



Simple Adaptive Strategy Wins the Prisoner's Dilemma

CHRISTOPH HAUERT*†‡§ AND OLAF STENULL§

†*Institut für Mathematik, Universität Wien Strudlhofgasse 4, A-1090 Vienna, Austria*, ‡*Department of Zoology, University of British Columbia, 6270 University Boulevard, Vancouver, B.C., Canada V6T 1Z4* and §*Institut für Theoretische Physik III Heinrich–Heine–Universität Düsseldorf Universitätsstraße 1, D-40225 Düsseldorf, Germany*

(Received on 22 November 2001, Accepted in revised form on 25 April 2002)

The prisoner's dilemma has become the leading paradigm to explain the evolution of cooperation among selfish individuals. Here, we present an adaptive strategy that implements new mechanisms to process information about past encounters. The history of moves is summarized in an internal state which then determines the subsequent move. This enables the strategy to adjust its decisions to the character of the current opponent and to adapt the most promising strategic behavior. For this reason, we call such strategies *Adaptor*. Through evolutionary simulations, we demonstrate that the concept of *Adaptor* leads to strategical patterns that are (a) highly cooperative when playing against kin, (b) stable in a sense that goes far beyond the concept of evolutionary stability, (c) robust to environmental changes, i.e. variations of the parameter values and finally (d) superior in performance to the most prominent strategies in the literature.

© 2002 Elsevier Science Ltd. All rights reserved.

1. Introduction

The evolution of cooperation and altruistic behavior among unrelated selfish individuals is one of the most fundamental questions in disciplines as diverse as biology and economics (Colman, 1995; Dugatkin, 1997; Kagel & Roth, 1995). Tremendous scientific effort has been dedicated to theoretical and experimental investigations to understand the superiority of mutually beneficial interactions despite their obvious vulnerability to exploitation through defective individuals. Across disciplines, the prisoner's dilemma (*PD*) (Axelrod & Hamilton,

1981) became a widely accepted mathematical framework to study cooperative behavior.

In the *PD*, individuals engage in pairwise interactions with two behavioral options. They must simultaneously decide whether to cooperate or to defect. Their joint behavior then determines their payoffs. A cooperative act involves a cost c to the donor and provides a benefit b to the recipient. For obvious reasons $b > c$ must hold. Thus, mutual cooperation pays a reward $R = b - c$, while mutual defection results in a punishment $P = 0$. If one player opts for D and the other for C , the former obtains the temptation to defect $T = b$ and the latter is left with the sucker's payoff $S = -c$. Hence, defectors are always better off regardless of their opponents decision. Consequentially, "rational" individuals, i.e. players attempting to

*Corresponding author. Tel.: +1-604-822 6726; fax: +1-604-822 2416.

E-mail address: hauert@zoology.ube.ca (C. Hauert).

maximize their short-term profits, always end up with the punishment P instead of the higher reward R for cooperation. The general definition of the PD requires the following rank ordering of the payoff values:

$$T > R > P > S. \quad (1)$$

Note that the notation involving costs c and benefits b always satisfies this ordering. In the following, we generally stick to this simpler notation with only two parameters.

In recent years, several scenarios have been proposed and experimentally verified to overcome this dilemma and to allow for cooperative behavior to emerge even among unrelated individuals. All of them are based on simple discrimination mechanisms that enable cooperative individuals to target their altruistic acts towards certain other individuals only. These mechanisms can be divided into three scenarios: (a) direct reciprocity, (b) indirect reciprocity and (c) spatial extension. In (a) the same individuals interact repeatedly. This enables them to trigger their behavior on the outcome of previous rounds and therefore to cooperate only with cooperative opponents (see below). In the case of indirect reciprocity (b), individuals interact only in one direction, i.e. there is no chance to reciprocate and return a service. Recent models (Leimar & Hammerstein, 2001; Nowak & Sigmund, 1998) suggest that individuals carry some sort of reputation that essentially summarizes their past actions. This enables cooperative individuals to direct altruistic acts selectively towards helpful individuals by following the experimentally verified rule *help and you shall be helped* (Wedekind & Milinski, 2000). Finally, in spatially extended systems (c) individuals interact once in a local neighborhood. This enables cooperators to prosper by forming clusters and thereby minimizing interactions with defective individuals (Hauert, 2001, 2002; Herz, 1994; Killingback *et al.*, 1999; Lindgren & Nordahl, 1994; Nowak & May, 1992).

1.1. DIRECT RECIPROCITY

The concept of direct reciprocity was summarized in a famous review by Trivers (1971) and has become very popular through game

theoretical computer tournaments organized by Axelrod (1984). In these tournaments, submitted strategies competed in the repeated or iterated prisoner's dilemma (IPD). The IPD imposes a second constraint on the payoff values R, S, T, P :

$$R > \frac{T + S}{2}. \quad (2)$$

This ensures that mutual cooperation pays the highest reward to the community. In particular, this prohibits that periodic alternations of C and D moves lead to more favorable outcomes. The notation with c and b always satisfies condition (2).

Axelrod's tournaments were won by a surprisingly simple strategy called *Tit-for-Tat* (TFT). This strategy cooperates in the first round and from then on imitates the opponents move of the previous round. Thus, TFT cooperates against kin, but retaliates against defectors. The weak side of TFT becomes apparent in noisy environments where erroneous moves occur. In long-lasting interactions, this reduces the payoff of TFT against itself to that of a random strategy which cooperates or defects by tossing a coin. This serious drawback of TFT was resolved by another simple strategy called *Pavlov* (Nowak & Sigmund, 1993). It follows the concept of win-stay, lose-shift: a payoff of R or T is considered as a win and *Pavlov* sticks to its decision, but switches whenever it gets S or P points. This simple rule enables *Pavlov* to readily correct errors when playing against kin and to return to cooperation.

Other prominent strategies include the always cooperate and always defect strategies *all C* and *all D*, *Grim* (Binmore & Samuelson, 1992), *generous TFT* (Nowak & Sigmund, 1992) and *Firm-but-Fair* (FbF) (Freen, 1994; Nowak & Sigmund, 1994). All these theoretical investigations were complemented by many biological field observations as well as experiments on human and animal behavior (Dugatkin, 1997; Milinski, 1987; Milinski & Wedekind, 1998; Packer, 1977; Wedekind & Milinski, 1996; Wilkinson, 1984).

The basic mechanism that promotes cooperation in repeated interactions is summarized by a quote of Axelrod stating that the *shadow of the*

future, i.e. the fear from retaliation, motivates individuals to cooperate in the present.

1.2. MEMORY ONE STRATEGIES AND BEYOND

All strategies introduced so far belong to the class of memory one strategies. This means that the conditional probabilities to cooperate require the memory of one round. The strategic behavior is encoded by a quadruple of parameters (p_R, p_S, p_T, p_P) denoting the probabilities to cooperate after receiving R, S, T or P points, respectively. For example, *TFT* corresponds to $(1, 0, 1, 0)$, *Pavlov* to $(1, 0, 0, 1)$, *Grim* to $(1, 0, 0, 0)$, *FbF* to $(1, 0, 1, 2/3)$, etc. If erroneous moves occur with a small probability p_{err} the following changes apply: $1 \rightarrow 1 - p_{err}$ and $0 \rightarrow p_{err}$.

Interestingly, generalizing the concept of memory one strategies to include longer memory sizes did not reveal substantially new strategical patterns (Hauert & Schuster, 1997). Apparently, the crucial information is covered by the knowledge about the outcome of the previous round.

In the following, we refer to memory one strategies in order to demonstrate the superiority of a new and conceptually different class of strategies. These strategies accumulate knowledge about the history of the game in an internal state which then determines the probability to cooperate in the following round. This approach enables strategies to adjust their strategical behavior to the character of the opponent and to adopt the most promising behavior. For this reason, we call these strategies *Adaptors*.

2. Concept of *Adaptor*

In the context of direct reciprocity, apparently the majority of people does not behave according to the rules prescribed to *homo oeconomicus*. A series of experiments has significantly contributed to the decline of this rationality concept (see, e.g. Fehr & Gächter, 1998, 1999; Henrich *et al.*, 2001; Nowak *et al.*, 2000; Wedekind & Milinski, 1996) and have demonstrated that other factors such as fairness considerations and emotions often play a dominating role in at least human interactions and decision making.

Memory one strategies represent nothing more but simple automata. The outcome of the previous round serves as an input that is

translated directly (by means of the rules defining the automaton) into an output, i.e. the subsequent move. But real players engaged in the *IPD* hardly fit into this simple framework. That is mainly because they process information and eventually develop an intuition and awareness about the character of their opponent.

Our motivation for *Adaptor* is to go beyond the limitations of memory one strategies towards a more realistic scenario by implementing a simple yet effective information processing scheme, i.e. to provide a framework for suitable learning rules. These rules are based on an internal state which can be viewed as memory and interpreted in terms of emotions.

Mathematically speaking, we consider strategies that trigger their next move on an internal state s^τ . This internal state summarizes the outcomes of the past rounds with one particular opponent up to time τ . The outcome of one round modifies s^τ according to the following rule:

$$s^{\tau+1} = s^\tau + \Delta^\tau, \quad (3)$$

where Δ^τ takes the values $\delta R, \delta S, \delta T, \delta P$ if round τ yields a payoff of R, S, T or P points, respectively. The average number of rounds, i.e. the average length l of an interaction with one particular opponent is specified by the continuation probability w : $l = 1/(1 - w)$.

The internal state s^τ specifies the probability to cooperate $p_c(s^\tau)$ in the subsequent round. We consider a particularly simple functional shape of $p_c(s^\tau)$ where a cooperative region is bounded by defecting ones (see Fig. 1):

$$p_c(s^\tau) = p_{err} + (1 - 2p_{err})[\Theta(s^\tau + 1) - \Theta(s^\tau - 1)], \quad (4)$$

where Θ denotes the Heaviside step-function and $p_{err} \ll 1$ a small amount of noise, i.e. the probability to misimplement a move. This means that only for $s^\tau \in [-1, 1]$ *Adaptor* most likely cooperates but defects otherwise. Since the interval of s^τ is not bounded, the chosen width of the cooperative region poses no restriction of generality. Its width relative to the values of $\delta R, \delta S, \delta T$, and δP determines the strategical characteristics of *Adaptor*. Actually, this defines

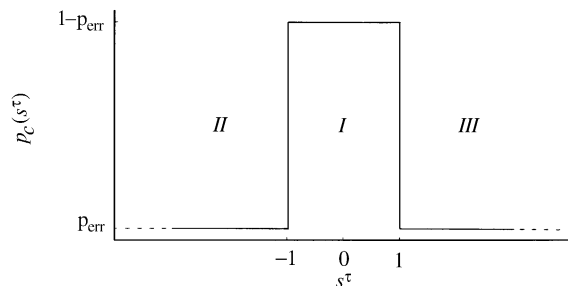


FIG. 1. The probability to cooperate p_c of *Adaptor* depends on its internal state s^τ in round τ . s^τ changes in the course of an *IPD* interaction [see eqn (3)] and triggers the subsequent move of *Adaptor*. As long as s^τ remains in region I, *Adaptor* most likely cooperates, but defects as soon as s^τ enters regions II or III. p_{err} indicates a small probability to mis-implement a move.

the learning rule of *Adaptor*, i.e. prescribes how to tune subsequent decisions and adapt to an opponent. The concept of *Adaptor* essentially allows to classify opponents either as dumb or unconditional cooperators, defectors or retaliators. If *Adaptor* realizes that an opponent can be exploited, the internal state may leave the cooperative regime turning *Adaptor* into a mean and unforgiving character. Similarly, against defective opponents, *Adaptor* may quickly adopt a defensive strategy and retaliate with defection. Only in balanced interactions, where erroneous defection is met by retaliation but apologizing handshakes are accepted, cooperation persists. In this sense, cooperative behavior represents a delicate equilibrium between testing the opponents strategy and the ability to correct errors. However, note that this summarizes only the key characteristics of *Adaptor*'s framework. It remains to be seen, to what extent successful learning rules actually take advantage of these possibilities.

Between memory one strategies and the concept of *Adaptor* a two-fold analogy exists: the outcome of the previous round serves as input and both depend on four parameters. The choice of the cooperation probabilities p_R, p_S, p_T, p_P is crucial for the success of memory one strategies, as is the choice of $\delta R, \delta S, \delta T$, and δP for the success of *Adaptor* and, in particular, of its error correcting mechanism. In addition, the analogy manifests itself in the fact that *Adaptor* reproduces many well-known memory

one strategies for specific choices of its four parameters.

Also note the conceptual differences to other strategies implementing an internal state like *contrite Tit-for-Tat (cTFT)* based on the standing of a player (Boerlijst *et al.*, 1997). *cTFT* needs to know the internal state of the opponent to find its next decision, whereas the internal state of *Adaptor* is invisible to the outside world. In that sense, relating the internal state to emotions, *cTFT* requires additional information on the standing of the opponent while *Adaptor*'s decisions are based solely on its own experiences. Another interesting concept in this context refers to strategies implementing an aspiration level (Posch, 1999; Posch *et al.*, 1999). The most prominent of these strategies is certainly the aforementioned *Pavlov*.

3. Evolving *Adaptor*

Successful learning rules of *Adaptor*, i.e. promising sets of parameter values are determined by extensive simulations. We evolve a large and generally heterogeneous population of *Adaptors* under the influence of selection and mutation over many generations. In each generation, all individuals engage in an *IPD* with all others. According to the rules of evolutionary game theory (Maynard Smith, 1982), the payoff achieved by an individual determines its fitness, i.e. its reproductive success. This means that only strategies achieving higher than average fitness will spread in the population and those with lower payoffs are bound to go extinct.

The selection reproduction scheme is governed by the difference equation:

$$x_i^{t+1} = x_i^t f_i^t / \sum_j x_j^t f_j^t, \quad (5)$$

where x_i^t denotes the frequency of strategy i and f_i^t its fitness at time t . f_i^t corresponds to the payoff achieved by strategy i when engaging in the *IPD* with all strategies present in the population. In the continuous time limit, eqn (5) reduces to the replicator equation (Hofbauer & Sigmund, 1998).

For the simulations, we set the cost of cooperation $c = 1$ and the benefit to $b = 3$. According to the characteristics of the replicator

dynamics, this is equivalent to setting $R = 3$, $S = 0$, $T = 4$ and $P = 1$. At time $t = 0$, we start with a homogeneous population of *Adaptors* with randomly drawn parameters $\delta R, \delta S, \delta T, \delta P$. The four parameter values are real numbers from the interval $[-2, 2]$. The restriction to this interval guarantees that the internal state may not cross from one defective domain to the other (regions II, III in Fig. 1, eqn (3)) without visiting the cooperative domain at least once (region I in Fig. 1). Because $p_c(s^r)$ is an even function [see eqn (4)], we may restrict either one of the four parameters to $[0, 2]$ without loss of generality. We exploit this symmetry to reduce the parameter space and choose $\delta T \in [0, 2]$. Finally, all *Adaptors* start their interactions with an initial value of the internal state $s^0 = 0$. This choice is motivated by our primary interest in cooperative solutions of the *IPD* and this makes sure that all *Adaptors* cooperate at least on their first move.

On average, once in 1000 generations a mutation occurs and a new brand of *Adaptor* with randomly drawn parameters is introduced. The initial frequency of the mutant is set to $f_{mut} = 0.11\%$. Whenever the frequency of any strategy drops below a certain noise level of $f_{noise} = 0.1\%$, it is removed. This corresponds

essentially to considering a population of size 1000. Ten such simulation runs are carried out over 10^8 generations. The results clearly indicate that only strategies with $\delta R \approx 0$ are capable of achieving payoffs close to the maximum R . For this reason, we did another ten simulation runs with $\delta R = 0$, fixed.

In a second stage, all strategies present at the end of the 20 simulation runs are merged into a new population. The initial frequencies of the strategies are set proportional to their frequencies at the end of the respective simulation run. This population is evolved in absence of mutation according to eqn (5) until only one strategy survives.

This entire procedure is carried out for two different values of the continuation probability w : brief interactions ($w = 0.9$) with an average length of $l = 10$ rounds and long interactions ($w = 0.999$) with $l = 1000$ rounds. Figure 2 shows the two runs that eventually lead to the most successful set of parameters for *Adaptor*. The resulting numerical values are summarized in Table 1.

Note that if we additionally include memory one strategies with randomly drawn $p_R, p_S, p_T, p_P \in [p_{err}, 1 - p_{err}]$, the following results remain

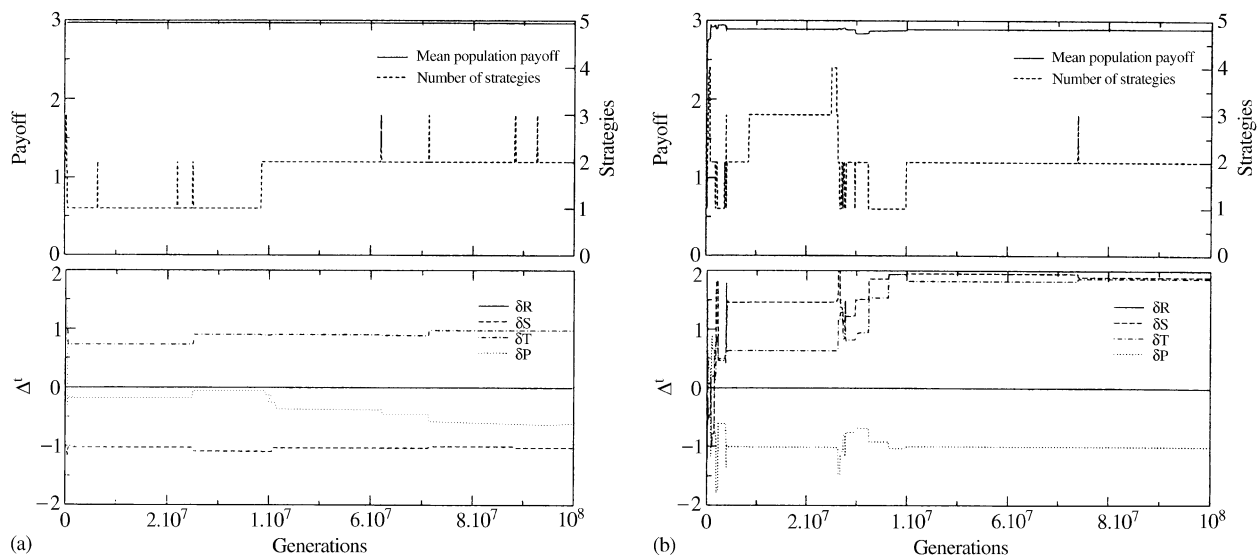


FIG. 2. Evolution of *Adaptor's* learning rule determined by the four parameter values $\delta T, \delta R, \delta P, \delta S$. The runs that eventually lead to the most successful strategical patterns are shown for (a) $w = 0.9$ and (b) $w = 0.999$. In both runs, the mean population payoff lies close to the maximum $R = 3$ indicating highly cooperative strategies. The main difference concerns the parameter δS which is negative for $w = 0.9$, but positive for $w = 0.999$. This leads to different strategical behavior and important changes in the error correction mechanisms.

TABLE 1

Results of evolving *Adaptor's* learning rule determined by the parameters $\delta S, \delta P, \delta R,$ and δT for two values of the continuation probability w . By definition $\delta T > 0$ holds, but the changing sign of δS gives rise to important differences in the strategical patterns of *Adaptor* in brief and long lasting interactions (see text)

w	δR	δS	δT	δP
0.9	0	-1.001505	0.992107	-0.638734
0.999	0	1.888159	1.858883	-0.995703

unchanged. Generally, *Adaptors* with the same strategical characteristics emerge. Only in few cases and only for $w = 0.999$ *Pavlov*-like strategies were found. But this is due to the finite simulation time, in Section 4 we actually show that *Adaptor* is able to outperform *Pavlov*.

Interestingly, significantly different strategical patterns evolve for brief and long-lasting interactions. This difference essentially results from δS and δT having opposite signs for $w = 0.9$, but are both positive for $w = 0.999$.

3.1. BRIEF INTERACTIONS

In brief interactions $w = 0.9$ the learning rule of *Adaptor* exploits all the features provided by the framework: exploitation of cooperators, retaliation against defectors and cooperation against retaliators and kin. Against defective opponents like *all D*, the internal state takes its values primarily in region II and hence *Adaptor* acts not cooperatively. Similarly, *Adaptor* tends to defect against unconditional cooperators like *all C* because its internal state primarily lies in sector III. Typical payoff values against prominent memory one strategies are gathered in Table 2. Another important feature of *Adaptor* is its error correcting capability when playing against kin. A typical sequence of moves following an error is depicted in Table 3. The number of defective moves played to correct an error is minimal. Compared to *Pavlov's* error correction scheme *Adaptor* plays no round of mutual defection. This is in favor of *Adaptor's* score against itself. At the same time *Adaptor* gets hardly exploited by *all D*, but is also less

TABLE 2

Payoff for several important strategies in the simultaneous IPD with $w = 0.9$ [c.f. Fig. 3(a)]*

$w = 0.9$	<i>Adaptor</i>	<i>c all C</i>	<i>d all D</i>	<i>c Pavlov</i>
<i>Adaptor</i>	2.966	2.994	0.916	2.826
<i>c all C</i>	2.944	2.982	0.035	2.636
<i>c all D</i>	1.322	3.964	1.018	2.570
<i>c Pavlov</i>	2.772	3.097	0.501	2.951

*The entries indicate the payoffs achieved by the strategy listed in the first column against the row strategy. The lowercase *c* and *d* in front indicates the initial move. The probability of mis-implementing moves is set to $p_{err} = 0.01$, i.e. the strategy quadruple for e.g. *all C* is (0.99, 0.99, 0.99, 0.99).

TABLE 3

Error correcting pattern of *Adaptor* against kin for $w = 0.9$ [☆]

<i>Adaptor</i> ₁	...	<i>C</i>	<i>D</i> *	<i>C</i>	<i>C</i>	...
<i>Adaptor</i> ₂	...	<i>C</i>	<i>C</i>	<i>D</i>	<i>C</i>	...

[☆]After playing an erroneous D, marked by an asterisk, the mistaken *Adaptor* offers a handshake by switching to *C* and accepts punishment for the erroneous move. The number of D's played is minimal, so that the loss during the error correction procedure is minimized. Once the error is corrected, the internal states are equal and almost unchanged. For short interaction lengths, it is therefore highly unlikely that another type of error correction scheme would be ever observed.

ready to exploit *all C*. It has to be mentioned, however, that *Adaptor's* error correction scheme cannot prevent conflicts from escalating if an unfortunate sequence of fast following erroneous moves occurs.

In order to investigate the robustness of this concept, we consider variations of two external parameters: the continuation probability w and the temptation to defect T . Figure 3(a) shows the payoff of several strategies as a function of w . Most importantly this demonstrates that (a) *all D* always scores significantly less against *Adaptor* than against *Pavlov* and (b) for long interactions ($l > 500$) *Adaptor's* payoff decreases when playing against its kin. This is a consequence of the small probability of escalations resulting in deadlocks of mutual defection.

Figure 4(a) similarly shows the payoff of the same strategies as a function of T . *Adaptor* and *Pavlov* do almost equally well: both perform

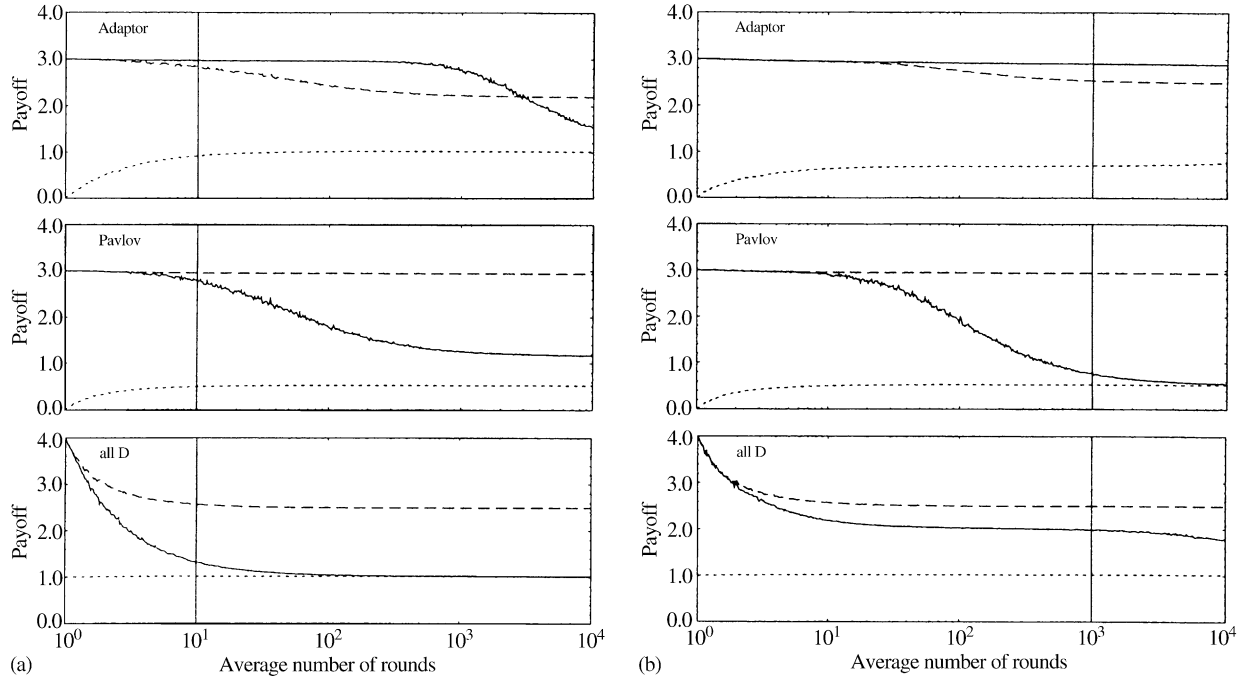


FIG. 3. Effects of the average interaction length on the payoffs of *Adaptor*, *Pavlov*, and *all D*. *Adaptor*'s learning rule is tuned for (a) short interactions ($w = 0.9$) and (b) for long-lasting encounters ($w = 0.999$) with $p_{err} = 0.01$. Vertical markers indicate the interaction lengths when evolving *Adaptor*'s learning rule. The three rows refer to the payoffs achieved (from top to bottom) by *Adaptor*, *Pavlov* and *all D*. Solid lines indicate payoffs against *Adaptor*, dashed lines against *Pavlov*, and dotted lines against *all D*. Key points to note: (1) When playing against its own kin, *Adaptor* achieves payoffs close to the maximum of $R = 3$, provided that the interaction length does not significantly exceed the original length (note the logarithmic scale of the ordinate). (2) *Adaptor* does systematically better against *Pavlov* than vice versa which is more pronounced for long interactions. (3) *All D* scores significantly less against *Adaptor* than against *Pavlov*.

slightly better when playing against kin than against the other strategy. However, *Adaptor* scores roughly twice as much against *all D*. This is also reflected in the score of *all D* against the two strategies: while the score against *Adaptor* increases very slowly with T , it increases significantly faster against *Pavlov*. It is the score of *all D* against *Pavlov* that delimits the parameter range where *Pavlov* is a limit ESS [see Section 4 and eqn (6)].

3.2. LONG INTERACTIONS

In long-lasting interactions $w = 0.999$, the efficient exploitation of unconditional cooperators as well as the relentless retaliation against defectors become apparently far less relevant compared to the enhanced importance of error correcting abilities. The most successful learning rule basically prevents the internal state from entering region II because δS and δT have the

same sign. The resulting strategical patterns are closely related to *Pavlov*. This is not too surprising because *Pavlov* implements probably the most robust error correction mechanism. Typical payoff values are compiled in Table 4 and the relevant error correction scheme of *Adaptor* is illustrated in Table 5. The first error is corrected identical to *Pavlov*, i.e. a single round of mutual defection is required to overcome the error. From then on, *Adaptor* plays two successive rounds of mutual defection before returning to cooperation (Lindgren, 1991). Occasionally, a more complicated sequence of moves may occur before cooperation is re-established. Unlike for brief interactions, the risk of escalations is almost negligible since it requires synchronized occurrences of erroneous moves.

The robustness of *Adaptor* is again investigated by varying w and T . Figure 3(b) shows that *Adaptor*'s payoff remains essentially unaffected by the interaction length. At the same

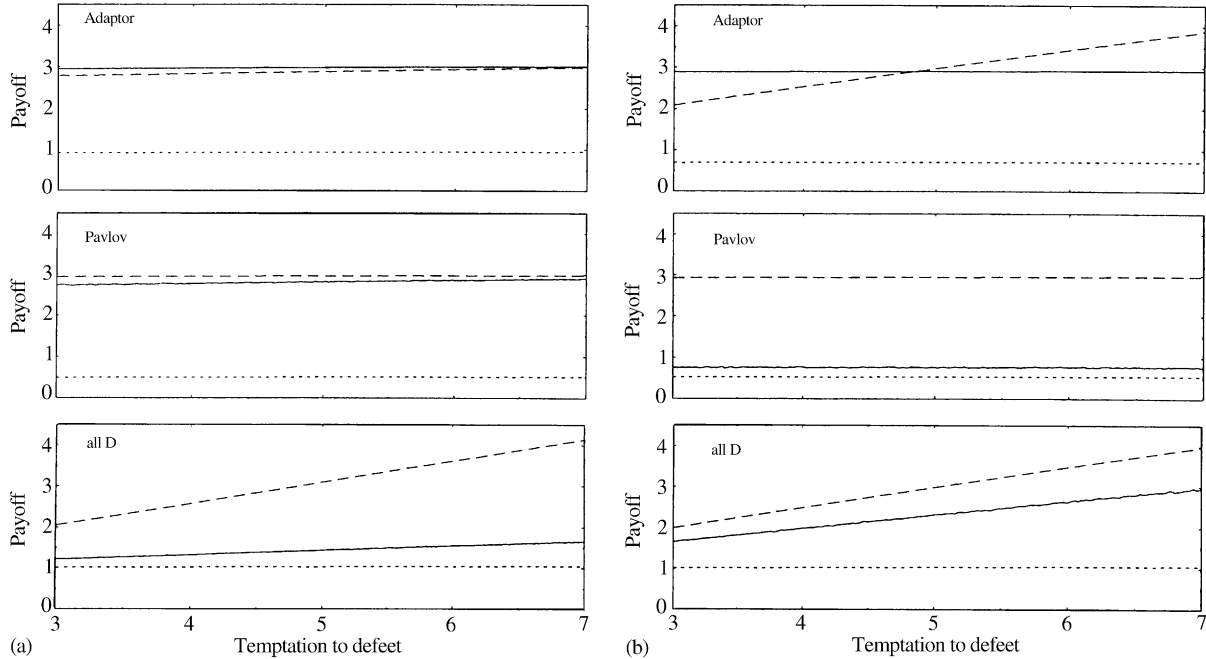


FIG. 4. Effects of varying the temptation on the payoffs of *Adaptor*, *Pavlov*, and *all D*. *Adaptor* is evolved for a temptation $T = 4$. For our choice of $R = 3, S = 0$ and $P = 1$, the range of acceptable values for T , i.e. satisfying the conditions of the *IPD* [see eqns (1), (2)] must lie in the open interval $T \in (3, 6)$. Arrangement of the graphs and legend is as in Fig. 3. Note that the payoffs of *Pavlov* and *all D* against kin and each other are almost identical for (a) $w = 0.9$ and (b) $w = 0.999$. As in Fig. 3, *Adaptor* always scores more against *Pavlov* than vice versa and *all D* does significantly better against *Pavlov* than against *Adaptor*.

TABLE 4

Payoff for several important strategies in the simultaneous *IPD* with $w = 0.999$ [c.f. Fig. 3(b)]. The entries are defined as in Table 2

$w = 0.999$	<i>Adaptor</i>	<i>c all C</i>	<i>d all D</i>	<i>c Pavlov</i>
<i>Adaptor</i>	2.900	3.912	0.698	2.534
<i>c all C</i>	0.184	2.980	0.040	1.549
<i>d all D</i>	1.986	3.960	1.020	2.491
<i>c Pavlov</i>	0.761	3.457	0.530	2.941

TABLE 5

Typical error correcting mechanism of *Adaptor* against its own kin for $w = 0.999$ [☆]

<i>Adaptor</i> ₁	...	<i>C</i>	<i>D</i> *	<i>D</i>	<i>D</i>	<i>C</i>	...
<i>Adaptor</i> ₂	...	<i>C</i>	<i>C</i>	<i>D</i>	<i>D</i>	<i>C</i>	...

[☆]The erroneous move is marked by an asterisk. The first error is corrected according to the well-known scheme of *Pavlov* (not shown). From then on, errors are followed by two rounds of mutual defection. Occasionally, errors are corrected involving more complicated sequences of moves requiring as much as seven rounds (not shown)

time, the score of *Pavlov* against *Adaptor* crucially depends on w . For $w = 0.999$, its payoff is almost as low as against *all D*. As for $w = 0.9$, *all D* scores less against *Adaptor* than against *Pavlov*.

Effects of variations of T are shown in Fig. 4(b). *Adaptor*'s performance against kin is almost independent of T and remains close to the maximum R . The payoffs of *Pavlov* and *all D* against kin and each other are largely

equal for brief and long interactions. Significant differences occur only in the payoffs of *Adaptor* against *Pavlov* and *all D* and vice versa. *Adaptor* is able to exploit *Pavlov* while *Pavlov* earns little more than against *all D*. Compared to $w = 0.9$, *Adaptor* does less well against *all D*, but the latter scores still significantly less against *Adaptor* than against *Pavlov*.

Important distinctions between *Adaptor* and *Pavlov* concern the formers ability to reduce its

loss against *all D*—note that this results from the second round of mutual defection in the error correcting scheme of *Adaptor*. At the same time *Adaptor* exploits *all C* more effectively. Even after an erroneous *D* of *all C*, *Adaptor* does not return to cooperation.

4. *Adaptor* vs. Memory one Strategies

To analyse *Adaptor's* performance in depth, we carry out evolutionary tournaments in which *Adaptor* competes with memory one strategies. One might argue that such a comparison is unfair because *Adaptor* has a higher memory capacity than memory one strategies. *Adaptor* is capable of storing a real number, whereas memory one strategies are storing merely an integer variable which takes on the values *R, S, T* or *P*. However, we think that such a comparison is valuable for two reasons. First, memory one strategies have gained most interest in the past and are best studied. Second, generalizing the concept of memory one strategies to include longer memory sizes apparently does not lead to substantial changes in behavior and performance compared to *Pavlov* (Hauert & Schuster, 1997). Among the memory one strategies, we restrict ourselves to those which may be specified by (p_R, p_S, p_T, p_P) with $p_i \in \mathfrak{S} = \{p_{err}, 1/3, 2/3, 1 - p_{err}\}$, $i \in \{R, S, T, P\}$ and initial moves of *C* or *D*. Note that most of the strategies well established in literature are covered by this choice.

For our parameter values *Pavlov* is a limit ESS (Leimar, 1997) for a sufficiently high continuation probability w :

$$w > \frac{T - R}{R - P} \quad (6)$$

(Nowak & Sigmund, 1995). However, note that *Pavlov* is not an evolutionarily stable strategy (Boyd & Lorberbaum, 1987; Lorberbaum, 1994) because strategies doing equally well could invade and possibly undermine the resident population through random drift. Evolutionary stability (Maynard Smith, 1982), however, requires that *no* mutant strategy introduced with an arbitrary small frequency is able to invade the resident population.

Our tournaments are based on the usual selection dynamics [see eqn (5)] augmented by mutation. Mutations occur on average every 1000 generations. However, instead of introducing mutants at a fixed initial frequency f_{mut} , we proceed in analogy to the optimization technique of simulated annealing: the initial frequency $\hat{f}_{mut}(t)$ assigned to mutants decreases with time t :

$$\hat{f}_{mut}(t) = \frac{1}{2} - \frac{1/2 - f_{noise}}{t_{max}} t, \quad (7)$$

where t_{max} specifies the total number of generations. Thus, early mutants are introduced with frequencies close to 1/2 while late mutants occur with frequencies close to the noise level $f_{noise} = 0.1\%$. If the frequency of any strategy drops below this fixed noise level, it is removed. New strategies are drawn with equal probabilities out of a set consisting of *Adaptor* and memory one strategies with $p_i \in \mathfrak{S}$. Since we consider only a finite number of strategies, each strategy is selected on average roughly 20 times ($t_{max} = 10^7$) at different frequencies of the mutant. Note that this estimation distinguishes between strategies playing *C* or *D* on their first move.

At time $t = 0$, we start with a homogeneous population playing a completely random strategy (0.5, 0.5, 0.5, 0.5). We perform these simulations for both $w = 0.9$ and 0.999. Typical simulation runs are depicted in Fig. 5. In all our simulations, we observed similar patterns where *Adaptor* takes over the population after some stage of equilibration. When introduced with high frequency mutants can invade the population until *Adaptor* takes over again. This interplay manifests itself in a sequence of meta-stable states. With decreasing $\hat{f}_{mut}(t)$, the population of *Adaptors* becomes noteworthy stable. For $w = 0.9$, all invasion attempts fail for $\hat{f}_{mut}(t) \lesssim 44\%$. The same holds for $w = 0.999$ with $\hat{f}_{mut}(t) \lesssim 38\%$. In both cases, the average payoff in homogeneous populations of *Adaptors* is close to the maximum value R , i.e. cooperative behavior dominates. The principle finding of our simulations is that *Adaptor* is capable of invading a population playing *Pavlov* whereas *Adaptor* is immune to invasions, in particular by *Pavlov*.

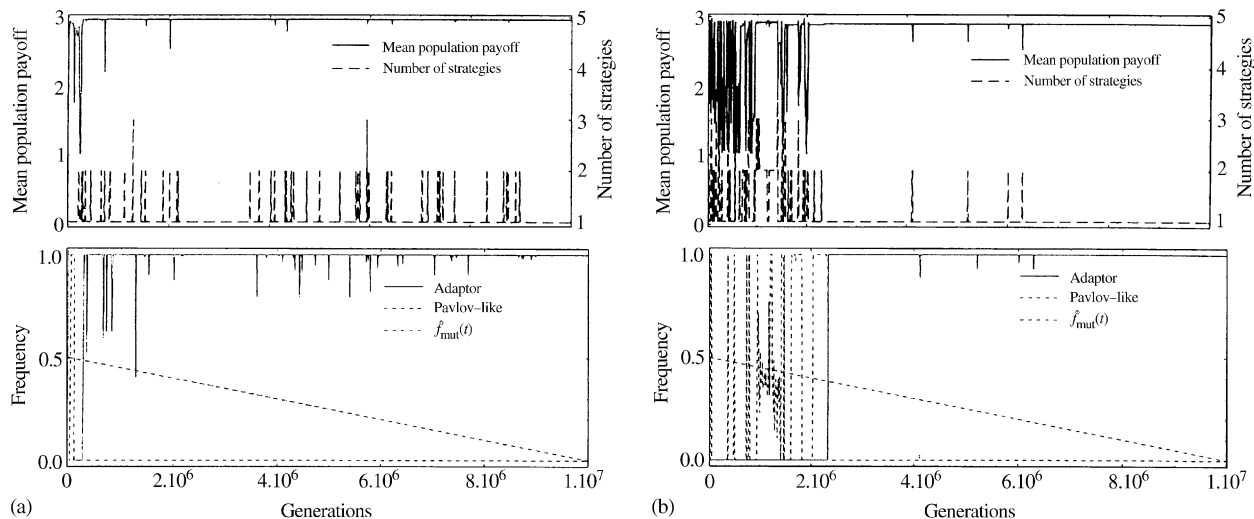


FIG. 5. Evolutionary simulations of *Adaptor* vs. memory one strategies for (a) $w = 0.9$ and (b) $w = 0.999$ in typical simulation runs with $p_{err} = 0.01$. The average population payoff and the number of strategies present in the population are displayed at the top, the initial frequency of mutants $f_{mut}(t)$ together with the frequency of *Adaptor* and *Pavlov*-like strategies at the bottom. If introduced with sufficiently high frequencies, mutants can temporarily invade the population until *Adaptor* takes over again. For $f_{mut}(t)$ below roughly 44% ($w = 0.9$) and 38% ($w = 0.999$), respectively, successful invasions are no longer observed.

5. Conclusions

The concept of *Adaptor* was motivated by experimental findings documenting the importance of emotions rather than rational reasoning in at least human decision making. The concept of the *Adaptor* strategy provides a powerful framework for learning rules attempting to model this process. *Adaptor* is determined by four parameters $\delta R, \delta S, \delta T$ and δP defining the learning rule together with an internal state summarizing the course of the game. This internal state triggers *Adaptor*'s subsequent moves and enables the strategy to adjust its decisions to the opponent and to adapt the most promising strategical behavior.

Through evolutionary simulations, we determine successful learning rules for brief and long-lasting interactions. In both cases, *Adaptor* turns out to be highly cooperative when playing against kin, implementing an efficient error correcting mechanism. However, as in real situations, it cannot prevent conflicts from escalating if unfortunate sequences of fast following erroneous moves occur.

The results clearly show that cooperative behavior strongly correlates with $\delta R \approx 0$, i.e.

the internal state remains unchanged after obtaining R points. This seems reasonable because it prevents the internal state from leaving the cooperative regime without cause. The other three parameters determine *Adaptor*'s error correcting scheme as well as its behavior when competing against other strategies. The resulting strategical patterns depend on the characteristic interaction length as determined by the continuation probability w . For brief interactions ($w = 0.9$), δS and δT have opposite signs. This enables *Adaptor* to distinguish between unconditional cooperators, defectors and retaliators reflected by the three regimes of the internal state. The strategical patterns suggest an interpretation in emotional terms as satisfied, frustrated and ruthless. Dumb cooperators are remorselessly exploited while defectors face tough retaliation. Cooperative behavior then resembles a tightrope walk between testing the opponents strategy and the ability to readily correct erroneous moves and re-establish cooperation.

The importance of efficient and reliable error correcting mechanisms is far more pronounced in long-lasting interactions ($w = 0.999$). This is reflected in δS and δT having the same signs

which leads to strategical patterns closely related to *Pavlov*. Consequentially, only two out of the three regimes of the internal state essentially determine *Adaptor's* behavior. This blurs the above distinction between ruthlessness and frustration. In that case, even against defectors, *Adaptor* occasionally offers a cooperative handshake hoping to re-establish cooperation.

The performance of *Adaptor* is surprisingly robust to environmental changes, i.e. variations of external parameters. Because of its parameterization, *Adaptor* is immune to changes of the payoff values R, S, T and P . Basically, this also holds for its performance against memory one strategies. With respect to changes in the average length of an interaction, the score of *Adaptor* against kin remains largely unaffected. However, for interaction lengths that considerably exceed the settings when evolving the learning rule, escalations leading to deadlocks in states of mutual defection become important.

Of particular interest was *Adaptor's* performance against memory one strategies well established in the literature. In another set of evolutionary simulations *Adaptor* competes against memory one strategies in a population where mutants are introduced with decreasing frequency. The main results of these simulations show: (a) *Adaptor* is superior in performance to memory one strategies. In particular, it outperforms *Pavlov* known as a most successful strategy in the *IPD*. One major reason for this is certainly its better ability to resist exploitation by *all D*. In long interactions, *Adaptor* resembles *Pavlov* but after an erroneous defection it resumes cooperation only after two consecutive rounds of mutual defection (see also Lindgren, 1991). (b) *Adaptor* displays a remarkable robustness, i.e. it is not only stable with respect to arbitrary small perturbations as requested by evolutionary stability, but also to mutants introduced with frequencies as high as roughly 40%. This makes *Adaptor* a stable cooperative strategy even in tiny populations or groups of as few as three individuals.

Many thanks to Karl Sigmund for inspiring discussions and helpful comments on the manuscript. We acknowledge support of the Sonderforschungsbereich 237 *Unordnung und große Fluktuationen*

of the Deutsche Forschungsgemeinschaft. Ch.H acknowledges support of the Swiss National Science Foundation.

REFERENCES

- AXELROD, R. (1984). *The Evolution of Cooperation*. New York: Basic Books.
- AXELROD, R. & HAMILTON, W. D. (1981). The evolution of cooperation. *Science* **211**, 1390.
- BINMORE, K. G. & SAMUELSON, L. (1992). Evolutionary stability in repeated games played by finite automata. *J. Econ. Theory* **57**, 278.
- BOERLIJST, M. C., NOWAK, M. A. & SIGMUND, K. (1997). The logic of contrition. *J. theor. Biol.* **185**, 281.
- BOYD, R. & LORBERBAUM, J. M. D. (1987). No pure strategy is evolutionary stable in the iterated prisoner's dilemma game. *Nature* **327**, 58.
- COLMAN, A. M. (1995). *Game Theory and its Applications in the Social and Biological Sciences*. Oxford: Butterworth-Heinemann.
- DUGATKIN, L. A. (1997). *Cooperation among Animals: an Evolutionary Perspective*. Oxford: Oxford University Press.
- FEHR, E. & GÄCHTER, S. (1998). Reciprocity and economics: the economic implications of homo reciprocans. *Europ. Econ. Rev.* **42**, 845.
- FEHR, E. & SCHMIDT, K. (1999). A theory of fairness, competition, and cooperation. *Quart. J. Econ.* **114**, 817.
- FREAN, M. R. (1994). The prisoner's dilemma without synchrony. *Proc. R. Soc. Lond. B* **257**, 75.
- HAUERT, C. (2001). Fundamental clusters in spatial 2×2 games. *Proc. R. Soc. Lond. B* **268**, 761.
- HAUERT, C. (2002). Effects of space in 2×2 games. *Int. J. Bifurcation and Chaos*. **12**.
- HAUERT, C. & SCHUSTER, H. G. (1997). Effects of increasing the number of players and memory size in the iterated prisoner's dilemma: a numerical approach. *Proc. R. Soc. Lond. B* **264**, 513.
- HENRICH, J., BOYD, R., BOWLES, S., CAMERER, C., FEHR, E., GINTIS, H. & McELREATH, R. (2001). Cooperation, reciprocity and punishment in fifteen small scale societies. *Am. Econ. Rev.* **91**, 73.
- HERZ, A. V. M. (1994). Collective phenomena in spatially extended evolutionary games. *J. theor. Biol.* **169**, 65.
- HOFBAUER, J. & SIGMUND, K. (1998). *Evolutionary Games and Population Dynamics*. Cambridge: Cambridge University Press.
- KAGEL, J. H. & ROTH, A. E. eds (1995). *The Handbook of Experimental Economics*. Princeton: Princeton University Press.
- KILLINGBACK, T., DOEBELI, M. & KNOWLTON, N. (1999). Variable investment, the continuous prisoner's dilemma, and the origin of cooperation. *Proc. R. Soc. Lond. B* **266**, 1723.
- LEIMAR, O. (1997). Repeated games: a state space approach. *J. theor. Biol.* **184**, 471.
- LEIMAR, O. & HAMMERSTEIN, P. (2001). Evolution of cooperation through indirect reciprocity. *Proc. R. Soc. Lond. B* **268**, 745.
- LINDGREN, K. (1991). In: *Artificial Life II* (Langton, C. G., Farmer, J. D., Rasmussen, S. & Taylor, C. (eds)) *Proceedings of the Santa Fe Institute Studies in the*

- Sciences of Complexity*, Vol. X, pp. 295–312. Reading, MA: Addison-Wesley.
- LINDGREN, K. & NORDAHL, M. G. (1994). Evolutionary dynamics of spatial games. *Physica D* **75**, 292.
- LORBERBAUM, J. M. D. (1994). No strategy is evolutionarily stable in the repeated prisoner's dilemma. *J. theor. Biol.* **168**, 117.
- 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**, 433.
- MILINSKI, M. & WEDEKIND, C. (1998). Working memory constrains human cooperation in the prisoner's dilemma. *Proc. Natl Acad. Sci. U.S.A.* **95**, 13 755.
- NOWAK, M. A. & MAY, R. M. (1992). Evolutionary games and spatial chaos. *Nature* **359**, 826.
- NOWAK, M. A. & SIGMUND, K. (1992). Tit for tat in heterogeneous populations. *Nature* **355**, 250.
- NOWAK, M. A. & SIGMUND, K. (1993). A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. *Nature* **364**, 56.
- NOWAK, M. A. & SIGMUND, K. (1994). The alternating prisoner's dilemma. *J. theor. Biol.* **168**, 219.
- NOWAK, M. A. & SIGMUND, K. (1995). Invasion dynamics of the finitely repeated prisoner's dilemma. *Games Econ. Behav.* **11**, 364.
- NOWAK, M. A. & SIGMUND, K. (1998). Evolution of indirect reciprocity by image scoring. *Nature* **393**, 573.
- NOWAK, M. A., PAGE, K. M. & SIGMUND, K. (2000). Fairness versus reason in the ultimatum game. *Science* **289**, 1773.
- PACKER, C. (1977). Reciprocal altruism in papio anubis. *Nature* **265**, 441.
- POSCH, M. (1999). Win-stay, lose-shift strategies for repeated games—memory length, aspiration levels and noise. *J. theor. Biol.* **198**, 183.
- POSCH, M., PICHLER, A. & SIGMUND, K. (1999). The efficiency of adapting aspiration levels. *Proc. R. Soc. Lond. B* **266**, 1427.
- TRIVERS, R. L. (1971) The evolution of reciprocal altruism. *Q. Rev. Biol.* **46**, 35.
- WEDEKIND, C. & MILINSKI, M. (1996). Human cooperation in the simultaneous and the alternating prisoner's dilemma: Pavlov versus generous tit-for-tat. *Proc. Natl Acad. Sci. U.S.A.* **93**, 2686.
- WEDEKIND, C. & MILINSKI, M. (2000). Cooperation through image scoring in humans. *Science* **288**, 850.
- WILKINSON, G. S. (1984). Reciprocal food-sharing in the vampire bat. *Nature* **308**, 181.