time step  $t$ , individuals simultaneously engage in social interactions. As a result of these interactions, individual  $i$  collects an *average payoff* given by:

$$u_i(t) = \frac{1}{k_i} \sum_{j \in N(i)} M(s_i(t), s_j(t)).$$

After interactions are completed, individual  $i$  randomly chooses one of its neighbours  $j \in N(i)$  as its *model* for cultural transmission. Imitation is assumed to be conformist-biased with probability  $\alpha$  and payoff-biased with probability  $1 - \alpha$ . Parameter  $\alpha$  thus weighs the importance of conformism relative to payoff-biased transmission.

The adoption of individual  $j$ 's strategy by the focal individual  $i$  depends on  $j$ 's *cultural fitness*  $w_{ij}$ . Cultural fitness (the direct analogue to biological fitness in genetic evolution) is a measure of the attractiveness or the transmissibility of a model's strategy. If transmission is payoff-biased,  $j$ 's cultural fitness is given by the difference of average payoffs between  $j$  and  $i$ :

$$w_{ij}(t) = u_j(t) - u_i(t).$$

If transmission is conformist,  $j$ 's cultural fitness is given by

$$w_{ij}(t) = q_{ij}(t) - \frac{1}{2},$$

where  $q_{ij}$  is the proportion of agents in  $N[i]$  having the same strategy as  $j$ . Notice that  $w_{ij}$  is positive whenever  $u_j > u_i$  (payoff-biased transmission) or  $j$  holds the strategy held by the majority of  $i$ 's neighbours (conformist transmission).

Agent  $i$  copies  $j$ 's strategy with a probability proportional to  $w_{ij}$ . Formally:

$$Pr(s_i(t+1) = s_j(t)) = f(w_{ij}),$$

<sup>3</sup>Without loss of generality, payoffs are assumed to be non-negative values.

where  $f$  is assumed to be a monotonically increasing function, in order for models with high cultural fitness to propagate their strategies more often than models with low cultural fitness. Three alternative definitions of  $f$  are considered in this paper, each one specifying a different imitation rule: (i) imitate-if-better (IIB); (ii) replicator dynamics 1 (RD1); and (iii) replicator dynamics 2 (RD2).<sup>4</sup>

The IIB rule is given by:

$$f_{IIB}(w_{ij}) = \begin{cases} 0 & \text{if } w_{ij} \leq 0 \\ 1 & \text{if } w_{ij} > 0 \end{cases},$$

whereas RD1 and RD2 are respectively defined by:

$$f_{RD1}(w_{ij}) = \begin{cases} 0 & \text{if } w_{ij} \leq 0 \\ \beta w_{ij} & \text{if } w_{ij} > 0 \end{cases},$$

and

$$f_{RD2}(w_{ij}) = \frac{1}{2}(1 + \beta w_{ij}).$$

Parameter  $\beta$  normalises  $w_{ij}$  such that  $0 \leq Pr(s_i(t+1) = s_j(t)) \leq 1$ . Thus,  $\beta = 2$  in the case of conformist transmission and

$$\beta = \frac{1}{\max\{a, b, c, d\} - \min\{a, b, c, d\}} \quad (1)$$

in the case of payoff-biased transmission. Fig. 1 depicts  $f$  for each imitation rule.

The three imitation rules described above have been traditionally used in the literature, either directly in evolutionary graph-theoretical models (e.g. RD1 by Hauert and Doebeli (2004) and Santos and Pacheco (2005)) or in order to derive population-level analytical models (e.g. RD2 by Henrich (2001) and Boyd and Richerson (2002)).

From the previous definitions it is possible to derive  $Pr(s_i(t+1) = A)$ , which is the probability of individual  $i$  holding strategy  $A$  at time step  $t+1$  after having chosen a neighbour  $j$  as a model. Individual  $i$ 's strategy will become or remain  $A$  whenever: *a*)  $A$  is the current strategy of both  $i$  and  $j$ ; *b*)  $i$ 's current strategy is  $A$ ,  $j$ 's current strategy is  $B$ , but  $i$  does not imitate  $j$ ; or *c*)  $i$ 's current strategy is  $B$ ,  $j$ 's current strategy is  $A$ , and  $i$  imitates  $j$ . The formal equation is shown in Fig. 2.

## Exact analysis for the case of large, well-mixed populations

### General games

Here we analyse the limiting case of a complete graph with large  $n$ , which is equivalent to having the large, well-mixed population that is traditionally assumed in standard evolutionary game theory.

<sup>4</sup>We give RD1 and RD2 these names because both imitation rules can be shown to recover the replicator dynamics in the well-mixed, 100% payoff-biased transmission case (Gintis, 2000; McElreath and Boyd, 2007).

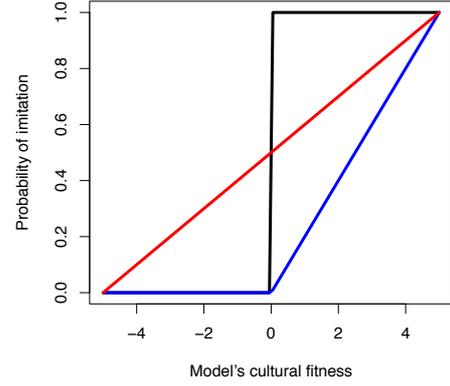


Figure 1: Imitation rules. IIB is shown in *black*, RD1 ( $\beta = 0.2$ ) in *blue* and RD2 ( $\beta = 0.2$ ) in *red*.

Let  $p_t$  denote the frequency of individuals with strategy  $A$  at time step  $t$ . For a complete graph with  $n \rightarrow \infty$ ,  $k_i = n - 1 \approx n \forall i$ , and

$$u_i(t) = \begin{cases} u_A(t) & \text{if } s_i(t) = A \\ u_B(t) & \text{if } s_i(t) = B \end{cases}$$

$\forall i$ , where  $u_A(t)$  and  $u_B(t)$  are the average payoffs collected by individuals with strategies  $A$  and  $B$  at time step  $t$ , respectively given by

$$u_A(t) = ap_t + b(1 - p_t), \quad (2)$$

and

$$u_B(t) = cp_t + d(1 - p_t). \quad (3)$$

Additionally, since  $N[i] = V \forall i$ :

$$q_{i,j}(t) = \begin{cases} p_t & \text{if } s_j(t) = A \\ 1 - p_t & \text{if } s_j(t) = B \end{cases} \quad \forall i, j.$$

Using these relations and RD2 as imitation rule, the equation of Fig. 2 can be shown to reduce to:

$$\Delta p = p_t(1 - p_t)\{(1 - \alpha)\beta[u_A(t) - u_B(t)] + \alpha(2p_t - 1)\}, \quad (4)$$

where  $\Delta p = p_{t+1} - p_t$  is the change in the proportion of individuals with behaviour  $A$  between time steps  $t$  and  $t + 1$ . The recursion of Eq. 4 is a modified replicator dynamics that had been already derived in related work on cultural transmission processes including both payoff-biased and conformist imitation (Henrich and Boyd, 2001; Henrich, 2001; Carpenter, 2004; Skyrms, 2005).

Let us first analyse the particular case when cultural transmission is payoff-biased only. Making  $\alpha = 0$ , Eq. 4 reduces to:

$$\Delta p = p_t(1 - p_t)\beta\{u_A(t) - u_B(t)\},$$

$$\begin{aligned}
Pr(s_i(t+1) = A) &= Pr(s_i(t) = A, s_j(t) = A) (1) \\
&+ Pr(s_i(t) = A, s_j(t) = B) \left\{ (1-\alpha) [1 - f(u_j(t) - u_i(t))] + \alpha [1 - f(q_{ij}(t) - \frac{1}{2})] \right\} \\
&+ Pr(s_i(t) = B, s_j(t) = A) \left\{ (1-\alpha) [f(u_j(t) - u_i(t))] + \alpha [f(q_{ij}(t) - \frac{1}{2})] \right\}
\end{aligned}$$

Figure 2: Probability of individual  $i$  having strategy  $A$  at time step  $t + 1$  after cultural transmission from model  $j$

which is the discrete-time equivalent of the standard replicator dynamics (Taylor and Jonker, 1978; Hofbauer and Sigmund, 1998; Gintis, 2000). Substituting Eq. 2 and 3 in the last expression and doing little algebra:

$$\Delta p = p_t(1-p_t)\beta \{(a-b-c+d)p_t + b-d\}. \quad (5)$$

Equilibria of this equation can be found by looking at the values of  $p_t$  that make  $\Delta p = 0$ . The two *pure equilibria* are given by  $p_t = 0$  and  $p_t = 1$ . In the following, these equilibria will be respectively called all- $B$  and all- $A$ . A third *internal equilibrium*, in which players with strategies  $A$  and  $B$  are present in the population, may exist. When this is the case, the proportion of individuals with strategy  $A$  in equilibrium is given by

$$p^* = \frac{d-b}{(a-c) + (d-b)}.$$

In general, the equilibrium  $p$  is stable<sup>5</sup> whenever

$$\left| \frac{dp_{t+1}}{dp_t} \right|_{p_t=p} < 1.$$

From this, it can be easily shown that

- all- $B$  is stable when  $b < d$ ,
- all- $A$  is stable when  $a > c$ , and
- $p^*$  is stable when both  $a < c$  and  $b > d$ .

Depending on the ranking of the entries of the payoff matrix, four different possibilities<sup>6</sup> for the imitation dynamics can thus be distinguished (Nowak, 2006):

1.  $a > c \wedge b > d$ : only all- $A$  is stable ( $A$  dominates  $B$ ).
2.  $a < c \wedge b < d$ : only all- $B$  is stable ( $B$  dominates  $A$ ).
3.  $a > c \wedge b < d$ : both all- $A$  and all- $B$  are stable ( $A$  and  $B$  are *bistable*). In this case, the internal unstable equilibrium  $p^*$  determines the sizes of the basins of attraction of the two pure equilibria. The equilibrium with the largest basin of attraction is called *risk dominant*. In particular

<sup>5</sup>The condition is necessary and sufficient for hyperbolic equilibria only. All- $B$  (resp. all- $A$ ) is non-hyperbolic when  $b = d$  (resp.  $a = c$ ).

<sup>6</sup>Actually, there is a fifth possibility:  $A$  and  $B$  are neutral when  $a = c$  and  $b = d$ . In this case there is no evolution since  $\Delta p = 0 \forall p_t$ .

- a) all- $A$  is risk dominant if  $d - b < a - c$ , and
- b) all- $B$  is risk dominant if  $d - b > a - c$ .

4.  $a < c \wedge b > d$ : pure equilibria are unstable and the internal equilibrium is stable ( $A$  and  $B$  coexist).

How this picture changes when cultural transmission has also a conformist component ( $\alpha > 0$ )? In order to answer to this question, an equilibrium analysis similar to the one done in the case  $\alpha = 0$  can be performed here for  $\alpha \neq 0$ . A second possibility is to rewrite Eq. 4 as

$$\Delta p = p_t(1-p_t)\{[(1-\alpha)\beta(a-b-c+d) + 2\alpha]p_t + (1-\alpha)\beta(b-d) - \alpha\},$$

and perform the following variable substitutions

$$\begin{aligned}
a' &= (1-\alpha)\beta a + \alpha, \\
b' &= (1-\alpha)\beta b, \\
c' &= (1-\alpha)\beta c, \\
d' &= (1-\alpha)\beta d + \alpha,
\end{aligned}$$

to obtain:

$$\Delta p = p_t(1-p_t)\{(a' - b' - c' + d')p_t + b' - d'\}. \quad (6)$$

Notice (see Eq. 5) that this recursion is equivalent to the discrete replicator dynamics of a population game with the following payoff matrix  $M'$ :

$$\begin{array}{c|cc}
 & A & B \\
\hline
A & a' & b' \\
B & c' & d'
\end{array}$$

Hence, in the framework of the replicator dynamics, *the addition of conformism to the cultural evolutionary process is equivalent to a transformation of the payoff matrix of the underlying game*. Observe that  $\alpha = 0$  recovers the original game and  $\alpha = 1$  completely transforms the original game into a pure coordination game with the following payoff matrix:

$$\begin{array}{c|cc}
 & A & B \\
\hline
A & 1 & 0 \\
B & 0 & 1
\end{array}$$

The addition of conformism to imitation dynamics can have considerable effects in the nature of equilibria of the modelled cultural evolutionary process (Boyd and Richerson, 1985; Henrich and Boyd, 2001; Henrich, 2001; Skyrms,

2005). In particular, since the entries of  $M$  are non-negative and  $0 \leq \alpha \leq 1$ ,

$$\begin{aligned} a &< c \not\Rightarrow a' < c' \\ b &> d \not\Rightarrow b' > d', \end{aligned}$$

which means that *a*) originally unstable pure equilibria could become stable and *b*) an originally stable internal equilibrium could become unstable. Furthermore, if  $A$  and  $B$  co-exist, the proportion of individuals with strategy  $A$  in equilibrium is now given by

$$p'^* = \frac{(1 - \alpha)\beta(d - b) + \alpha}{(1 - \alpha)\beta\{(a - c) + (d - b)\} + 2\alpha}.$$

Not everything changes in the dynamics of the game when conformism is introduced. In particular,

$$\begin{aligned} a &> c \Rightarrow a' > c', \\ b &< d \Rightarrow b' < d', \end{aligned}$$

which means that originally stable pure equilibria will continue to be stable in the transformed game. Moreover,

$$\begin{aligned} d - b < a - c &\Rightarrow d' - b' < a' - c', \\ d - b > a - c &\Rightarrow d' - b' > a' - c', \end{aligned}$$

which means that, if  $A$  and  $B$  are bistable, the risk dominant equilibrium of the transformed game will be the same as the one of the original game.

The new conditions for stability are

1. All- $B$  is stable if

$$\alpha > \frac{\beta(b - d)}{1 + \beta(b - d)} \quad (7)$$

2. All- $A$  is stable if

$$\alpha > \frac{\beta(c - a)}{1 + \beta(c - a)} \quad (8)$$

3. The internal equilibrium, when it exists, is stable if neither Eq. 7 nor Eq. 8 holds.

### Social dilemmas

Let us now focus on the effect of conformist biases in games reflecting social dilemmas, such as the PD, SD and the SH. In order to simplify the analysis for these games, it is customary to rescale their payoff matrices so that they depend on a single parameter. For the PD, we follow Nowak and May (1992) and make  $T = b$ ,  $R = 1$ ,  $P = \epsilon \approx 0$  and  $S = 0$ , where  $1 < b < 2$  characterises the advantage of defectors against cooperators. For the SD game, we follow Hauert and Doebeli (2004) and make  $T = \gamma > 1$ ,  $R = \gamma - 1/2$ ,  $S = \gamma - 1$  and  $P = 0$ , such that the cost-to-benefit ratio

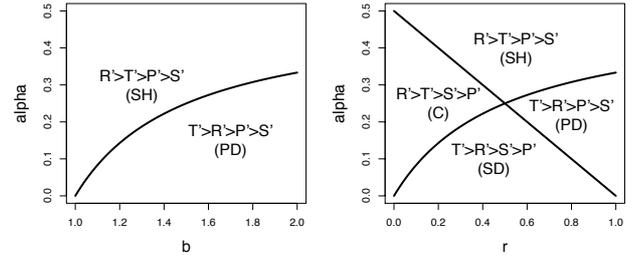


Figure 3: Effect of conformist bias in the PD (*left*) and the SD game (*right*).

of mutual cooperation is given by  $r = 1/(2\gamma - 1)$ , with  $0 \leq r \leq 1$ . For the SH we make  $T = P = 1$ ,  $R = g$  and  $S = 0$ , with  $1 < g < 2$ . With these settings,  $\beta = 1/b$  for the PD,  $\beta = 1/\gamma$  for SD and  $\beta = 1/g$  for the SH in the case of payoff-based biased imitation (see Eq. 1).

As it has been previously analysed, the effect of conformist transmission may be interpreted as a transformation in the payoff matrix that can alter the original ordering of its entries. This in turn can drastically change the nature of the game played. In the PD with conformism, the all- $C$  equilibrium (unstable in the original game) can become stable if  $R' > T'$ . This holds when

$$\alpha > \frac{b - 1}{2b - 1}.$$

The resulting ordering of the payoffs ( $R' > T' > P' > S'$ ), and the fact that all- $D$  is always the risk-dominant equilibrium, effectively converts the game into a SH (see Fig. 3).

In the case of the SD game, the ordering of the entries of the transformed payoff matrix  $M'$  can be different from that of the original matrix  $M$  if  $R' > T'$  (all- $C$  becomes stable),  $P' > S'$  (all- $D$  becomes stable) or both conditions hold. For the rescaled version of this game,  $R' > T'$  whenever

$$\alpha > \frac{r}{1 + 2r},$$

and  $P' > S'$  when

$$\alpha > \frac{1 - r}{2}.$$

There are thus 4 different possibilities for the SD game with conformist transmission (see Fig. 3):

1.  $T' > R' > S' > P'$  (the game is still a SD),
2.  $R' > T' > S' > P'$  ( $C$  dominates  $D$ ),
3.  $T' > R' > P' > S'$  (the game becomes a PD), and
4.  $R' > T' > P' > S'$  (the game becomes a SH). In this last case the game is a proper SH ( $C$  and  $D$  are bistable and all- $D$  is the risk-dominant equilibrium) when

$r > 0.5$ . When  $r < 0.5$ , all- $C$  is both payoff and risk dominant.

Finally, in the case of the SH the ordering of the payoffs is not importantly affected, but the unstable equilibrium moves towards  $p = 1/2$ , thus reducing the basin of attraction of all- $D$ , i.e. the riskiness of all- $C$ .

Broadly speaking, conformist transmission can promote cooperation in the PD by turning it into a SH, and in the SH by diminishing the basin of attraction of all- $D$ . In the SD game, results are dependent on the cost-to-benefit ratio of mutual cooperation. For  $r < 0.5$ , cooperation is generally favoured: all- $C$  can become the only stable equilibrium (when  $R' > T' > S' > P'$ ), or the risk dominant equilibrium (when  $R' > T' > P' > S'$ ). For  $r > 0.5$  the opposite happens, with all- $D$  possibly becoming the only stable equilibrium (when  $T' > R' > P' > S'$ ) or the risk-dominant equilibrium (when  $R' > T' > P' > S'$ ).

Although conformist transmission opens the possibility of a cooperative equilibrium in the PD and diminishes the riskiness of engaging in cooperative actions in the SH, populations with an initial majority of defectors are always doomed to a non-cooperative equilibrium in these two games. In the SD case, defection prevails for  $r > 0.5$ , and this for any amount of conformism. In this sense, conformist transmission alone is unable to sustain cooperation in both PD and SH, and it promotes cooperation for the SD game only when  $r < 0.5$ . For cooperation to be sustained, other mechanisms are necessary to be present along with conformism. Punishment has been suggested as one such possible mechanism (Henrich and Boyd, 2001). In the next section, we explore another mechanism: graph reciprocity.

### Simulation results for the case of medium-sized, spatially structured populations

Here, the evolutionary dynamics of the three social dilemmas discussed above are studied by means of computer simulations for the case of medium-sized populations (1024 individuals) organised into a  $32 \times 32$  square lattice with periodic boundary conditions. For the three games, the rescaled versions presented in the last section were used<sup>7</sup>.

Square lattices were implemented using both Moore and von Neumann neighbourhoods with ranges equal to 1. Simulations were conducted using each of the three imitation rules previously defined (IIB, RD1, RD2), varying values of the game parameters ( $b$  in the PD,  $r$  in SD and  $g$  in the SH) and different amounts of conformism ( $\alpha \in \{0.0, 0.125, 0.25, 0.375, 0.5\}$ ). Agents were updated synchronously.

For each simulated condition, 50 runs were executed. Each simulation was initialised with 50% cooperators and

terminated whenever the population converged to any of the two absorbing states (all- $C$ , all- $D$ ) or after 3000 simulation steps. In this last case, the equilibrium proportions of cooperators were calculated by averaging over the last 1000 time steps of each run, well after transients have passed.

Fig. 4 shows the average level of cooperation in equilibrium for the Moore neighbourhood case. Results for the von Neumann neighbourhood case are qualitatively similar and are not reproduced here for reasons of space. In the figures corresponding to the SD game, the dashed lines represent the equilibrium fraction of cooperators predicted by Eq. 6 (the well-mixed case).

Fig. 4 shows how cultural transmission including a conformist component consistently promotes higher levels of cooperation than payoff-based biased transmission alone for both the PD and the SH. Moreover, the larger the amount of conformism, the larger the proportion of cooperators at equilibrium, as it can be seen from the nice ordering of the curves for different values of  $\alpha$ . For the SD game, the addition of conformist bias results in higher frequencies of cooperators for small  $r$  but also in lower frequencies of cooperators for large  $r$ . Thus, the general observations made for the effects of conformist transmission on the well-mixed case continue to hold for the case of spatially structured populations, i.e. that conformism promotes cooperation in the PD and the SH for the whole range of their game parameters, and that it promotes cooperation in the SD game for  $r < 0.5$  while inhibiting cooperation for  $r > 0.5$ .

Regarding the effects of embedding the population in a lattice, our results confirm those already classic in evolutionary game theory: spatial structure promotes cooperation in the PD (Nowak and May, 1992; Nowak et al., 1994) and the SH (Skyrms, 2003), but can inhibit cooperation in the SD game (Hauert and Doebeli, 2004). In general, for the SD game, cooperators in a lattice do better than their counterparts in a well-mixed population for a)  $\alpha < 0.25$  and small  $r$ , and b)  $\alpha > 0.25$  and large  $r$ .

Notice that these qualitative results do not depend on the specific imitation rule being used. However, quantitative results do depend on the specificities of these rules. For instance, the higher stochasticity of the RD2 with respect to the other two imitation rules seems to hinder the evolution of cooperation in the PD and SH games, where only moderate levels of cooperation can be sustained, and only for very small  $b$  or very large  $g$ .

### Conclusions

We have augmented traditional evolutionary graph-theoretic models with conformist transmission (the tendency to imitate common behaviours) and studied the effects of this extension on the evolutionary dynamics of social dilemmas. From a replicator dynamics perspective, the addition of conformism is equivalent to a simple transformation of the payoff matrix favouring the stability of pure equilibria. In par-

<sup>7</sup>We effectively set  $P = \epsilon = 0$  in the PD.

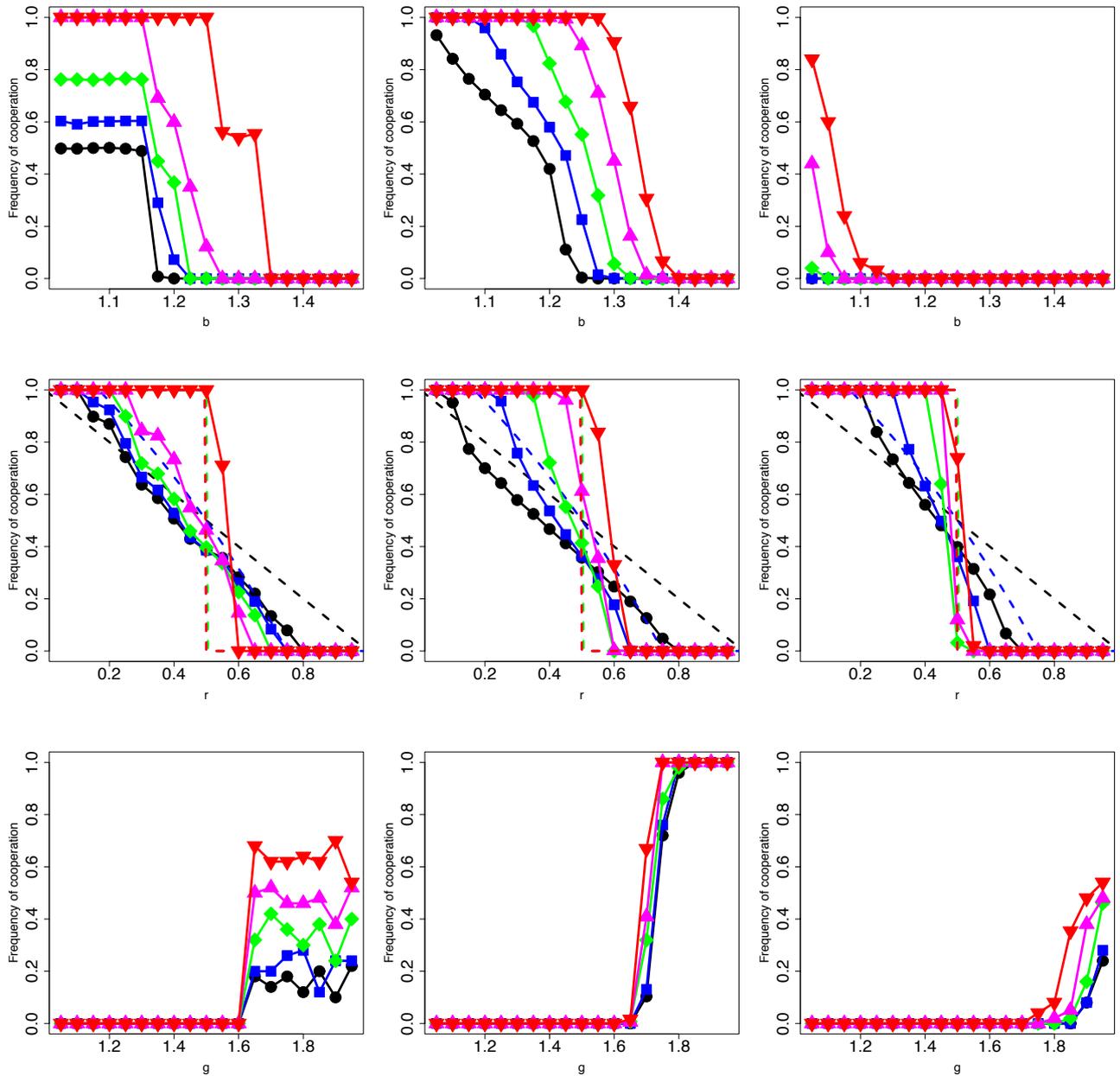


Figure 4: Average values of the equilibrium proportion of cooperators as a function of the game parameter for the PD (*first row*), the SD game (*second row*) and the SH (*third row*). Results are given for IIB (*first column*), RD1 (*second column*) and RD2 (*third column*) imitation rules and different amounts of conformism:  $\alpha = 0.0$  (*black*),  $\alpha = 0.125$  (*blue*),  $\alpha = 0.25$  (*green*),  $\alpha = 0.375$  (*magenta*) and  $\alpha = 0.5$  (*red*). For the SD game, the corresponding proportions of cooperators in well-mixed populations for each value  $\alpha$  are also reported (*dashed lines*).

ticular, a Prisoner's Dilemma can become a Stag Hunt, and a Snowdrift can become a Stag Hunt, a Prisoner's Dilemma or a game in which cooperation dominates defection. In the Stag Hunt case, where both pure equilibria are already stable, conformist transmission moves the unstable equilibrium towards  $p = 1/2$ , thus reducing the basin of attraction of the non-cooperative equilibrium. Although unable to sustain cooperation by its own when cooperators are not the majority at the beginning of the evolutionary process, conformist transmission enhances cooperation when other mechanisms, such as spatial locality, are also present in the model, at least for the PD and the SH cases. For the spatial SD, conformism can also be shown to promote higher levels of cooperative behaviour, but only for small cost-to-benefit ratios.

**Acknowledgements.** This work is funded by the Future and Emerging Technologies programme IST-STREP of the European Community, under grant number IST-034632 (PERPLEXUS). The author would like to thank Isis Fuchs, Olivier Jorand, Andrés Pérez-Urbe and Henri Volken for valuable discussions, and the three anonymous reviewers for their comments on this paper.

### References

- Asch, S. E. (1951). *Groups, leadership and men*, chapter Effects of group pressure upon the modification and distortion of judgment, pages 177–190. Pittsburgh, PA: Carnegie Press.
- Boyd, R. and Richerson, P. J. (1985). *Culture and the evolutionary process*. Chicago: University of Chicago Press.
- Boyd, R. and Richerson, P. J. (2002). Group beneficial norms can spread rapidly in a structured population. *Journal of Theoretical Biology*, 215(3):287–296.
- Carpenter, J. P. (2004). When in rome: conformity and the provision of public goods. *Journal of Socio-Economics*, 33(4):395–408.
- Coultas, J. C. (2004). When in rome... an evolutionary perspective on conformity. *Group Processes Intergroup Relations*, 7(4):317–331.
- Gintis, H. (2000). *Game theory evolving*. Princeton: Princeton University Press.
- Hauert, C. and Doebeli, M. (2004). Spatial structure often inhibits the evolution of cooperation in the snowdrift game. *Nature*, 428(6983):643–646.
- Henrich, J. (2001). Cultural transmission and the diffusion of innovations: Adoption dynamics indicate that biased cultural transmission is the predominate force in behavioral change and much of sociocultural evolution. *American Anthropologist*, 103:992–1013.
- Henrich, J. and Boyd, R. (1998). The evolution of conformist transmission and between-group differences. *Evolution and Human Behavior*, 19:215–242.
- Henrich, J. and Boyd, R. (2001). Why people punish defectors: conformist transmission stabilizes costly enforcement of norms in cooperative dilemmas. *Journal of Theoretical Biology*, 208:79–89.
- Hofbauer, J. and Sigmund, K. (1998). *Evolutionary Games and Population Dynamics*. Cambridge: N. Y.
- Lieberman, E., Hauert, C., and Nowak, M. A. (2005). Evolutionary dynamics on graphs. *Nature*, 433(7023):312–316.
- Maynard Smith, J. (1982). *Evolution and the Theory of Games*. Cambridge Univ. Press, Cambridge.
- McElreath, R. and Boyd, R. (2007). *Mathematical Models of Social Evolution: A Guide for the Perplexed*. University Of Chicago Press.
- Nowak, M. (2006). *Evolutionary Dynamics*. Cambridge, MA: Harvard University Press.
- Nowak, M., Bonhoeffer, S., and May, R. (1994). Spatial Games and the Maintenance of Cooperation. *Proceedings of the National Academy of Sciences of the United States of America*, 91(11):4877–4881.
- Nowak, M. A. and May, R. M. (1992). Evolutionary games and spatial chaos. *Nature*, 359(6398):826–829.
- Santos, F. C. and Pacheco, J. M. (2005). Scale-free networks provide a unifying framework for the emergence of cooperation. *Phys. Rev. Lett.*, 95(9):098104–4.
- Skyrms, B. (2003). *The Stag-Hunt Game and the Evolution of Social Structure*. Cambridge: Cambridge University Press.
- Skyrms, B. (2005). Dynamics of conformist bias. *The Monist*, 88(2):259–269.
- Szabo, G. and Fath, G. (2007). Evolutionary games on graphs. *Physics Reports*, 446(4-6):97–216.
- Taylor, P. and Jonker, L. (1978). Evolutionary stable strategies and game dynamics. *Mathematical Biosciences*, 40:145–156.