A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game

Martin Nowak* & Karl Sigmund†

*Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK 
†Institut für theoretische Biologie, Universität Wien, Strudlhofgasse 4, A-1090 Vienna, Austria

The Prisoner's Dilemma is the leading metaphor for the evolution of cooperation in large populations of competing agents, especially since the well-known computer tournaments of Axelrod's and their application to biological communities. In Axelrod's simulations, the simple strategy tit-for-tat did outstandingly well and subsequently became the major paradigm for reciprocal altruism. Here we present extended evolutionary simulations of heterogeneous ones of probabilistic strategies including mutation and selection, and report the unexpected success of another protagonist: Pavlov. This strategy is as simple as tit-for-tat and embodies the fundamental behavioural mechanism win-stay, lose-shift, which seems to be widespread in nature. Pavlov has two important advantages over tit-for-tat: it can correct occasional mistakes and exploit unconditional cooperators. This second feature prevents Pavlov populations from being undermined by unconditional cooperators, which in turn invites defectors. Pavlov seems to be more robust than tit-for-tat, suggesting that cooperative behaviour in natural situations may often be based on win-stay, lose-shift.

Two players engaged in the Prisoner's Dilemma have to choose between cooperation (C) and defection (D). In any given round, the two players receive $R$ points if both cooperate and only $P$ points if both defect; but a defector exploiting a cooperator gets $T$ points, while the cooperator receives $S$ (with $T > P > S$ and $2R > T + S$). Thus in a single round it is always best to defect, but cooperation may be rewarded in an iterated (or spatial) Prisoner's Dilemma.

The conspicuous success of the tit-for-tat (TFT) strategy (start with a C, and then use your co-players previous move) relies partly on the collective unfitness of a deterministic world. In natural populations, errors occur. TFT suffers from stochastic perturbations in two ways: (1) A TFT population can be "softened" up by random noises in introducing unconditional cooperators, which allows exploiters to grow (TFT is not an evolutionary stable strategy); and (2) occasional mistakes between two TFT players cause long runs of mutual backbiting. (Such mistakes abound in real life, even humans are apt to vest frustrations upon innocent bystanders.) Within the restricted world of strategies reacting only to the co-players previous move, TFT has an important, but transitory role; in a real world of complex interactions, defectors, but then bow down to a related strategy, genorus tit for tat (GFT) cooperates after a co-player's C, but also with a certain probability after a D.

But as soon as one admits strategies which take into account the moves of both players, a new round, evolution becomes much less transparent. We first conjectured that GFT (or variants thereof) would win the day, but are forced to admit, after extensive simulations, that the strategy Pavlov did much better in the long run. A Pavlov player cooperates if and only if both players cooperated on the previous move. The name stems from the fact that this strategy embodies the alluring reflex-like response to the payoff: it repeats its former move if it was rewarded by R or T points, but switches behaviour if it was punished by P or S points. This strategy, which went by the name of "simpleton" fares poorly against invertebrate defectors: in every second round, it switches to cooperation. It cannot gain a foothold as a defector's world; defectors have to be invaded by other strategies, like TFT. But Pavlov has two important advantages over TFT, (1) an inadvertent mistake between players using TFT causes a drawn-out battle; between two Pavlovians, it causes one round of mutual defection followed by a return to joint cooperation. Thus Pavlov is fairly tolerant, like GFT, and can correct mistakes. (2) Whereas TFT and GFT can be invaded by DFT by all-out cooperators (to the eventual profit of exploiters), Pavlov has no qualms in exploiting a sucker, once it has discovered (after an accidental mistake) that it need not fear any retaliation.
Selten cannot sustain a Pavlov population. In this sense, cooperation based on Pavlov is a safer bet than TFT. Pavlov’s advantage shows best among nice strategies.

For our simulations, we consider all (stochastic) strategies with memory one (that is, recalling only the previous round). This yields the conditional probabilities \( p_i, p_j, p_k, p_l \) where cooperation, given the outcome of the previous round, is \( R, S, F, O \) or \( P \), respectively. The game between two such strategies can be formulated as a Markov process, and its stationary distribution specifies the payoffs for the infinitely iterated Prisoner’s Dilemma. Owing to noise, the initial state has no effect on the final state. For mathematical simplicity we retain the assumption of the infinitely iterated game, but note that the outcome is essentially unchanged if we consider sufficiently long iterated games. In our evaluation TFT is given by \( (1, 0, 0, 1) \) and Pavlov by \( (1, 0, 0, 1) \), but mistakes in implementing the move change 1 to 0 and 0 to 1, where \( z \) is a small number specifying the minimal amount of noise. This is closely related to the ‘bumping hand’ in Selten’s game theoretical notion of perfect equilibrium. We start each simulation with the random strategy \( (0.5, 0.5, 0.5, 0.5) \). Every 100 generations (on average), we introduced a small amount of a randomly chosen mutant strategy, the frequencies of strategies spread according to the usual game dynamics, reflecting natural selection. Strategies with higher payoffs produce more offspring. Strategies whose frequency dropped below a certain threshold were discarded. Each run was observed for 10^7 generations, generating a total of about 10^8 different mutant strategies. (Note that the timescale is arbitrary, because the difference equation can be seen as an approximation of a differential equation. It is, however, very important to study the long-term dynamics and to try many mutants.) The evolutionary dynamics display an extreme diversity. Nevertheless, they allow some clear and simple conclusions: (1) the plot for the average payoff in the population is a smooth, well-behaved equilibrium (Fig. 1). Most of the time, this payoff is very close to one of the external values \( P \) (a regime of defection) or \( B \) (overall cooperation). The time for switching from one of the regimes to the other is usually extremely short (only a few generations). The periods of stability last for millions of generations. The later in the run, the longer they last. But the threat of a sudden collapse may never abate. (2) There is a clear tendency for cooperation (Fig. 2). After \( r = 10^7 \) generations, only 27.5% of the runs exhibit cooperation (population payoff > 2.95); but 91.0% at \( r = 10^9 \). There is also a clear tendency towards Pavlov, which dominates 19% of the runs at \( r = 10^9 \), but

![Diagram](image-url)

**FIG. 1** An anti-coalition simulation (including mutation and selection) of all strategies that consider the previous move in the Iterated Prisoner’s Dilemma. Such strategies are defined by four probabilities \( p_i, p_j, p_k, p_l \). TFT cooperates after having received payoff \( P, S, F, O \) in the previous round. We start with the random strategy \( (0.5, 0.5, 0.5, 0.5) \). For technical reasons, we use the U-shaped density distribution \( \exp(-x^2) \) for sampling the \( p_i \) in the interval \( 0,1 \). This distribution is close to the empirical distribution, helps to achieve the correct frequency, and is more efficient. We assume that there is a minimal amount of noise, but we reduce 0.05% to 0.05% for sampling.

The frequencies of strategies \( p_i \) according to the usual game dynamics. Strategies with frequencies below 0.001 are removed. In this way, the initial strategy for cooperation is unsuccessful. An AID-like strategy emerges as winner and dominates until \( r = 10^9 \). Then a TFT mutant invades and establishes a regime of cooperation dominated by TFT. This population is underdetermined by noise and more forgiving strategies (increasing \( p_i \)), which leads to a faster and overall cooperation and another period of defection, now dominated by the severe retaliator (DFW) \( (0.999, 0.001, 0.001, 0.001) \) again TFT invades and stabilizes the rate of cooperation. After the small adjustments of \( p_i, p_j, p_k, p_l \), there seems to be a long-lasting period of cooperation dominated by the Pavlov-like strategy \( (0.999, 0.001, 0.001, 0.0007, 0.948) \), which persists at least until \( r = 10^9 \) (not shown).

The figure shows the average cooperation payoff. The number of strategies present at a given time, and the distribution of the probabilities \( p_i, p_j, p_k, p_l \). The latter is given in generations of the difference equation, but the timescale is arbitrary if the difference equation is understood as an approximation of a differential equation. Payoff values: \( R = 3, S = 0, F = 5, P = 1 \).
Pavlov-like behaviour does not seem to be restricted to strategies, which only remember the last move. In other evolutionary runs, where mutations can extend the memory length, similar strategies have been found: typically, they rezone cooperation after two rounds of mutual defection. Like Pavlov, they correct for mistakes and exploit unconditional cooperators. There is, of course, no limit in the complexity of conceivable strategies. But it may be expected that the simple, natural rule of win-stay, lose-shift performs well under a variety of sophisticated conditions. Pavlov is no simpleton.