

# The prisoner's dilemma without synchrony

MARCUS R. FREAN†

*The Physiological Laboratory, Downing Street, Cambridge CB2 3EG, U.K.*

## SUMMARY

There are many situations in which biological organisms cooperate despite obvious incentives to do otherwise. Such situations are commonly modelled by using a paradigm known as the prisoner's dilemma. In this way cooperative behaviour has previously been shown to emerge in a model population of strategies. If players can make probabilistic choices, taking into account their co-player's previous action, a strategy known as 'generous tit for tat' dominates the long-term behaviour of such a population. If they can also take into account their own previous action, a strategy of 'win stay, lose shift' dominates instead. These models assumed that participants make their decisions in synchrony, which seems improbable in many biological situations. Here we show that the timing of decisions is critical in determining which strategy emerges in the long run. If individuals make their decisions at different times, neither of the above strategies survives given the usual payoffs. In the former case, generous tit for tat succumbs to inveterate defectors, and in the latter a new strategy takes over. This 'firm but fair' strategy is retaliatory yet highly cooperative. In particular, continued exploitation of a sucker is no longer a successful behaviour.

## 1. INTRODUCTION

Animals frequently show cooperative behaviour in situations where it appears better not to on an individual basis. Examples include predator inspection (in which groups of fish approach a predator together), reciprocal grooming, mutual restraint of conflict, and food sharing: all these share the characteristic that individual exploitation of some kind would seem to be more advantageous than mutual cooperation. Because evolution has presumably tended to select behaviours that benefit the individual as opposed to the group, mutual cooperation in this context is somewhat surprising.

Situations such as this can be modelled as a game in which two players interact, and each opts either to cooperate or defect (i.e. not cooperate). If both players cooperate each gets a payoff of  $R$  points (a 'reward'), compared with the lesser  $P$  points should both defect ('punishment'). If they choose different behaviours, the player who defects receives the maximum  $T$  points ('temptation'), whereas the 'sucker' who cooperated gets the minimum score of  $S$ . With  $S < P < R < T$  and  $S + T < 2R$ , the dilemma becomes apparent: regardless of the other player's choice, an individual is better off defecting in a game consisting of a single round. This leads to mutual defection even though mutual cooperation carries a greater reward.

If the game is iterated, however, and players can base their decision on previous encounters, cooperation may become worthwhile. The striking success of 'tit for tat' in Axelrod's computer tournaments (Axelrod &

Hamilton 1981; Axelrod 1984; Axelrod & Dion 1988) showed how this could occur, and this strategy has subsequently been observed in a variety of biological contexts (Lombardo 1985; Milinski 1987; Dugatkin 1991; Hart & Hart 1992; Godard 1993). Tit for tat (TFT) starts by cooperating, and then simply copies whatever the other player did last. This proves remarkably effective: for instance, in a population of strategies which all play each other and reproduce in relation to their mean payoff, TFT is almost alone in being able to invade a population of AD ('always defect'). However, if two TFT's play one another, any error leads to a long run of switching between cooperation and defection. This intolerance to noise allows other strategies to take over (Nowak & Sigmund 1992), notably generous tit for tat (GTFT), which cooperates with some non-zero probability even after a defection by the other player. 'Turning the other cheek' in this way prevents long runs of back-biting but at the cost of vulnerability to inveterate defectors.

TFT and its probabilistic variants are 'opponent-driven' in the sense that they only consider the action of the other player in making their decisions. The idea of a genetic tournament has recently been extended to include probabilistic strategies which can also consider their own previous action (Nowak & Sigmund 1993): these might be called 'state-driven' strategies. Within this larger space of possibilities, a strategy of 'win stay, lose shift' (WSLS) dominates. Like TFT, WSLS cooperates after a mutual cooperation and defects if it is a sucker. Unlike TFT, however, it continues to defect against a sucker, and cooperates after a mutual defection.

An implicit assumption in this work has been that the two individuals make their decisions at effectively the same moment, which is uncharacteristic of many

† Present address: Department of Zoology, University of Otago, P.O. Box 56, Dunedin, New Zealand.

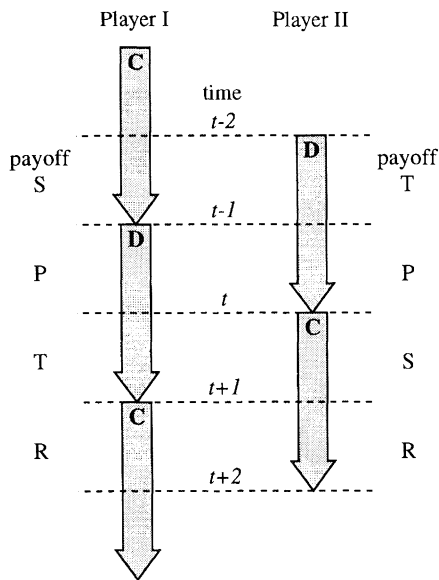


Figure 1. An alternating form of the prisoner's dilemma: a particular sequence is shown for illustration. For example, at time  $t$  player II makes a decision (in this case C) based on player I's decision at time  $t-1$  (D) and its own decision at  $t-2$  (also D). Equivalently, its decision is simply based on the immediate behaviours of both players. The payoff for each individual is determined by the joint behaviour over the interval between decisions. For example, between time  $t$  and  $t+1$ , player I defects while player II cooperates, so the latter receives the 'suckers' payoff  $S$ .

biological situations. For example, in the case of reciprocal food sharing between vampire bats (Wilkinson 1984), a bat that has found food shares with one that has gone without, which clearly should not be modelled as a simultaneous exchange. Another frequently quoted example concerns 'predator inspection', in which fish approach a potential predator in pairs (Milinski 1987), even though hanging back and observing the other's fate is the safer option for either individual. This occurs under conditions in which the fish are constantly interacting with one another: although their behaviour might appear to be synchronous there is no evidence that they actually make simultaneous decisions as opposed to rapid but asynchronous ones.

In the following we consider an asynchronous alternative in which players 'take turns'. That is, each individual waits for the other to respond before reassessing its own decision, instead of making its decision at the same moment. (Another alternative, not shown here, is that players make their decisions at random and independent times: the results in this case closely parallel those of alternating games. Notably the payoff under each of these two dynamics for a strategy playing against itself is the same.) This does not change the dilemma of a one-shot game, as the payoff is greatest for a defection whatever the other player has done (or is about to do). However, the outcome of the iterated prisoner's dilemma is substantially altered.

Nowak & Sigmund (1994) have considered an alternating game dynamics in which the payoff is entirely determined by the 'leader' (the individual making the current decision and acting on it). This

models biological situations such as that of the vampire bat, as a well-fed bat can decide whether to share food or not, whereas a hungry bat is unable to influence the payoff of either party. They showed that this was equivalent to prisoner's dilemma with payoffs constrained to obey  $R+P=S+T$ .

From the point of view of timing, it is useful to distinguish between decisions about how to act and the behaviour that results from such decisions. Decisions are made at discrete times, whereas behaviours can be discrete (as in feeding by vampire bats) or continuous (as in body positioning by fish). This paper addresses the latter situation, shown schematically in figure 1. Players alternate in their decision making, and their actions are extended over time. As in the synchronous case, players make decisions based on their own previous decision and that of their co-player. The payoff is determined by the current joint behaviour of both players over the next interval of their interaction. This models situations in which individuals adopt a given behaviour for some period of time (such as approaching a predator in the case of fish), and the payoff they receive (such as a lowered risk of predation) results from the joint behaviour of both animals over that time. The overlap in behaviours means that each decision actually results in two payoffs: one due to the other player's previous decision, and one due to their subsequent one. Note that under these conditions the restriction  $R+P=S+T$  does not apply.

## 2. METHODS

We use the same approach as Nowak & Sigmund (1993). A strategy is defined by four parameters  $p_1, p_2, p_3, p_4$ , these being its probabilities of cooperation following each of the possible previous joint behaviours CC, CD, DC and DD, respectively (where the first letter refers to the strategy in question and the second to the other player). For example, if player I defected at time  $t-1$ , and this is followed by player II cooperating at  $t$ , player I will decide to cooperate at  $t+1$  with probability  $p_3$ . Unless stated otherwise, Axelrod's payoffs of  $R=3, S=0, T=5$  and  $P=1$  are used. Note that opponent-driven strategists have  $p_1=p_3=p$  and  $p_2=p_4=q$ , responding to the other player's cooperation or defection, respectively. Under this notation TFT can be written as  $(1, 0, 1, 0)$ , and WLS as  $(1, 0, 0, 1)$ . In simulations, we limit all these probabilities to be between 0.001 and 0.999, meaning that every strategy, no matter how 'pure', is subject to some uncertainty. The interaction between two players can then be seen as a Markov process, and the payoff in the limit of an infinitely long game calculated in the following way. The strategy adopted by player I is denoted by  $(p_1, p_2, p_3, p_4)$  and that of player II by  $(p'_1, p'_2, p'_3, p'_4)$ . We can then write down a matrix in which each row gives the probabilities of the four joint behaviours after a decision, with the columns indexing the joint behaviour before the decision. If it is player I's turn to decide, this matrix is

$$M_I = \begin{pmatrix} p_1 & 0 & 1-p_1 & 0 \\ 0 & p_2 & 0 & 1-p_2 \\ p_3 & 0 & 1-p_3 & 0 \\ 0 & p_4 & 0 & 1-p_4 \end{pmatrix}.$$

For instance, the probability of DC and CC is simply the chance that player I defects after a mutual cooperation, namely  $1-p_1$ . The zero entries refer to transitions that are

impossible given that only player I is making a decision. If it is player II's turn to decide, the corresponding matrix is

$$\mathbf{M}_{II} = \begin{pmatrix} p'_1 & 1-p'_1 & 0 & 0 \\ p'_2 & 1-p'_2 & 0 & 0 \\ 0 & 0 & p'_3 & 1-p'_3 \\ 0 & 0 & p'_4 & 1-p'_4 \end{pmatrix}.$$

For a round consisting of a pair of decisions in which player I decides first followed by player II, the transition matrix is therefore  $\mathbf{M}_I \mathbf{M}_{II}$ . This has a unique left eigenvector whose elements are the relative frequencies of each of the four joint behaviours in a game of infinite length. We can then simply add up the four payoffs, weighted by their frequencies, to obtain the expected payoff per round for a long interaction. This is how Nowak & Sigmund (1994) calculate the payoffs in their model. It corresponds to counting each round in which player I decides first, followed by player II. However, as noted above, if behaviours are extended over time each action results in two payoffs, corresponding to the other player's previous and subsequent action. To include both we need to repeat the above procedure for the matrix  $\mathbf{M}_{II} \mathbf{M}_I$  as well as  $\mathbf{M}_I \mathbf{M}_{II}$ , and add the resulting payoffs together.

To put the game into an evolutionary context, payoff is equated with genetic fitness. Simulations begin with a single strategy of (0.5, 0.5, 0.5, 0.5), and we keep track of relative proportions of strategies rather than numbers of individuals. At each successive generation all the strategies play one another (each game being an infinitely long sequence of interaction), and their proportions in the population then change according to the usual dynamics (Maynard Smith, 1982). Under these dynamics any strategy gaining a net payoff which is above the average will increase its relative share of the population. There is a small probability (1%) in any generation that a small amount (0.2%) of a new mutant strategy enters the population, and those making up less than 0.1% of the total population are removed. Simulations need to be run for large numbers of generations for a consistent 'winner' to emerge. In part this reflects the fact that a large number of mutants need to be tried out to see that one in particular stands out from the crowd. Moreover, the mutant probabilities were generated from a uniform distribution, which unavoidably means that huge numbers of very poor random strategies are tried out.

### 3. RESULTS

We first note the effect of asynchrony on opponent-driven strategists (see figure 2). To check for robustness, a variety of values of  $R$  (the payoff for mutual cooperation) were used, between the limits  $2R > S + T$  and  $R < T$  required of a prisoner's dilemma. For each value of  $R$ , 50 runs were made, and in each run the values of  $p$  and  $q$  were recorded after  $10^7$  generations. Runs for larger numbers of generations do not give significantly different results. In the synchronous case a strategy of GTFT dominates the long-term population over a wide range of values for the payoff  $R$ :  $p$  is almost 1 and the 'generosity'  $q$  closely follows the theoretical optimum value (Molander 1985; Nowak & Sigmund 1990) of  $\min [1 - (T - R)/(R - S), (R - P)/(T - P)]$ . Without synchrony, for  $R \leq 3.5$  no TFT-like strategists survive and AD is the clear victor (hence TFT might not have won Axelrod's tournaments had the dynamics been asynchronous). By contrast, for  $R \geq 4$  the highly cooperative GTFT strategy with  $q \approx 0.8$  (as in the synchronous case) wipes out AD completely. The

transition is not smooth. At  $R = 3.8$ , GTFT and AD are present in about equal numbers (non-extremal values of  $p$  are almost never present).

The distribution of probabilities obtained by state-driven strategists is shown in figure 3. Virtually every strategy which survives in the long term is cooperative after a mutual cooperation ( $p_1$ ), and likely to retaliate after being made a sucker ( $p_2$ ), i.e. if they cooperated in a given round, they play 'tit for tat' in the sequel. The probability of cooperating after exploiting the other player,  $p_3$ , is strikingly different: whereas wSLS would 'win and stay', successful strategists in alternating games strongly prefer to switch to cooperation, despite having just received the maximum payoff  $T$  for defecting. Hence they are unable to make the most of naive cooperators. Cooperation following a mutual defection ( $p_4$ ) is favoured in both cases. In the synchronous case, as expected, the long-term population is dominated by wSLS, but in the alternating case a new strategy emerges. This strategy could be called 'firm but fair' (FBF): it is 'firm' in that it retaliates by defecting if it was a sucker in the previous round. It is also 'fair' in that it does not retaliate against a defector if it defected itself, and it cooperates with suckers rather than continuing to exploit them. In the notation above, FBF is  $(1, 0, 1, \alpha)$ , with  $\alpha$  significantly greater than zero. Note that if  $\alpha$  were to be near zero the strategy would just be TFT.

Again it is important to test for robustness by varying  $R$ . At  $R = 2.6$ , where cooperation is scarcely better than oscillating between payoffs  $T$  and  $S$ , wSLS's domination drops to less than 10% in the synchronous case under the same conditions. However, without synchrony, FBF strategists (with  $\alpha > 0.3$ ) still made up over 98% of the pooled populations. For  $R = 4.5$

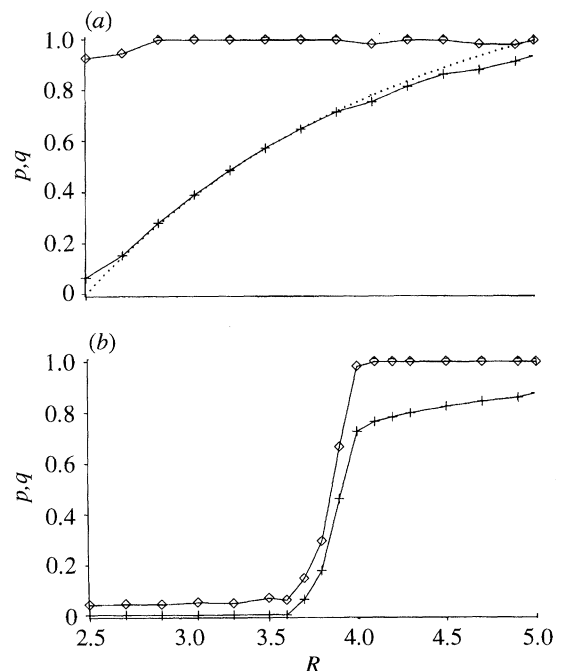


Figure 2. Opponent-driven strategists. The figures show average values of  $p$  (diamonds) and  $q$  (crosses) in the long term population against  $R$ , the reward for mutual cooperation. (a) Synchronous, (b) alternating. In (a) the dotted line shows the theoretical optimum of  $q$ .

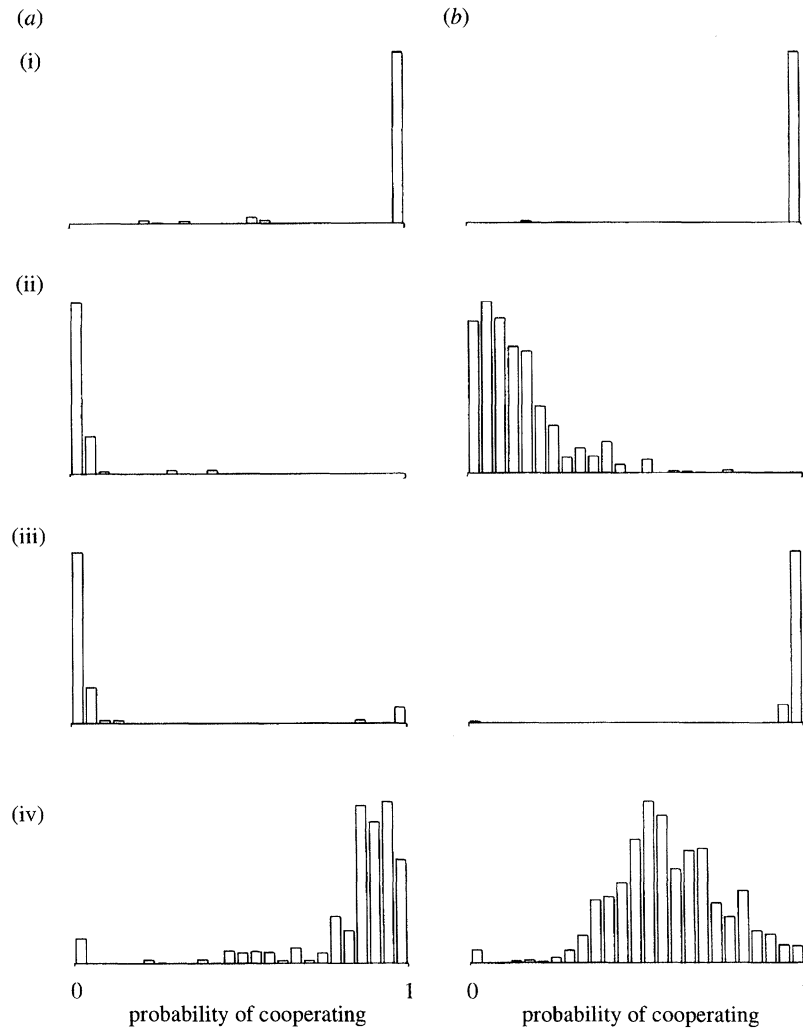


Figure 3. State-driven strategists. The histograms show the prevalence (relative to the maximum) of  $p_1$ ,  $p_2$ ,  $p_3$  and  $p_4$  in a population after  $10^9$  generations, pooled from 100 such runs. (a) Synchronous, (b) alternating. (i)  $p_1$ , (ii)  $p_2$ , (iii)  $p_3$ , (iv)  $p_4$ .

cooperation is favoured very strongly: the result is FBF (with  $\alpha > 0.5$ ) in 40% of cases, with the remainder preferring to cooperate in all four situations. Hence FBF is more robust in the asynchronous case than is WSLs in the synchronous one.

#### 4. DISCUSSION

The prisoner's dilemma has found wide acceptance as a model for the evolution of cooperation amongst selfish agents. In a one-shot game it makes no difference what the order of play is. In the iterated form it has always been the case that individuals are modelled as making synchronous decisions. Synchronicity here really means that they must determine their behaviour in the light of the co-player's previous action rather than the current one. However, in many situations of interest to biologists, individuals make decisions at different times, and in doing this take their co-player's present behaviour into account. Here we have examined the consequences of one such dynamic, in which individuals make decisions one after another. In contrast to the usual prisoner's dilemma, this models situations in which an animal acts and then waits for a response before reassessing its decision.

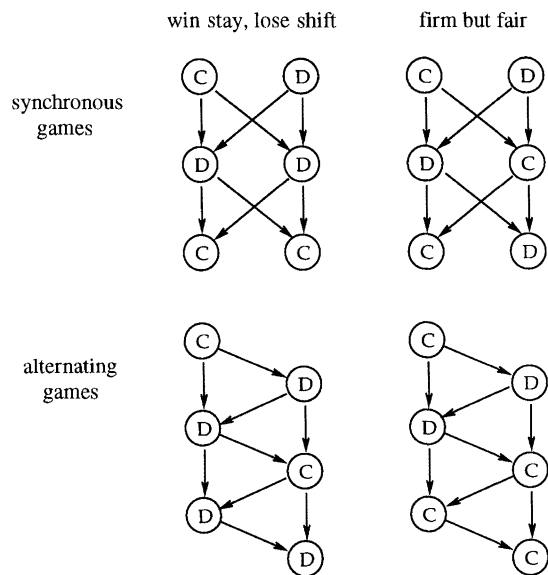


Figure 4. Getting out of fights. The diagrams show the first few steps following a defection, where the strategy named at the top is playing itself.

Why do these different playing sequences produce different strategies? If a strategy is to dominate a population it needs to spend as much time as possible cooperating with itself. Suppose that two individuals using the same strategy play each other and a single defection occurs. In the alternating case errors are even more detrimental to TFT than with synchrony: rather than a run of oscillations yielding an average payoff of  $(S+T)/2$  per round, a run of mutual defection ensues, with only  $P$  points per round. The only way for an opponent-driven strategy to escape is to become very 'generous', but this increases its vulnerability to attack by hardened defectors. However, state-driven strategists can be generous while remaining difficult to exploit. Figure 4 shows the situation for wsls and fbf after an unprovoked defection. If the game dynamics are synchronous, the injured party is very likely to be retaliating, so attempting to cooperate with them as in fbf (or TFT) is not a route back to mutual cooperation at all. Instead, the wsls strategy continues to defect, which leads almost always to mutual defection. From there two wsls's will switch to mutual cooperation in one step. Hence the significance of wsls's defection against suckers is that this enables it to return to cooperation with similar strategies one round later (rather than an adaptation for exploiting naive cooperators, which make only brief appearances in the population). After a mutual defection both wsls players must switch to cooperation at the same moment because if only one player is generous retaliation will follow. If players take turns, wsls's behaviour leads to mutual recriminations. Instead, apologizing becomes a good idea: it only takes one individual to be generous, and mutual cooperation will follow. Notice that in an alternating game it only takes one individual to choose cooperation and the other will follow under fbf. This may be why  $p_4$  is very close to 1 in wsls and much more variable in fbf.

To summarize, in the iterated prisoner's dilemma, strategies which can produce long runs of relatively uninterrupted mutual cooperation can overthrow defectors. It is vital to retaliate after a defection by the other player to avoid being exploited. However, this opens up the possibility of runs of oscillations or mutual defections, and in these cases the order of play is crucial. We have considered long interactions between strategists with very short memories. If such players do

not make their decisions at the same time it no longer follows that continuing to exploit a sucker will be successful. Guarded generosity may indeed be preferable for selfish individuals.

I thank Roland Baddeley, Horace Barlow, Arthur Pece and particularly David MacKay for helpful discussions, and the referees for helpful comments. This work was carried out under grants from the Medical Research Council, the Newton Trust and the James S. McDonnell Foundation.

#### REFERENCES

- Axelrod, R. 1984 *The evolution of cooperation*. New York: Basic Books.
- Axelrod, R. & Dion, D. 1988 The further evolution of cooperation. *Science, Wash.* **242**, 1385–1390.
- Axelrod, R. & Hamilton, W. D. 1981 The evolution of cooperation. *Science, Wash.* **211**, 1390–1396.
- Dugatkin, L. A. 1991 Do guppies play tit for tat during predator inspection visits? *Behav. Ecol. Sociobiol.* **25**, 395–399.
- Godard, R. 1993 Tit-for-tat among neighboring hooded warblers. *Behav. Ecol. Sociobiol.* **33** (1), 45–50.
- Hart, B. L. & Hart, L. A. 1992 Reciprocal allogrooming in impala, *aepycerosmelampus*. *Anim. Behav.* **44** (6), 1073–1083.
- Lombardo, M. P. 1985 Mutual restraint in tree swallows: A test of the tit for tat model of reciprocity. *Science, Wash.* **227**, 1363–1365.
- Maynard Smith, J. 1982 *Evolution and the theory of games*. Cambridge University Press.
- Milinski, M. 1987 Tit for tat in sticklebacks and the evolution of cooperation. *Nature, Lond.* **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, Lond.* **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, Lond.* **364**, 56–58.
- Nowak, M. & Sigmund, K. 1994 The alternating prisoner's dilemma. *J. theor. Biol.* (In the press.)
- Wilkinson, G. S. 1984 Reciprocal food-sharing in the vampire bat. *Nature, Lond.* **308**, 181–184.

Received 20 April 1994; accepted 6 May 1994