



## The Evolution of Degrees of Cooperation

MARCUS FREAN

*Artificial Intelligence Laboratory, Departments of Computer Science and Zoology,  
University of Otago, Dunedin, New Zealand*

(Received on 11 January 1996, Accepted on 14 June 1996)

The Prisoner's Dilemma has been widely studied as a model for the evolution of cooperation, and most of this work has dealt with agents who either cooperate or not. In this paper we look at the consequences of allowing agents to have intermediate levels of cooperation, and to update these levels over time. The familiar strategy of "tit for tat" emerges as a robust mode of behaviour, yet there are important differences between this case and that of "all or nothing" cooperation.

© 1996 Academic Press Limited

### 1. Introduction

The Prisoner's Dilemma is widely used in models for the evolution of co-operation in the presence of selfish incentives. In such evolutionary models interactions between individuals take the form of "games" in which each participant can choose either to cooperate or to defect (not cooperate) at any one time. The two players then receive pay-offs amounting to fitnesses which depend on their joint behaviour. By choosing appropriate values for these pay-offs one can model situations in which cooperation is best for the group but not for the individual.

In a Prisoner's Dilemma, if both players cooperate they each receive a "reward" pay-off  $R$  (say 3 or 4) whereas if they both defect they get a lesser "punishment" pay-off  $P$  (usually set at 1). Thus, mutual cooperation is favoured over mutual defection. The so-called dilemma arises in situations where defecting against a cooperator gives a "temptation" pay-off  $T$  which is higher than  $R$ , and the cooperator in this situation is made a "sucker" and received a meagre  $S$  points ( $T = 5$  and  $S = 0$  being typical). In general the Prisoner's Dilemma is characterised by two relationships amongst these pay-offs:  $T > R > P > S$  and  $2R > T + S$ . The first ensures that defection is favoured over cooperation, while the second gives it a paradoxical quality: the combined total pay-off is highest for mutual cooperation. In fact

there is no "paradox" or "dilemma", in that it is always better to defect in a one-shot game. That is, if the other player chooses to cooperate, the pay-off for defection is greater than for cooperation ( $T > R$ ), and if the other player defects, it is better to defect than to cooperate and be made a sucker ( $P > S$ ). In a Prisoner's Dilemma situation, no matter what the other player has chosen to do, defecting brings the higher reward.

### 2. Modeling the Evolution of Cooperation

There are many examples of animals which cooperate even though they appear to be receiving Prisoner's Dilemma pay-offs (Dugatkin, 1991; Milinski, 1987; Godard, 1993; Wilkinson, 1984; Hart & Hart, 1992), yet such behavior would not be expected to survive Darwinian selection pressure for long. One answer is that in many of these cases the "game" is not one-shot but consists of a series of interactions over time, and that the animals concerned have some memory for previous decisions and their outcomes. One now deals with the evolution of strategies, in that decisions for the current round are conditioned on the previous round or rounds. There are strategies in this iterated Prisoner's Dilemma which are cooperative and yet able to resist defectors, the best known of which is "Tit for tat" (TFT) (Axelrod, 1984; Axelrod

& Hamilton, 1981). TFT starts by cooperating and thereafter simply copies what the other player did last time: it is a “memory 1” strategy in that it reacts only to the other player’s latest action. Despite this simplicity it is very successful, being virtually unique in its ability to invade (that is, take over a population of) the strategy “always defect” (AD). However TFT suffers in situations with “noise”. In games between two TFT’s, a single accidental defection leads to an endless cascade of reprisals and counter-reprisals (Nowak & Sigmund, 1990). TFT is extremely provokable (which after all is what enables it to avoid exploitation by AD and thereby invade it), however this very property prevents it cooperating with itself in uncertain environments. Once TFT is established, a better strategy to adopt is “generous tit for tat”, which reciprocates cooperation but does not retaliate against defection every time. Instead, with some probability it cooperates even when the other player has defected. Provided this probability isn’t too high, it combines Strict-TFT’s robustness against defectors with an ability to cooperate with itself.

Nowak & Sigmund (1992) showed how Generous-TFT could evolve in simulations. They modelled an infinite population made up of a number of strategists (typically less than ten at any one time), each of which consisted of two parameters,  $p_C$  and  $p_D$ . These give the probability of cooperating after a  $C$  (cooperate) or  $D$  (defect) by the other player. Two such strategists can then play one another and accumulate pay-offs over a long game; in the limit of a game of infinite length the pay-off (per time-step) is readily calculated. Population densities of particular strategies were altered using a simple approximation to evolutionary dynamics, with small amounts of new random strategies being added occasionally and nearly extinct ones being weeded out. In these simulations Generous-TFT ( $p_C = 1$  and  $p_D \approx 1/3$  given the usual pay-offs) appeared to be a stable end state, as almost any starting condition converged to it provided the run was long enough. This agreed precisely with theoretical work (Molander, 1985; Nowak & Sigmund, 1990) which showed the “optimal” value of generosity  $p_D$  to be the minimum of  $(R - P)/(T - P)$  and  $1 - (T - R)/(R - S)$ , which is  $1/3$  for  $R = 3$  and  $3/4$  for  $R = 4$ , for example. “Optimal” in this case means that among strategies which are immune to invasion by less cooperative strategies, this is the one with the highest average pay-off when playing itself.

This treatment is readily extended to “memory 2” strategists, in which a player bases their decision not only on the other player’s action but on their own previous action as well. Each player now consists of four probabilities of cooperating,  $p_{CC}$ ,  $p_{CD}$ ,  $p_{DC}$ ,  $p_{DD}$ ,

reflecting the four states (ie. latest decisions) in which it could find itself (using the notation  $p_{\text{co-player, player}}$ ). Note that the memory 1 case is recovered by restricting  $p_{CC} = p_{CD}$  and  $p_{DC} = p_{DD}$ . Nowak & Sigmund (1993) performed simulations of this case in the same way as before and, to their surprise, found a different strategy dominating the long-term behavior. They called this strategy “Pavlov” or “win-stay, lose-shift” (WSLS) because if it receives a good pay-off (either  $T$  or  $R$ ) it repeats its previous action ( $D$  in the former and  $C$  in the latter). Conversely if it receives a low pay-off ( $P$  or  $S$ ) it prefers to change its behavior next time. The WSLS which “evolved” typically had values close to  $p_{CC} = 1$ ,  $p_{DC} = 0$ ,  $p_{CD} = 0$  and  $p_{DD} \approx 0.8$ . In particular this means it usually cooperates after mutual defection and continues to defect against a “sucker”. Another view is that it plays TFT if it cooperated last time, and “anti-TFT” if it defected.

All this work assumed that the agents in question make their decisions simultaneously, whereas this is unlikely to be true for most biological systems. Frean (1994) looked at agents which make their decisions in alternation. That is, player  $A$  makes a decision, player  $B$  responds to it,  $A$  then responds to that, and so on. This goes some of the way towards mimicking situations in which animals are continuously adjusting their decisions on the basis of the other animal’s responses. It might be thought that this would make little difference, but in fact it markedly changes the advantages enjoyed by all the well-known strategies. Among memory 1 strategists, AD now dominates over Generous-TFT (for  $R < 4$ ), and in the memory 2 case a new and more robust strategy displaces WSLS. This “firm but fair” strategy (FBF) is like Generous-TFT but is only generous after mutual defection: after being exploited it retaliates without exception. In this way it exhibits a “guarded” generosity, by venturing cooperation with a co-player whose defection might itself have been purely retaliatory (in other words it reacts differently to unprovoked and provoked defections). Nowak & Sigmund (1994) have dealt with a closely related situation and come to similar conclusions.

### 3. Degrees of Cooperation in the Prisoner’s Dilemma

It may actually be quite rare to find situations in which animals are faced with a choice between simply cooperating or defecting: in many cases it is more accurate to consider an action involving some degree of cooperation, which could be anywhere between 0 and 100%. For example, Milinski (1987) has studied “predator inspection” behavior in fish, in which a

small number (often only a pair) of stickleback approach a potential predator, presumably to ascertain whether it poses a threat or not. The stickleback faces a Prisoner's Dilemma in this situation because on the face of it it is better for an individual fish to hang back rather than to go first, and indeed Milinski has shown that some form of "tit for tat" strategy is used by such fish. In fact predator inspection is not a single decision but the product of many lesser decisions about how much to approach the (possible) predator. It could perhaps be considered the end result of many yes/no decisions made at every step of the approach, but it seems more natural to model the decisions as being ones of degree (how much to lead or follow the other fish). Another example is the voluntary regurgitation of food for hungry vampire bats by other unrelated bats who have managed to find food (Wilkinson, 1984). Some portion of the contents of a full bat's stomach is donated to the other bat, hence it is again a matter of degree and not a simple yes/no decision.

Are the results from the binary case robust if we relax this strong assumption, and consider agents which can adopt degrees of cooperation between zero and one? For instance, does the continuous case also show dominant strategies, do they have the same character, and is timing a crucial factor as it is in binary situations? In order to answer such questions we must first have a representation of what is meant by pay-offs and strategies in the continuous case.

3.1. PAY-OFFS

Suppose we have two agents  $X$  and  $Y$ , who adopt degrees of cooperation  $x$  and  $y$  respectively. We take these degrees to be one for complete cooperation and zero for defection. The joint state of affairs can be pictured as a point  $(x, y)$  in the unit square. The pay-offs  $R, S, T$  and  $P$  refer to the reinforcement received at the four corners of this square. The most natural way to specify the pay-off for values of  $x$  and  $y$  other than zero or one is to simply interpolate between the extremal pay-offs of the binary case. For instance if  $X$  is totally uncooperative, the pay-off  $Y$  receives depends on its degree of cooperation, and the most obvious way to do this is to interpolate linearly between punishment and sucker pay-offs  $P$  and  $S$ . Denoting the pay-off  $Y$  receives by  $G_Y(x, y)$ , we have

$$G_Y(0, y) = P + y(S - P)$$

Similarly if  $X$  cooperates fully,  $Y$ 's pay-off is between temptation  $T$  and reward  $R$ :

$$G_Y(1, y) = T + y(R - T)$$

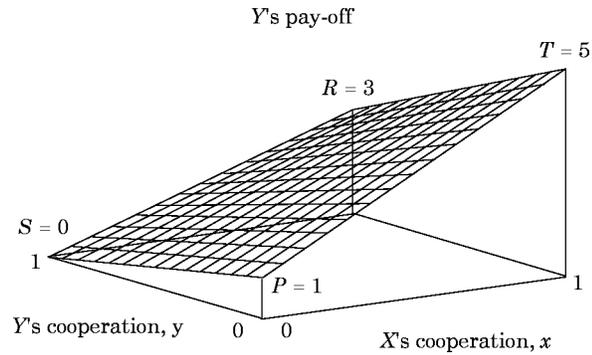


FIG. 1. The pay-off surface for degrees of cooperation. The "dilemma" shows itself in the slope of this surface: no matter what level the  $X$  player adopts,  $Y$ 's pay-off always increases if it cooperates less. However, if both agents are constrained to adopt the same behavior then states are restricted to the line  $x = y$ , and the pay-off slopes upward towards full cooperation: this is the generalisation of the binary preference for mutual cooperation over mutual defection. The first Prisoner's Dilemma condition  $T > R > P > S$  ensures that, for an individual, less cooperation gives a higher pay-off. In general the surface is quadratic, but if  $R + P = S + T$  (for example if  $R = 4$  given the usual values for the other pay-offs) it is simply an inclined plane of constant slope. With  $R = 3$  it is steepest at the "temptation" corner. The pay-off surface for  $X$  is the same as for  $Y$  but with axes reversed.

One can then interpolate between these two to include  $X$ 's level of cooperation:

$$G_Y(x, y) = G_Y(0, y) + x(G_Y(1, y) - G_Y(0, y))$$

It is useful to rewrite this as

$$\begin{aligned} G_Y(x, y) = & Rxy & (1) \\ & + Tx(1 - y) \\ & + S(1 - x)y \\ & + P(1 - x)(1 - y) \end{aligned}$$

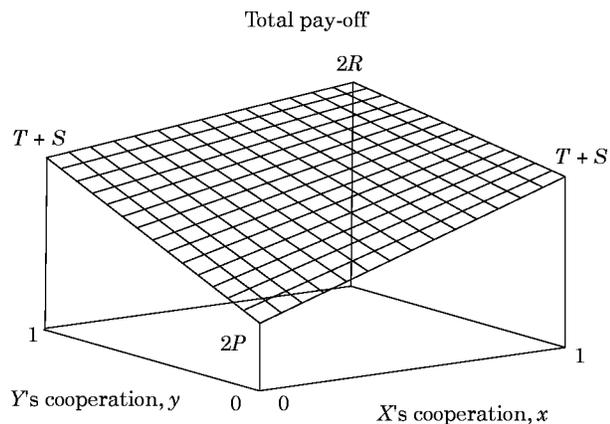


FIG. 2. The sum of both players' pay-offs as a function of  $(x, y)$ . The second Prisoner's Dilemma condition that  $2R > T + S$  ensures that increased cooperation by either player increases the total pay-off. The maximum possible average pay-off is  $R$ , attained only by mutual cooperation.

Figure 1 shows this pay-off surface, while Fig. 2 shows the combined pay-off the players receive in total.

3.2. STRATEGIES

Recall that in the all-or-nothing case a strategy can be specified by the four numbers  $p_{CC}, p_{CD}, p_{DC}$  and  $p_{DD}$ . As with the pay-offs, these can be visualised as the heights of points above the four corners of a unit square representing the present states of the players. In the continuous case one needs to specify a degree of cooperation  $y'$  to follow every possible joint behavior, so a strategy is in effect a surface in this picture:  $y' = f(x, y)$ . In general any single-valued function with a range of zero to one is possible. An obvious simplification to make is to restrict our attention to strategies for which  $f$  amounts to linear interpolation from levels of cooperation specified at the four corners. Suppose that for agent  $Y$  these corner values are denoted  $Y_{11}, Y_{10}, Y_{01}$  and  $Y_{00}$ , where the first index refers to the other player's present behavior and the second to  $Y$ 's (ie.  $Y_{01}$  plays the role of  $p_{DC}$  in the binary case, for instance). In the same manner as for continuous pay-offs,  $Y$ 's response can be written

$$\begin{aligned}
 y' &= Y_{11}xy \\
 &+ Y_{10}x(1 - y) \\
 &+ Y_{01}(1 - x)y \\
 &+ Y_{00}(1 - x)(1 - y).
 \end{aligned}
 \tag{2}$$

4. Memory 1 Strategists

Memory 1 strategists can make no use of their own previous action. They are simply curves mapping the other player's degree of cooperation at time  $t$  to their own at time  $t + 1$ . Writing  $Y_{01} = Y_{00} = Y_0$  and  $Y_{11} = Y_{10} = Y_1$  reflects the fact that  $Y$ 's own previous behavior is not to be taken account of: these are simply the response to  $x = 0$  and to  $x = 1$  respectively. We are then left with linear memory 1 strategies

$$y' = Y_1x + Y_0(1 - x) \tag{3}$$

and  $x' = X_1y + X_0(1 - y)$ .

For example,  $Y_0 = 0, Y_1 = 1$  is the continuous analogue of tit-for-tat, as it simply copies the other player's level.

Behaviors  $x$  and  $y$  could be updated either in synchrony or in alternation: in fact either process converges quickly to the stable point  $y' = y, x' = x$ , ie.

$$y = \frac{1}{\Delta} (Y_0 + X_0(Y_1 - Y_0)) \tag{4}$$

and  $x = \frac{1}{\Delta} (X_0 + Y_0(X_1 - X_0))$

where  $\Delta = 1 - (X_1 - X_0)(Y_1 - Y_0)$ . There is a simple graphical interpretation of this convergence (see Fig. 3). The strategies are straight lines which can both be drawn on the unit square. For instance  $Y_0$  and  $Y_1$  are the intercepts of  $Y$ 's line with the boundaries  $x = 0$  and  $x = 1$ . Note that TFT amounts to the line  $x = y$ . The present joint state is a point which successive updates move towards the intersection of the two lines. This happens regardless of the particular dynamics of the game. For short games one would expect the trajectory to be of some significance since it causes different pay-offs, however for longer games the pay-off at the intersection point dominates the total pay-off, and this is the same in both cases.

We can now relate this situation to the binary case. Consider two binary strategies  $X$  and  $Y$  playing a long game in which their overall probabilities of cooperation  $P(X)$  and  $P(Y)$  are independent. Completely "memoryless" binary strategies cooperate with probabilities which are independent, trivially. However as Nowak & Sigmund (1990) point out, this is

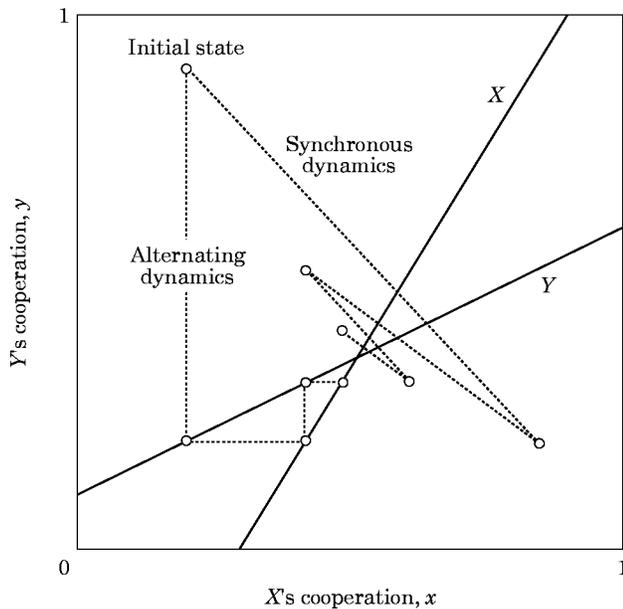


FIG. 3. Convergence of degrees of cooperation in games between linear memory 1 strategists. In the alternating case shown here,  $Y$  goes first, then  $X$ , and so on. In the synchronous case both agents adopt new levels at the same time, so the trajectory is different, however, convergence to the intersection of the lines is rapid and inevitable in either case. Notice that if agents were to update their decisions independently at random times the trajectory is the same as that of the alternating case.

also true of long games between memory 1 strategists making synchronous decisions (note however that it does not hold for games with alternating dynamics, or those between memory 2 strategies). Knowing this, the following must be true:

$$P(Y) = p_c^Y P(X) + p_d^Y(1 - P(X)) \quad (5)$$

and  $P(X) = p_c^X P(Y) + p_d^X(1 - P(Y))$ .

The probability of both  $X$  and  $Y$  cooperating is  $P(X)P(Y)$ , the probability of  $X$  cooperating and  $Y$  defecting is  $P(X)(1 - P(Y))$ , and so on. The average pay-off resulting from such a game must then be

$$\begin{aligned} & RP(X)P(Y) \\ & + TP(X)(1 - P(Y)) \\ & + S(1 - P(X))P(Y) \\ & + P(1 - P(X))(1 - P(Y)). \end{aligned} \quad (6)$$

Comparing eqns (5) and (6) with (3) and (1), we can see that (a) the probabilities converge to the same values as degrees of cooperation and (b) that at these values the pay-offs are the same. The results of the binary probabilistic case (for synchronous decision makers) therefore carry over directly to the linear case (synchronous or not). The dominant strategy must therefore be Generous-TFT: This was confirmed in the simulations described in Section 5.

An agent's pay-off is determined by the intersection of the two lines as in Fig. 3, and we further know that this pay-off has the form of Fig. 1. It is reasonable to ask then, where is the best place to place say  $Y$ 's line so that the intersection gives the highest possible pay-off? Clearly this depends on the line adopted by  $Y$ 's opponent, but it is interesting to note that if either player adopts TFT behavior, it forces the joint state to lie on the line  $x = y$ , for which the pay-off slopes up rather than down with cooperation: TFT thus makes it "selfish" for the other player to cooperate. Note however that if a strategy  $Y$  plays itself the two lines also intersect along  $x = y$ , and hence in the absence of defectors there will always be a temptation towards increased generosity, since it moves the intersection point higher.

Higher levels of generosity mean greater immunity to noise for TFT-like players, resulting in evolution selecting for higher and higher values of  $Y_0$ . This is tempered by the risk of invasion by defectors if the generosity gets too high. One can get a feeling for why  $Y_0$  ends up at the value it does by considering the introduction of a small amount of a completely unreactive strategy into a population dominated by Generous-TFT. Assume the unreactive newcomer

simply adopts a static degree of cooperation  $x$ , irrespective of its opponent's behavior. The maximum pay-off  $G_x$  such a strategy receives when playing a TFT with low generosity is obviously just  $R$ , obtained only for  $x = 1$ , its maximum value. We can calculate the gradient of this pay-off as generosity  $Y_0$  increases, and evaluate it at  $x = 1$ : this comes out to be

$$\left. \frac{\delta G_x}{\delta x} \right|_{x=1} = R - T + (R - S)(1 - Y_0).$$

As expected, against TFT (with  $Y_0 = 0$ ) this gradient is positive (since  $2R > T + S$  is a condition of a prisoner's dilemma), indicating that  $x < 1$  variants cannot earn pay-offs as high as  $R$ . However, above

$$Y_0 = 1 - \frac{T - R}{R - S}$$

it is negative, indicating that there are  $x < 1$  variants which get higher pay-offs than those with  $x = 1$  (which get  $R$ ) and therefore more than  $Y$  itself. If even a small amount of such a variant appears in a population of overly generous TFT strategists it can begin to invade it.

We can also consider the score that AD gets against a majority of Generous-TFT players. AD cooperates at the level  $x = 0$ , while Generous-TFT responds with the level  $Y_0$ . From eqn (1) (note we want the pay-off for  $X$  instead of  $Y$ ) AD receives  $TY_0 + P(1 - Y_0)$ . This increases with  $Y$ 's generosity, reaching  $R$  at

$$Y_0 = \frac{R - P}{T - P}$$

at which point AD can begin to invade.

Thus, the pay-off a  $Y$  player gets when playing itself increases with its generosity, but those with generosity above

$$Y_0^{\text{crit}} = \min\left(1 - \frac{T - R}{R - S}, \frac{R - P}{T - P}\right)$$

are susceptible to invasion even by unresponsive opponents or stalwart defectors. Hence it is not surprising to find the strategy  $Y_1 \approx 1$ ,  $Y_0 = Y_0^{\text{crit}}$  consistently dominating the population in the long term.

### 5. Memory 2 Strategists

Memory 1 strategists are relatively easy to understand but limited: any actual agent would be expected to know its own previous behavior and to use this in determining its subsequent action. Although such memory 2 strategies are to step more realistic, this self-dependence makes them more

difficult to understand. In particular the long-term pay-off between memory 2 agents is problematic to calculate analytically. Instead, we simulate actual games between the two agents.

Section 5.1 describes how the pay-offs in individual games are found, and Section 5.2 describes how strategies can be played off against one another in an “evolutionary tournament”.

### 5.1. SIMULATING LONG GAMES BETWEEN CONTINUOUS STRATEGISTS

Unlike the binary case, interactions between continuous strategists need not lead to a unique stable solution. One (very slow) method of evaluating the pay-offs is to simply run the simulated game for a great many iterations and find the average pay-off for each player. Instead, in the simulations described here each pair of players meets for 50 iterations (each iteration being one decision by both players) and then the pay-offs are averaged over the next ten: these are taken to be the pay-offs that would be received over a very long game. The averaging over ten iterations is much more important for highly nonlinear strategies where cycles or even chaotic behavior are likely.

As in the binary case, the presence of at least some noise in perceptions or in carrying out actions can be expected to play a part in the efficacy of strategies. Rather than model the noise explicitly, its effect is captured here by averaging over starting conditions. Instead of two players playing a single game, they play 25 games starting from initial states  $(0, 0)$ ,  $(0, 0.25)$ ,  $\dots$ ,  $(1, 0.75)$ ,  $(1, 1)$ . This mimics the effect that occasional but drastic errors would have on the average long-term pay-offs attained.

### 5.2. EVOLUTIONARY SIMULATIONS

Following Nowak & Sigmund (1992, 1993), we model a finite number of strategies in a large population. This is done by keeping track of the proportion of each strategy over time: we begin with a single “indifferent” strategy  $Y_{11} = Y_{10} = Y_{01} = Y_{00} = 0.5$  occupying 100% of the population. On average every 100 generations a new randomly generated strategy appears, initially making up 0.2% of the total population. These new strategies are generated by choosing each of the four parameters from the U-shaped distribution  $(\pi\rho(1-\rho))^{-1/2}$ , which favours parameters close to zero or one (a uniform distribution can be used of course, but the simulation then wastes most of its time trying out very poor contenders).

With each new generation the following steps are carried out. Firstly the expected pay-off of each

strategy present is obtained by playing long games against each of the other strategies (and itself). The scores from these games are then weighted by prevalence in the population of the other player. The proportions of all the strategies are then altered by making the reproductive rate of each strategy proportional to the average pay-off it accrues (Taylor & Jonker, 1979; Maynard Smith, 1982), giving

$$\rho'_Y = \frac{\rho_Y G_Y}{\sum_X \rho_X G_X} \quad (7)$$

where  $\rho_Y$  is the proportion and  $G_Y$  is the total expected pay-off of strategy  $Y$ . Equation (7) ensures the growth of any strategy whose total pay-off exceeds the average. Finally any strategy making up less than 0.1% of the population is deemed “extinct” and weeded out.

The overall picture is then as follows: random strategies appear in a model population at random times. Most of these are destined to die out fairly quickly but occasionally one arrives which can increase its proportion  $\rho_Y$  and become the dominant strategy in the population. We are interested in the composition of such a population in the long term.

### 5.3. THE EVOLUTION OF A CONTINUOUS FORM OF GENEROUS-TFT

As an initial check on the simulation, new strategies were restricted to be memory 1. After  $10^6$  generations, the vast majority of surviving strategists were indeed Generous-TFT, with an average generosity exactly that predicted (namely  $1/3$  for  $R = 3$  and  $3/4$  for  $R = 4$ ). As expected, the timing of the interaction did not matter: whether the players updated their states in synchrony or in alternation made no difference to their final pay-offs and hence the dominant strategy.

Figure 4 shows what happens in the memory 2 case for  $R = 3$ . To make the plot, every 2000 generations a strategy was drawn from the population with a probability equal to its prevalence. The data shown here comes from a particular run, but many different starting conditions, different random seeds, and both alternating and synchronous dynamics were tried out, and they all converge rapidly to the same values.

$Y_{11}$  converges to 1.0 (or as close to this value as possible) and remains there.  $Y_{10}$  indicates what level of cooperation an agent adopts immediately after maximally exploiting the other player by completely defecting when they cooperate. “Apologetic” cooperation is strongly favoured, just as it is in simple TFT. Note that this is in sharp contrast to the

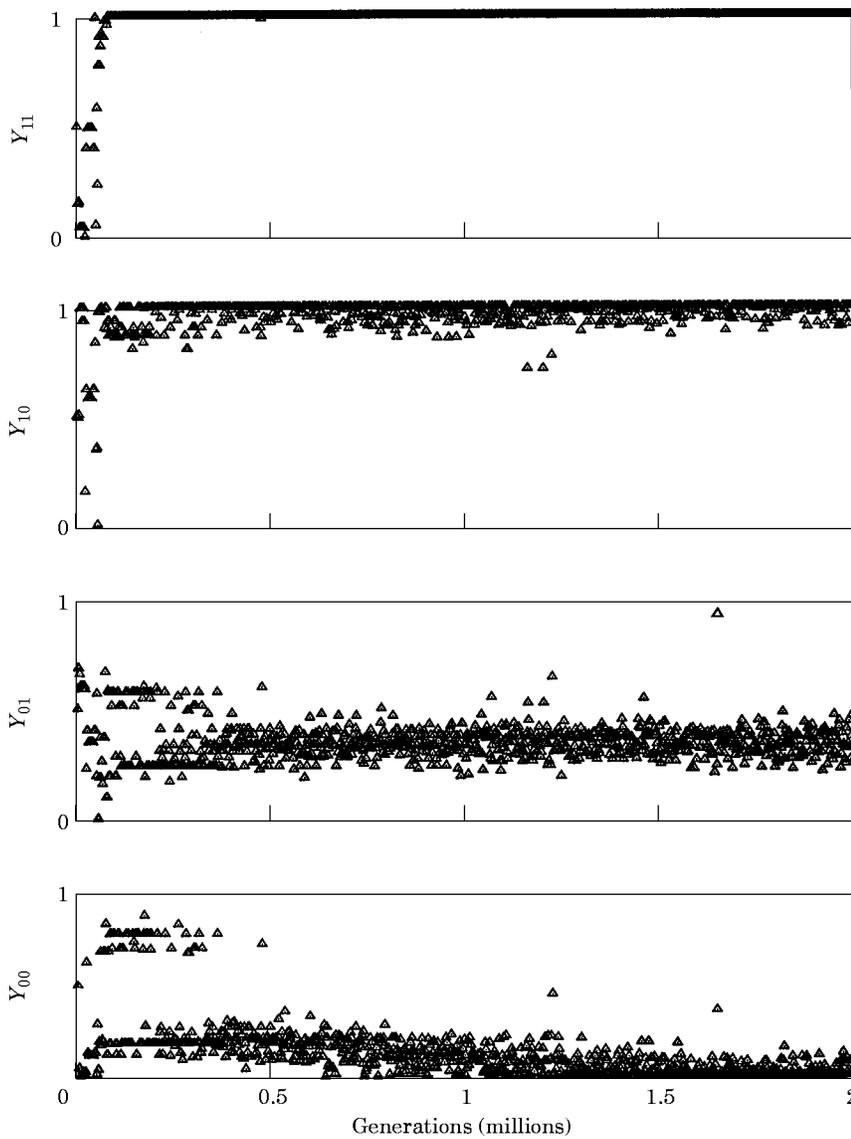


FIG. 4. The timecourse of the evolution of memory 2 strategists.

“win-stay, lose-shift” strategy found in the synchronous binary case, where successful agents continue to exploit a sucker in this situation. Together  $Y_{11}$  and  $Y_{10}$  mean that once cooperation is established, strategies which do not reciprocate with a cooperator are strongly selected against.  $Y_{01}$  denotes the level that follows being made a sucker (ie. where ones own cooperation meets with complete defection). Intriguingly the final value is centered on  $1/3$ , which coincides with the optimal generosity for memory 1 strategists. This was confirmed for  $R = 4$ , i.e. the value always arrives at  $3/4$  in that case. Again this is unlike the all-or-nothing strategies WSLs and FBF, which prefer never to cooperate in this situation. Finally, one

might expect that  $Y_{00}$  would follow  $Y_{01}$  just as  $Y_{10}$  followed  $Y_{11}$  (this would be the familiar Generous-TFT strategy) or perhaps that  $Y_{00}$  would be close to one, as it is in the binary case (WSLS and FBF both have this character). In fact it simply plummets to zero and stays there: mutual defection leads not to a “fresh start” or generosity at all, but only to more defection.

Figure 5 confirms that the average level of cooperation is high and that a large number of (albeit similar) strategies co-exist in the long run.

This highly successful strategy (henceforth denoted  $Y^*$ ) amounts to

$$y' = x + Y_{01}^*(1 - x)y \tag{8}$$

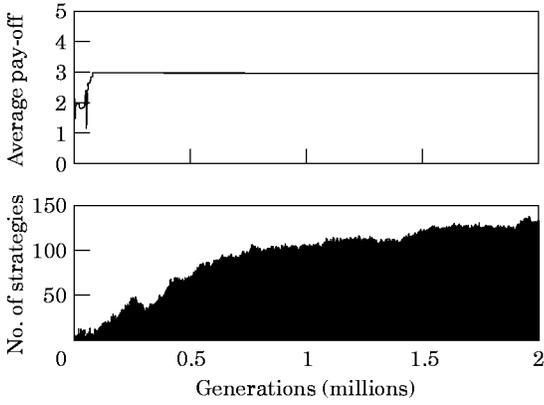


FIG. 5. The timecourse of average pay-off and the total number of strategies present.

[from eqn (2)] and is shown pictorially in Fig. 6. To summarise, if  $Y^*$  was in an uncooperative state before, it responds to  $X$  with Strict-TFT. On the other hand if it was in a cooperative state it responds with Generous-TFT, with a generosity which appears to match that of the simpler memory 1 case.

Consider what happens when  $Y^*$  meets an opponent  $X$  playing Strict-TFT. If  $Y^*$  adopts an initial degree of cooperation  $y$ ,  $X$  responds with  $x = y$ , and  $Y^*$ 's new level is then simply  $x + Y_{01}^*(1 - x)x$ , this being the height of the surface of Fig. 6 along the line  $x = y$ .  $Y^*$  increases its level at each iteration and will reach full mutual cooperation quickly, yet it does so without the risk of being generous at  $x = 0$  (see Fig. 7).

As a second example, consider an interaction between  $Y^*$  and an unreactive player with static level  $x$  independent of  $y$ . Equation (8) then iterates to a final value of  $y = x/(xY_{01}^* - Y_{01}^* + 1)$ , which has the same character (see Fig. 7):  $Y$  behaves like Generous-TFT, but avoids unnecessary generosity with persistent defectors by taking its own previous action into account.

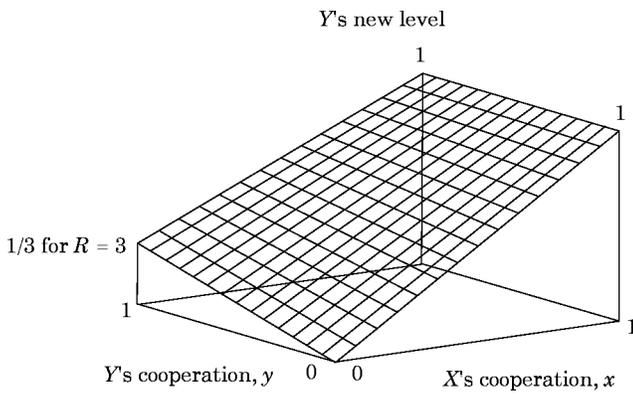


FIG. 6. The evolutionarily stable strategy,  $Y^*$ .

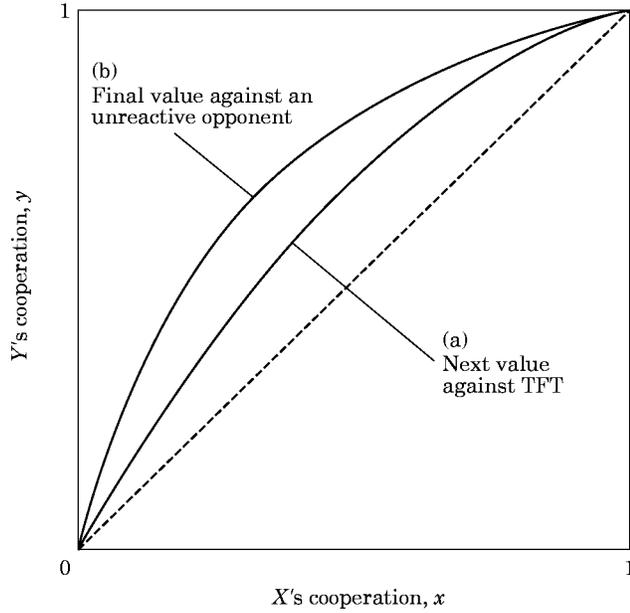


FIG. 7. (a) shows the response of  $Y^*$  to Strict-TFT: the curve shows the new level adopted by  $Y^*$ . The level will quickly rise to full cooperation in a few iterations. (b) shows the long-term level adopted against an unreactive strategy with level  $x$ . Both lines are for  $Y_{01} = 3/4$ , the optimal value where  $R = 4$ . The dotted line is  $y = x$ , to which both curves tend as  $R$  tends towards its lower limit of  $(T + S)/2$  (which is 2.5 given the usual values).

Just as in the memory 1 case, we can consider how the gradient  $\delta G_x / \delta x$  changes with  $Y_{01}$ , given the other values  $Y_{11}^* = Y_{10}^* = 1$ ,  $Y_{00}^* = 0$ . At  $Y_{01} = 0$  (Strict-TFT) the maximum pay-off must be  $R$ , (attained only at  $x = 1$ ), and the gradient evaluated at  $x = 1$  comes out to be exactly the same as that for  $Y_0$  in the memory 1 case, with a critical value at  $1 - (T - R)/(R - S)$ .

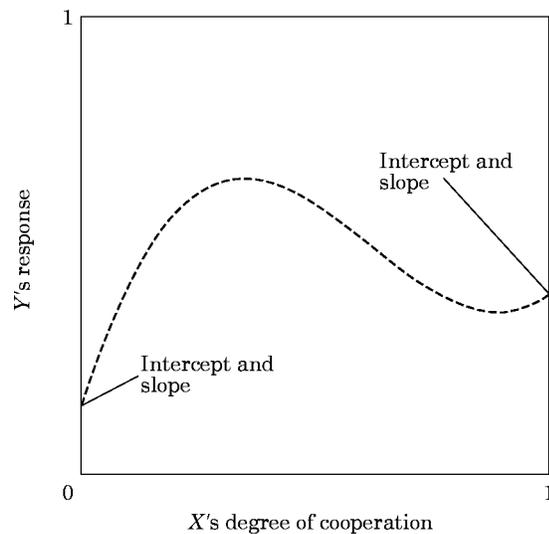


FIG. 8. Shows how nonlinear strategies can be defined by specifying two slopes in addition to the two intercepts.

In the same way, the pay-off AD gets against  $Y^*$  is readily calculated, and as before it has a critical value at  $Y_{01} = (R - P)/(T - P)$ . We therefore expect  $Y_{01}$  to be stable at a value close to

$$Y_{01}^{\text{crit}} = \min\left(1 - \frac{T - R}{R - S}, \frac{R - P}{T - P}\right)$$

exactly as for  $Y_0$  with memory 1 strategies. This is the largest value  $Y_{01}$  can take which makes it invasion-proof against both AD and “cooperative” unreactive strategies. This agrees well with the value found in simulations for a variety of prisoner’s dilemma pay-off settings.

### 6. Nonlinear Strategists

So far we have looked at linear strategists, which simply interpolate linearly between their extremal behaviors to arrive at general responses. This section describes a way to parameterise nonlinear strategies and shows a particular example of its evolution.

Nonlinear memory 1 strategies consist of curves rather than straight lines. A wide variety can be captured by using four parameters to define a cubic curve (see Fig. 8). Two of these,  $Y_0$  and  $Y_1$ , denote the values at  $x = 0$  and  $x = 1$  respectively, as before. A linear strategy would then be the straight line joining these points, and would have a slope of  $Y_1 - Y_0$ .

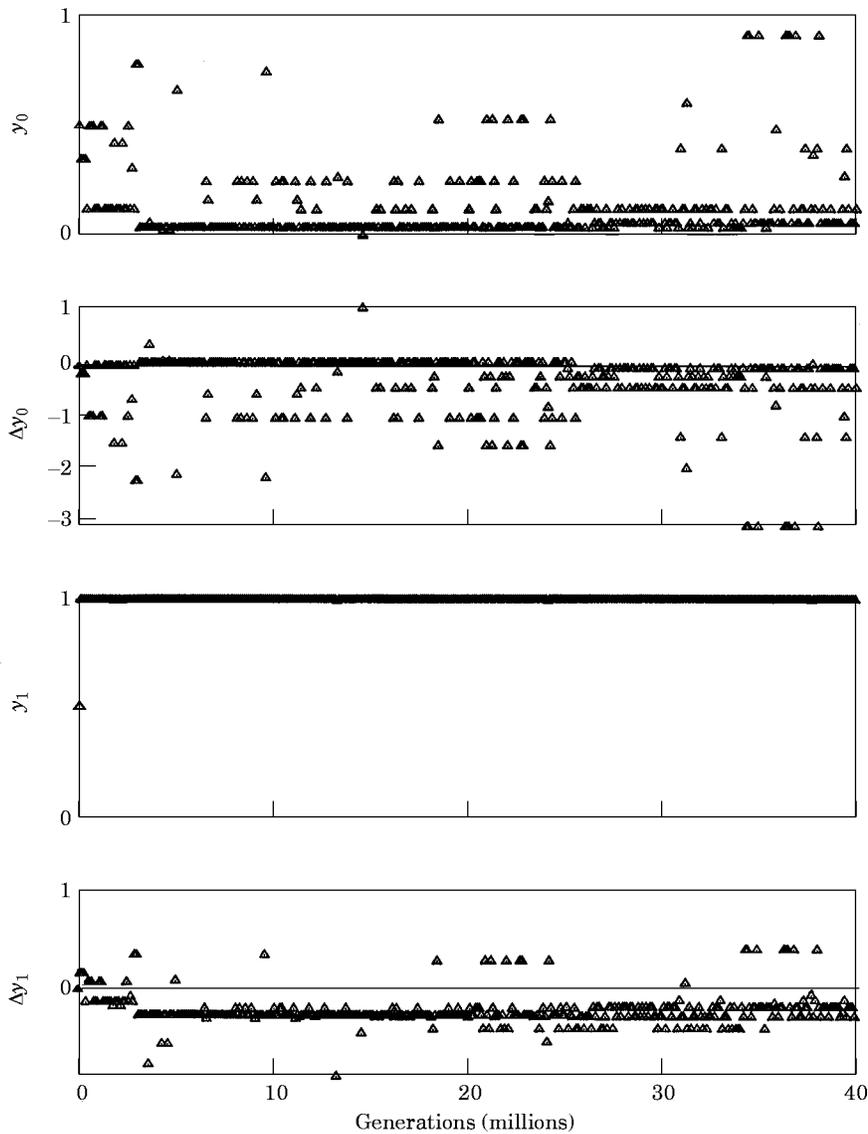


FIG. 9. An example of the timecourse for evolution of nonlinear memory 1 strategies.

Two further parameters,  $\Delta Y_0$  and  $\Delta Y_1$ , are used to specify the deviation of the actual slope from this value. For instance if  $\Delta Y_0 = 2.1$  the steepness of the line at  $x = 0$  is  $Y_1 - Y_0 + 2.1$ . The intercepts were chosen in the same way as before, while the slopes were chosen from the distribution  $((1 + \rho^1)(1 + \rho^{-1}))^{-1}$ . This allows large slopes of either sign but favours those close to zero. An algebraic check needs to be made that the curve stays in the correct range, with those that do not stay in the range  $[0, 1]$  over the domain being rejected.

Figure 9 shows a typical simulation where  $R = 3$ . Not shown is the average pay-off, which reaches  $R$  quickly and remains there for virtually the entire simulation. The number of strategies present stabilises at around ten. As Fig. 9 indicates, no single strategy is as clearly favoured as linear Generous-TFT (or  $Y^*$  is in the memory 2 case). Nevertheless, the average level of cooperation remains high throughout, so although not stable, as a group these strategists are cooperative. Figure 10 shows a selection of those that made up the population after 40 million generations. Note that any strategy which remains above the line  $x = y$ , as these all do, is guaranteed to reach full cooperation with itself. Unlike the linear case, there is no real convergence to a steady level of generosity to defectors. Instead there is a restless interplay of invasions and re-invasions between strategists which are all broadly cooperative. The particular dynamics (alternating or synchronous) seemed to make little difference to this outcome.

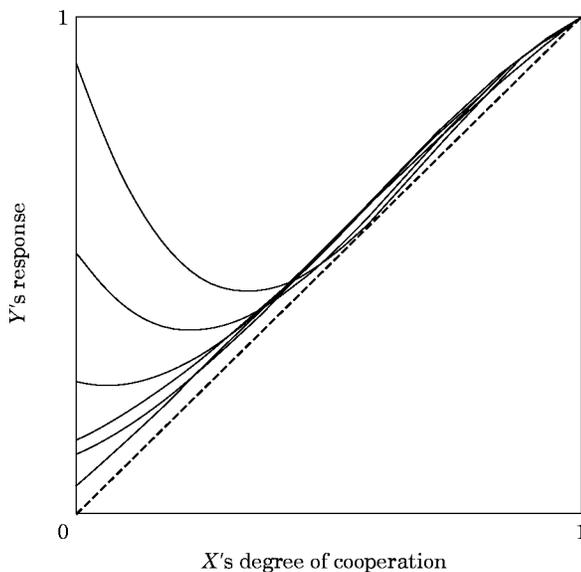


FIG. 10. A selection of strategies dominating long-term populations of memory 1 nonlinear strategies.

## 7. Discussion

One fascinating aspect of the Prisoner's Dilemma is that it potentially gives evolutionary situations in which the average pay-off decreases, even though the highest scoring individuals are being advanced at every generation. In the iterated prisoner's dilemma however cooperation can be stabilised as these simulations (and many others before) have demonstrated. This paper has examined whether the conclusions from the all-or-nothing case carry over to that of degrees.

There are strong similarities between the two scenarios. The continuous analogue of tit-for-tat presents itself in various guises, and its status as a robust mode of behavior is confirmed. In particular there is an exact correspondence in the memory 1 case, in that binary synchronous agents essentially face the same evolutionary pressures as linear continuous agents, with results that agree analytically and in simulations.

This however highlights one marked difference: with degrees of cooperation possible, the details of timing are largely irrelevant. This is certainly true for linear memory 1 players and appears to be true in practice for the other cases too. In particular if it is known that interactions are not synchronous, successful strategies in the continuous Prisoner's Dilemma are not the same as those of the all-or-nothing case. In the former we expect some form of Generous-TFT while in the latter AD is the predicted outcome. There is also a striking difference between successful memory 2 strategies in the two cases. In the binary case both WSLs and FBF cooperate after mutual defection, and never cooperate immediately after being a sucker. Without the binary probabilistic assumption both these behaviors are different. In particular, mutual defection leads only to more defection, at least in the short term. The reason for this may be that in the binary case the only route back to mutual cooperation is via such a switch whereas in the degrees case players can work their way back to cooperation step-by-step in relative safety.

Axelrod's seminal studies established the significance of tit-for-tat, but this position has waned as the original simulations were made more realistic. Accounting for "noise" in actions (via occasional errors) and in timing (via asynchronous dynamics) has made tit-for-tat look less significant than would have been thought at first. Including partial degrees of cooperation has the reverse effect, and would seem to reinstate tit-for-tat in its former role in models for the evolution of cooperation in competitive environments.

## REFERENCES

- AXELROD, R. (1984). *The Evolution of Cooperation*. New York: Basic Books.
- AXELROD, R. & DION, D. (1988). The further evolution of cooperation. *Science*, **242**, 1385–1390.
- AXELROD, R. & HAMILTON, W. D. (1981). The evolution of cooperation. *Science*, **211**, 1390–1396.
- DUGATKIN, L. A. (1991). Do guppies play tit for tat during predator inspection visits? *Behav. Ecol. Sociobiol.* **25**, 395–399.
- FREAN, M. R. (1994). The prisoner's dilemma without synchrony. *Proc. Roy. Soc. Lond. B*, **257**, 75–79.
- GODARD, R. (1993). Tit-for-tat among neighboring hooded warblers. *Behav. Ecol. Sociobiol.* **33** (1), 45–50.
- HART, & HART, L. A. (1992). Reciprocal allogrooming in impala, *aepyceromelampus*. *Anim. Behav.* **44** (6), 1073–1083.
- MAYNARD SMITH, J. (1982). *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.
- MILINSKI, M. (1987). Tit for tat in sticklebacks and the evolution of cooperation. *Nature*, **325**, 434–435.
- MOLANDER, P. (1985). The optimal level of generosity in a selfish, uncertain environment. *J. Conflict Resolution* **29**, 611–618.
- NOWAK, M. & SIGMUND, K. (1990). The evolution of stochastic strategies in the prisoner's dilemma. *Acta appl. math.* **20**, 247–265.
- NOWAK, M. & SIGMUND, K. (1992). Tit for tat in heterogeneous populations. *Nature*, **355**, 250–253.
- NOWAK, M. & SIGMUND, K. (1993). A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. *Nature*, **364**, 56–58.
- NOWAK, M. & SIGMUND, K. (1994). The alternating prisoner's dilemma. *J. theor. Biol.* **168**, 219–226.
- TAYLOR, P. & JONKER, L. (1979). Evolutionarily stable strategies and game dynamics. *Math. Biosc.* **40**, 145–156.
- WILKINSON, G. S. (1984). Reciprocal food-sharing in the vampire bat. *Nature*, **308**, 181–184.