

## Error and Discounting in the Iterated Prisoner's Dilemma

D. W. STEPHENS, K. NISHIMURA<sup>†</sup> AND K. B. TOYER

*Nebraska Behavioral Biology Group, School of Biological Sciences, University of Nebraska,  
Lincoln, NE 68588, U.S.A.*

(Received on 6 July 1994, Accepted in revised form on 25 May 1995)

Since 1981 a large body of research has focused on the Iterated Prisoner's Dilemma as a "basic paradigm" for the study of non-kin cooperation. Current evidence, however, shows that animals consistently defect in controlled Prisoner's Dilemmas. In this paper, an attempt is made to understand this by studying the effects of error and discounting (the tendency to devalue future rewards) on the stability of two strategies, tit-for-tat and Pavlov, against the clear experimental winner, all-defection. When considering strategic error, it is found that there are some payoff combinations in which the "cooperative strategy" (tit-for-tat or Pavlov) can never be stable against all defection, and others where low levels of temporal discounting and/or large levels of game repetition can stabilize the cooperative strategy, as the conventional view suggests. These no-cooperation regions are characterized and compared for tit-for-tat and Pavlov. When tit-for-tat is pitted against all defection, however, there is also a third set of payoff combinations in which tit-for-tat is stable against all defection at intermediate levels, but unstable at both very low and very high levels of temporal discounting. For both strategies, increasing error rates increase this no-cooperation region. Similarly, as error increases weaker temporal discounting is required to stabilize a cooperative strategy against all defection. The most modest requirements for stability occur when the "temptation to defect" is negligible and the benefits of mutual cooperation greatly exceed the benefits of mutual defection. Estimates of the relevant discounting parameters are presented and discounting rates that are at least an order of magnitude smaller than values that seem plausible under the conventional "game repetition" view of the Iterated Prisoner's Dilemma are estimated. The prediction is that animal cooperation in Prisoner's Dilemmas will be restricted to a very small set of payoff combinations.

© 1995 Academic Press Limited

### Introduction

Since the publication of Axelrod and Hamilton's (1981) classic paper, empirical and theoretical studies of cooperation among unrelated animals (Lombardo, 1985; Milinski, 1987; Dugatkin, 1988; Boyd & Lorberbaum, 1987; Boyd, 1989; Nowak, 1990; Nowak & Sigmund, 1992, 1993) have been dominated by the Iterated Prisoner's Dilemma (IPD) game and variations of a strategy called tit-for-tat (Axelrod, 1984), in which an animal cooperates or "defects" (i.e. does not cooperate) if its opponent has done likewise at a previous encounter. Rigorous empirical justification for this preoccupation is, however,

difficult to find. There is currently no evidence of non-human animals cooperating in an empirically verifiable Prisoner's Dilemma (PD). The results of experiments conducted in controlled PDs paint an especially bleak picture with animals including rats (Flood *et al.*, 1983; Gardner *et al.*, 1984), humans (Scodel *et al.*, 1959; Rapoport & Chammah, 1965; Rapoport, 1974; Colman, 1982), and passerine birds (starlings, Reboreda & Kacelnik, personal communication; blue jays, Clements & Stephens, 1995) all favoring mutual defection.

We feel that animals fail to cooperate in IPDs because of two well substantiated properties of animal behavior: discounting and error. Animals have extraordinarily strong preferences for immediate reward (Kagel *et al.*, 1986; Logue, 1988); in economic

<sup>†</sup> Present address: Faculty of Fisheries, Hokkaido University, 3-1-1 Minato-cho, Hako-date, Hokkaido 041, Japan.

jargon, they discount the future heavily. Biological models of cooperation have relied exclusively on repetition to make cooperation “worthwhile” in IPD. This emphasis on repetition is especially striking in Nowak and Sigmund’s recent work using the asymptotic stationary distribution of Markov chains to calculate the benefits of given strategic interactions (Nowak, 1990; Nowak & Sigmund, 1992, 1993).

Similarly, it is exceedingly unlikely that an animal will follow a rigid behavioral rule like tit-for-tat or all-defection without variation or error. One might expect that error and discounting have similar effects on the value of alternative strategies. For example, in classical error-free tit-for-tat one supposes that cooperating animals forego the temptation to defect, (in part) because *certain* retaliation, costs more in the long-run than a player can net by defecting now. Discounting will tend to make defection now more profitable because it devalues the player’s expectation of “the long-run”. Similarly, strategic error weakens both the threat of retaliation and the reliability of future cooperation.

In this paper we analyze the interactions of strategic error and temporal discounting in the IPD. First, we present preliminary results and formalisms, including a straightforward generalization of the geometric series to the case where elements of the series are successive steps of a Markov chain. Second, we use the resulting formalism to analyze the effects of error and discounting on the venerable tit-for-tat strategy. Third, we study error and discounting in the version of the Pavlov strategy recently advocated by Nowak & Sigmund (1993). Finally, we present experimentally derived estimates of the relevant discounting parameters, and discuss their implications for cooperation in IPDs. In the interest of clear exposition, we comment on the implications of our results as they are developed, rather than deferring all these points to the discussion.

**Preliminaries**

SCALING AND DISCOUNTING

*The game*

The Prisoner’s Dilemma is a symmetric two-player game defined by the game matrix:

$$\begin{matrix} & C & D \\ C & \begin{bmatrix} R & S \end{bmatrix} \\ D & \begin{bmatrix} T & P \end{bmatrix} \end{matrix}$$

where *C* represents a strategy called “cooperate” and *D* represents “defection” (don’t cooperate). A PD

exists when  $T > R > P > S$ , where these payoffs are given the alluring names: the temptation to defect (*T*); the reward for mutual cooperation (*R*); the punishment for mutual defection (*P*); and the sucker’s payoff (*S*). The dilemma is that although *D* is a dominating strategy for a single play of the game, both players could do better if they cooperated by jointly choosing *C*, simply because  $R > P$ . It is important to note that the inequality  $R > P$  is the *defining* property of cooperation in the simple dictionary definition sense of “joint action for mutual benefit”. In addition, we follow biological tradition (Mesterton-Gibbons & Dugatkin, 1992; Mesterton-Gibbons, 1992) by requiring that  $R > (T + S)/2$  to ensure that mutual *C* is the only cooperative choice; because otherwise an alternating strategy might lead to higher mutual benefit.

*Scaling*

In determining the relative merits of alternative strategies in the Prisoner’s Dilemma, we will compare expected payoffs of sequences of plays. These comparisons, like the comparisons of expected utilities in classical micro-economic theory (Raiffa, 1968), are not affected by linear transformations of the payoffs. We capitalize on this insensitivity by reducing any PD to the standard form:

$$\begin{matrix} & C & D \\ C & \begin{bmatrix} \rho & 0 \end{bmatrix} \\ D & \begin{bmatrix} 1 & \beta\rho \end{bmatrix} \end{matrix}$$

with the rectangular parameter space  $1/2 < \rho < 1$  and  $0 < \beta < 1$ , where  $\rho = (R - S)/(T - S)$  and  $\beta = (P - S)/(R - S)$ . We call this parameter space the PD rectangle. The parameter  $\beta$  measures how cooperative the situation is (Fig. 1). If  $\beta$  is near one, there is only a trivial difference in payoff between mutual cooperation and mutual defection. Similarly,  $\rho$  reflects the relative magnitude of the temptation to defect. If  $\rho$  is small (near 1/2) the temptation is great; although, if  $\rho$  is large (near 1), the temptation is weak. In Fig. 1 we use this rectangular parameter space to created a crude classification of Prisoner’s Dilemmas.

*Discounting*

We consider a geometric discounting process in which the value of a reward obtained at the *n*-th step of the process is proportion  $\alpha^{n-1}$  ( $0 < \alpha < 1$ ) of the value of the same reward obtained at the first step of the process. Hence, if we know the sequence of rewards,  $\{x_1, x_2, \dots\}$ , then the total value of the process is

$$x_1 + \alpha x_2 + \alpha^2 x_3 + \dots$$

We refer to  $\alpha$  as the *discount factor*. We find it convenient in the calculations below to use the substitutions

$$\alpha = \frac{z}{z+1}$$

$$z = \frac{\alpha}{1-\alpha}$$

where  $z$  ranges from 0 to infinity, and may be thought of as the “perceived” repetition of the process; we call  $z$  the *subjective repetition*. We suppose that  $\alpha = vw$ , where  $w$  is the probability that the game will continue from one play to the next, and  $v$  is the proportion of value that is preserved from one play to the next. As we noted above, much of the biological literature of the IPD only considers repetition of the process; in this symbolism, it supposes that  $v = 1$ , so  $\alpha = w$ . Viewing  $\alpha$  as the product of two processes—value degradation and process termination—makes little practical difference to the mechanics of our calculations. The practical significance of this formulation is that one expects  $\alpha$  to be *smaller* in this more realistic combined process than in either a pure “value degradation” or a pure “process termination” interpretation of  $\alpha$ .

MARKOVIAN STRATEGIES

Following Nowak & Sigmund (1993) we consider the class of strategies represented by a 4-tuple of probabilities of cooperating after received each of the four possible payoffs in the Prisoner's Dilemma,  $\vec{\mathcal{P}} = (t, r, p, s) = (\text{Pr}(C|T), \text{Pr}(C|R), \text{Pr}(C|P), \text{Pr}(C|S))$ . In this notation, tit-for-tat corresponds to  $(1, 1, 0, 0)$ , Pavlov corresponds to  $(0, 1, 1, 0)$ , generous tit-for-tat corresponds to  $(1, 1, g, g)$ ,  $g \in (0, 1)$ . Then the rewards to  $\vec{\mathcal{P}}_1 = (t_1, r_1, p_1, s_1)$  against  $\vec{\mathcal{P}}_2 = (t_2, r_2, p_2, s_2)$  are determined by the Markov transition matrix,

$$\mathbf{M}_{\vec{\mathcal{P}}_1, \vec{\mathcal{P}}_2} = \begin{pmatrix} s_2(1-t_1) & (1-r_1)r_2 & (1-p_1)p_2 & (1-s_1)t_2 \\ s_2t_1 & r_1r_2 & p_1p_2 & s_1t_2 \\ (1-s_2)(1-t_1) & (1-r_1)(1-r_2) & (1-p_1)(1-p_2) & (1-s_1)(1-t_2) \\ (1-s_2)t_1 & r_1(1-r_2) & p_1(1-p_2) & s_1(1-t_2) \end{pmatrix}$$

where the entries in the first column represent the conditional probabilities receiving the payoffs  $T, R, P$  and  $S$  respectively in the next round given that payoff  $T$  was received in the current round. The second, third and fourth columns similarly represent probabilities conditioned on  $R, P$  and  $S$  probabilities in the obvious

way.

To completely specify a strategy one must specify not only the vector  $\vec{\mathcal{P}}$ , but also initial behavior; e.g. tit-for-tat is represented by  $\vec{\mathcal{P}} = (1, 1, 0, 0)$  and initial cooperation. We represent the probability of initial cooperation as  $c$ . So a strategy for player  $i$  is a 5-tuple  $A_i = \{\vec{\mathcal{P}}_i, c_i\}$ . Hence any two interacting strategies specify a matrix  $\mathbf{M}_{\vec{\mathcal{P}}_1, \vec{\mathcal{P}}_2}$ , and an initial distribution of payoffs

$$\mathbf{y}_{\vec{\mathcal{P}}_1, \vec{\mathcal{P}}_2} = \begin{pmatrix} (1-c_1)c_2 \\ c_1c_2 \\ (1-c_1)(1-c_2) \\ (1-c_2)c_1 \end{pmatrix}$$

If we let  $\mathbf{v}$  be the vector of payoffs  $(T, R, P, S)$ , then the expected benefit from a sequence of  $n$  plays with discount factor  $\alpha$  is given by

$$E_n = \mathbf{v} \cdot \mathbf{y} + \alpha \mathbf{v} \cdot \mathbf{M} \cdot \mathbf{y} + \alpha^2 \mathbf{v} \cdot \mathbf{M}^2 \cdot \mathbf{y} + \dots + \alpha^{n-1} \mathbf{v} \cdot \mathbf{M}^{n-1} \cdot \mathbf{y}$$

$$= \mathbf{v} \cdot \underbrace{(\mathbf{I} + \alpha \mathbf{M} + \alpha^2 \mathbf{M}^2 + \dots + \alpha^{n-1} \mathbf{M}^{n-1})}_{\mathbf{s}_n} \cdot \mathbf{y}$$

where we interpret  $\mathbf{v}$  to be row vector and  $\mathbf{y}$  to be a column vector. We recognize that as  $n$  increases to infinity the term in parenthesis,  $\mathbf{s}_n$ , which is analogous to a geometric series, may be written as

$$\mathbf{s}_\infty = (\mathbf{I} - \alpha \mathbf{M})^{-1}$$

if  $\alpha < 1$  (see, for example, lemma 8.19 of Burden & Faires, 1985). Applying this result, we see that the total benefit that a player adopting the strategy  $A_i = \{\vec{\mathcal{P}}_i, c_i\}$  obtains when playing against an opponent adopting strategy  $A_j = \{\vec{\mathcal{P}}_j, c_j\}$  from an indefinite sequence of plays can be written as

$$E_{i,j} = \mathbf{v} \cdot (\mathbf{I} - \alpha \mathbf{M}_{i,j})^{-1} \cdot \mathbf{y}_{i,j} \tag{1}$$

This formalism allows us to calculate the total benefit that derived by one arbitrary strategy playing another.

We rely on this formalism in the remainder of the paper. We note that in contrast to analyses using stationary distributions, this result does not require that  $\mathbf{M}$  be an ergodic (or mixing) transition matrix (Feller, 1950).

*Error*

Error can be added to any strategy  $\{(t, r, p, s), c\}$  in a straightforward way. Suppose that the probability of error is  $\epsilon$ , and that an error-free strategy specifies a “intended” probability of cooperation,  $\xi$ , in a given circumstance, ( $\xi \in \{t, r, p, s, c\}$ ) then the achieved probability of cooperation  $\xi_\epsilon$  is

$$\begin{aligned} \xi_\epsilon &= \xi(1 - \epsilon) + (1 - \xi)\epsilon \\ &= \xi(1 - 2\epsilon) + \epsilon \end{aligned}$$

The reader can confirm that choices of “intended” cooperation probabilities 0, 1 and 1/2, lead to realized cooperation probabilities of  $\epsilon$ ,  $1 - \epsilon$  and 1/2, respectively, as one expects.

**Tit-for-tat vs. AllD**

We consider the combined effects of error and discounting on two well-known strategies, tit-for-tat and Pavlov, in this section and the next. We consider what seems to us to be the most basic question that one can ask of a strategy in the IPD: under what conditions is it stable against all defection (AllD). This puts us at odds with the biological fashion of seeking a champion to unseat tit-for-tat (Boyd & Lorberbaum, 1987; Nowak & Sigmund, 1993). Again, our analytical tactics stem from empiricism. AllD is the most widely observed tactic in verifiable PDs.

Tit-for-tat is the strategy  $\mathcal{T}_0 = \{(1, 1, 0, 0), 1\}$ , and AllD is the strategy  $\mathcal{D}_0 = \{(0, 0, 0, 0), 0\}$ . When error is added to these strategies we have

$$\begin{aligned} \mathcal{T}_\epsilon &= \{(1 - \epsilon, 1 - \epsilon, \epsilon, \epsilon), 1 - \epsilon\} \\ \mathcal{D}_\epsilon &= \{(\epsilon, \epsilon, \epsilon, \epsilon), \epsilon\} \end{aligned}$$

For a game with only these two strategies, tit-for-tat ( $\mathcal{T}_\epsilon$ ) is stable against all-defection ( $\mathcal{D}_\epsilon$ ) when tit-for-tat

gains more playing against other tit-for-tat strategists than AllD gains playing against tit-for-tat. In our symbolism, this is

$$\mathbf{v} \cdot (\mathbf{I} - \alpha \mathbf{M}_{\mathcal{T}_\epsilon, \mathcal{T}_\epsilon})^{-1} \cdot \mathbf{y}_{\mathcal{T}_\epsilon, \mathcal{T}_\epsilon} > \mathbf{v} \cdot (\mathbf{I} - \alpha \mathbf{M}_{\mathcal{D}_\epsilon, \mathcal{T}_\epsilon})^{-1} \cdot \mathbf{y}_{\mathcal{D}_\epsilon, \mathcal{T}_\epsilon}$$

where our transition matrices and initial distributions are readily calculated from the results in the previous section. This is equivalent to studying the sign of the difference

$$\begin{aligned} \Delta &= \mathbf{v} \cdot \left( \mathbf{I} - \frac{z}{z+1} \mathbf{M}_{\mathcal{T}_\epsilon, \mathcal{T}_\epsilon} \right)^{-1} \cdot \mathbf{y}_{\mathcal{T}_\epsilon, \mathcal{T}_\epsilon} \\ &\quad - \mathbf{v} \cdot \left( \mathbf{I} - \frac{z}{z+1} \mathbf{M}_{\mathcal{D}_\epsilon, \mathcal{T}_\epsilon} \right)^{-1} \cdot \mathbf{y}_{\mathcal{D}_\epsilon, \mathcal{T}_\epsilon} \quad (2) \end{aligned}$$

where we have used the substitution  $\alpha = z/(z + 1)$ , and  $\mathbf{v}$  is defined to be  $(1, \rho, \beta\rho, 0)$ , following the re-scaling introduced in the previous section. The Appendix shows that this problem can, in turn, be reduced to the well-understood problem of studying the roots of a cubic polynomial in  $z$ . If there is a single positive simple root then we have a result that agrees qualitatively with Axelrod and Hamilton, i.e. there is a single level of repetition above which tit-for-tat is stable against AllD. However, two positive simple roots would mean that tit-for-tat is stable against AllD at intermediate levels of repetition, but unstable at very low and very high levels of repetition. To investigate the possibility that there may be more than one root, we used the numerical root counting algorithm provided with Wolfram’s Mathematica software (Wolfram, 1991) (based on Sturm’s Theorem), to count the number of simple roots in the interval  $[0, \infty]$ . Figure 2 shows the results of counting roots at 10 000 equally spaced points in the PD rectangle at two different values of error ( $\epsilon = 0.1, \epsilon = 0.01$ ).

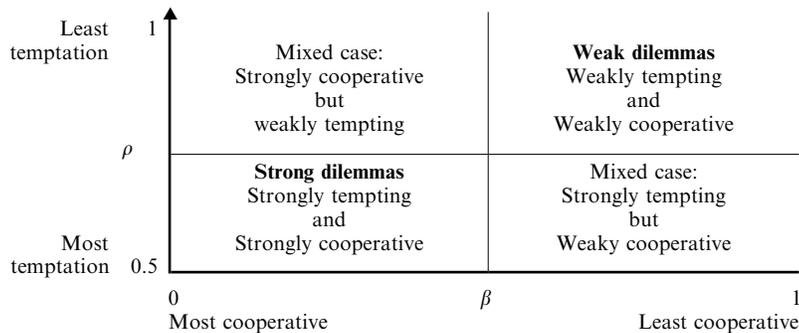


FIG. 1. The rectangular parameter space we use to study Prisoner’s Dilemmas. A heuristic categorization of the parameter space is superimposed.

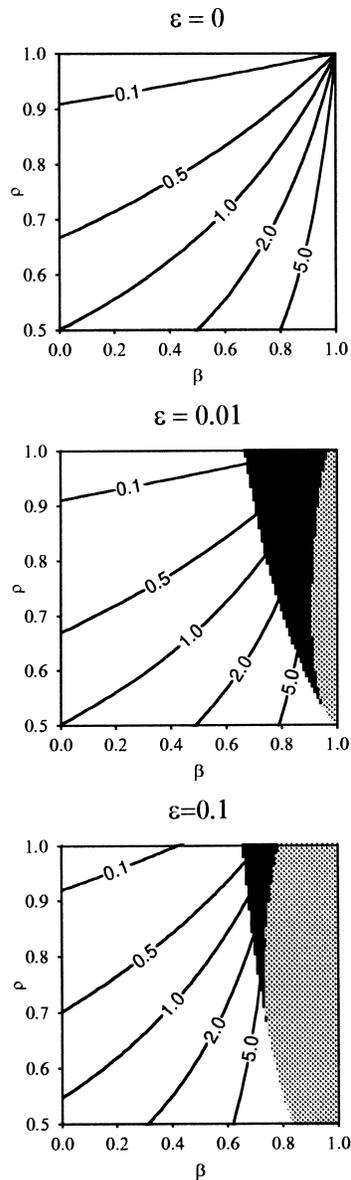


FIG. 2. Effect of error on the stability on tit-for-tat vs. AllD. The figures shows level curves for the critical amount of subjective repetition  $z$  required for tit-for-tat to be stable against AllD. Level curves for  $z$  values of  $\{0.1, 0.5, 1, 2, 5\}$  are show for three levels of error. In addition, regions of the PD rectangle are shaded according to the number of positive real roots of the polynomial  $v$ . There is one root in the unshaded region. In this unshaded region the conventional claim that tit-for-tat is stable against AllD given sufficient repetition applies. However, this claim does not apply to the other two regions. The light gray region shows parameter values where tit-for-tat is never stable against AllD (zero positive real roots). The black region shows parameter values where tit-for-tat is stable against AllD at intermediate amounts of repetition but unstable at both very high and very low amounts of repetition.

These calculations suggest (Fig. 2) that we can divide the PD rectangle into three qualitatively distinct regions. In most of the PD rectangle there is a single root, and hence a single amount of subjective repetition

above which tit-for-tat is stable against AllD. Thus the error-free result of Axelrod and Hamilton holds qualitatively in leftmost two-thirds of the PD rectangle. In contrast, when  $\beta$  is near 1, (the game is only weakly cooperative) there exists a region where no amount of subjective repetition is sufficient to make tit-for-tat stable against AllD. Nowak (1990) derived a similar result using stationary distributions. The third region consists of parameters at which two positive roots occur; at these parameter values tit-for-tat is stable against AllD only at *intermediate* levels of repetition and is unstable against AllD at both the lowest and the highest levels of repetition. Notice that this “two root” region is largest for  $\epsilon$  small. Mathematically, this occurs because there is a “bifurcation” in the behavior of  $v$  at  $\epsilon = 0$ . Intuitively, the existence of this two root region reflects two properties of tit-for-tat: (i) *some* repetition is required for stability against AllD to possible, but (ii) in the long-run tit-for-tat loses much of its intermediate-term advantage by playing “randomly” (the stationary distribution of  $\mathcal{T}_\epsilon$  vs.  $\mathcal{T}_\epsilon$  is  $\{1/4, 1/4, 1/4, 1/4\}$ , see Molander 1985).

Both the zero and two root regions of the PD rectangle represent conditions in which tit-for-tat is not stable against AllD in the *long run*. Considering this zero-root or two-root region as a whole, we note that the long-run stability of tit-for-tat is more strongly influenced by the “cooperativeness” of the game  $\beta$  than by the “temptation” of the game  $\rho$ . Tit-for-tat tends to be unstable against AllD in weakly cooperative ( $\beta$  high) situations.

These root counting calculations reveal the qualitative nature of the solution but say nothing about the amount of repetition required to make tit-for-tat stable against AllD when a single root exists. The Appendix shows that one can readily calculate level curves in the PD rectangle that represent conditions where a given  $z$  value is critical. Figure 2 also shows these level curves. Each curve represents the set of points in the PD rectangle at which a given amount of repetition is critical. These level curves show the following:

- Regardless of the amount of error, games that are strongly cooperative ( $\beta$  low) but only weakly tempting ( $\rho$  high) require the least subjective repetition to stabilize tit-for-tat against AllD. This is the upper left hand corner of the PD rectangle.
- Although the degree of repetition required to stabilize tit-for-tat against AllD, increases with error, the effect is not striking.

- The most widely studied payoff values ( $T=5$ ,  $R=3$ ,  $P=1$ ,  $S=0$ ) are located at  $\beta=1/3$ ,  $\rho=3/5$  in the lower left corner of the PD rectangle. The fact that both bluejays (Clements & Stephens, 1995) and humans (Scodel *et al.*, 1959) defect at these parameter values might be accounted for by “perceived” values of repetition ranging from 1 (with no error) to 1.4 (with 10% error).

Finally, an examination of Fig. 2 suggests that there may be a critical value of  $\beta$  below which the “one root” result holds. The Appendix shows that for small error rates the “one root” result can be expected to hold when  $\beta < 2/3$ .

**Pavlov vs. AllD**

Recently Nowak & Sigmund (1993) have popularized a strategy called Pavlov. The advertised advantages of Pavlov are (i) it replaced tit-for-tat in Nowak and Sigmund’s population simulations (in which fitness consequences of strategic interaction are based on stationary distributions) and (ii) that it superficially resembles Thorndike’s “law of effect” (see Staddon, 1983, for an engaging historical review of the law of effect and related issues). The significance of the name Pavlov seems to be that both Thorndike and Pavlov were animal psychologists.

INITIAL BEHAVIOR OF PAVLOV

To analyze the stability of Pavlov vs. AllD, we must specify the initial behavior of a “Pavlov” strategist. Nowak and Sigmund’s stationary distribution approach does not depend on initial behavior, and they do not appear to have specified any. We will suppose that Pavlov cooperates initially, simply because Pavlov has been advanced as a cooperative strategy. That is, we take Pavlov to be  $\mathcal{P} = \{(0, 1, 1, 0), 1\}$ . We remark that the amount of “repetition” required to stabilize error-free Pavlov against AllD is the same regardless of whether we take  $c=0$  or  $c=1$ .

ERROR-FREE ANALYSIS

Since the Pavlov strategy is not as well known as tit-for-tat, we begin with an elementary game matrix analysis without strategic error that is analogous to the well-known tit-for-tat vs. AllD analysis.

	Pavlov	AllD
Pavlov	$\rho(1+z)$	$\frac{\beta\rho z(1+z)}{1+2z}$
AllD	$\frac{(1+z)(1+z+\beta\rho z)}{1+2z}$	$\beta\rho(1+z)$

The matrix is similar to the tit-for-tat vs. AllD matrix:

the cells on the diagonal are the same, and AllD is a strong Nash equilibrium (hence an ESS) for all feasible parameter combinations. We find that Pavlov can be stable against AllD when

$$z > \frac{1-\rho}{\rho(2-\beta)-1} \tag{3}$$

There exist, however, degrees of “cooperativeness” ( $\rho$ ) and “temptation” ( $\beta$ ) where no amount of repetition can make error-free Pavlov stable against AllD. This is clearly seen in expression (3), because when

$$\rho < \frac{1}{2-\beta} \tag{4}$$

the denominator on the r.h.s. of expression (3) is negative, and the direction of the inequality is reversed. Since  $z$  can never be less than a negative number, expression (4) describes a region in which Pavlov can never be stable against AllD. One can readily confirm that expression (4) is equivalent to the condition  $2R < T + P$  given by Nowak & Sigmund (1993: 58). Our formulation, however, makes it clear that this represents a rather large proportion of the Prisoner’s Dilemma parameter space (Fig. 3).

ADDING ERROR

Closely following our analysis of tit-for-tat, we study the combined effects of “subjective repetition” and error by studying the difference

$$\Delta = \mathbf{v} \cdot \left( \mathbf{I} - \frac{z}{z+1} \mathbf{M}_{\mathcal{P}_c, \mathcal{P}_c} \right)^{-1} \cdot \mathbf{y}_{\mathcal{P}_c, \mathcal{P}_c} - \mathbf{v} \cdot \left( \mathbf{I} - \frac{z}{z+1} \mathbf{M}_{\mathcal{Q}_c, \mathcal{Q}_c} \right)^{-1} \cdot \mathbf{y}_{\mathcal{Q}_c, \mathcal{Q}_c}$$

As in the case of tit-for-tat, we show in the Appendix that one can study the sign of this difference by studying the roots of a polynomial. The appendix presents the calculations required to plot a figure like Fig. 2 for Pavlov (Fig. 3). The most striking feature of Fig. 3 is the large region in which Pavlov cannot be stable against AllD. Calculations in the appendix show that the “no Pavlov” region ranges from 39% of the PD rectangle in the no error case, to 58% of the PD rectangle in the 10% error case.

COMPARISON WITH NOWAK AND SIGMUND’S RESULTS

As mentioned above, Nowak & Sigmund (1993) acknowledge that Pavlov cannot be stable against AllD when condition (4) holds. They diminish this fact,

however, by claiming that when condition (4) is an exact equality, then a *Pavlov-like* strategy can be stable against *AllD*. Of course, this says nothing about the remainder of the large “no Pavlov” region that is not exactly on the boundary defined by condition (4).

Nowak and Sigmund studied  $R$  values ranging from 2.6 to 4.5, but held the other three values at  $T=5$ ,  $P=1$ ,  $S=0$ . In the PD rectangle this implies the relationship

$$\rho = \frac{1}{5\beta},$$

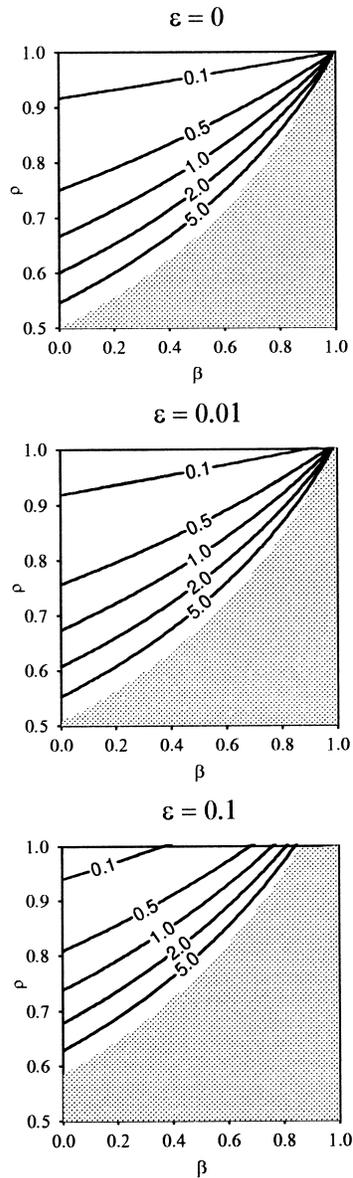


FIG. 3. Effect of error on the stability of Pavlov vs. *AllD*. This figure shows information for Pavlov that is similar to Fig. 2. The gray area shows regions of the PD rectangle where Pavlov cannot be stable against *AllD*. Pavlov can be stable against *AllD* in the white regions. We show level curves for the amount of subjective repetition,  $z$ , required to stabilize Pavlov in the white region.

where  $\beta \in [0.22, 0.38]$ . Figure 4 places Nowak and Sigmund's results in the PD rectangle. Two noteworthy points emerge from Fig. 4. First, despite very significant differences in analytical procedure and assumptions (e.g. we consider only two strategies, Pavlov and *AllD*, while Nowak and Sigmund introduce random “mutants” at random times; we consider only one “Pavlov”, Nowak and Sigmund consider two) there is a remarkable qualitative agreement between our elementary boundary calculation and Nowak and Sigmund's reported levels of cooperation. Second, Nowak and Sigmund's reported results represent a very small part of the PD rectangle, the narrow range of  $\beta$  considered seems especially worrying. One wonders what picture would emerge at high  $\beta$  values, where error weakens tit-for-tat based strategies.

We remark that although the relationship  $\rho = 1/5\beta$  is an arbitrary one, there is one important situation where an orderly relationship between  $\rho$  and  $\beta$  arises from first principles. One appealing way to construct a PD is to imagine that the action  $C$  delivers  $k_c$  units of benefit to the actor (the “keep” amount), and  $g_c$  units of benefit to the opponent (the “give” amount); similarly the action  $D$  delivers amounts  $k_d$  and  $g_d$  to the actor and opponent, respectively. PDs constructed in this way are said to be “decomposable” (Mesterton-Gibbons, 1992; Mesterton-Gibbons & Dugatkin, 1992). Decomposable PDs imply a hyperbolic relationship between  $\rho$  and  $\beta$ :  $\rho = 1/(1 + \beta)$ . This curves lies almost entirely within the white region of Fig. 2; hence we conclude that Axelrod and Hamilton's single root result is more robust for decomposable PDs than for PDs generally. The effect of decomposable PDs on Pavlov is less striking.

GENERAL PATTERN

The qualitative pattern that emerges from our analysis of Pavlov is similar to our tit-for-tat analysis in some ways, but quite different in others. For example, the neighborhood of the upper left corner,  $\rho=1$  and  $\beta=0$  is the region where “cooperative behavior” can be maintained even in the face of relatively strong discounting in both analyses. Putting this another way, the isoclines of critical  $z$  values show a qualitatively similar pattern in both analyses. The regions of “no stability” are, however, quite different. Although tit-for-tat is least likely to be stable at high  $\beta$  values (weakly cooperative situations), Pavlov's stability is affected roughly equally by both  $\beta$  and  $\rho$ : the stability of Pavlov requires a relatively “cooperative

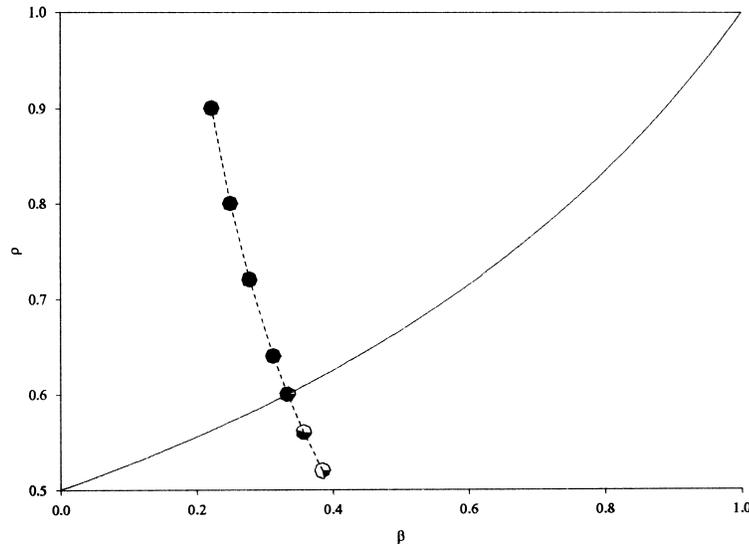


FIG. 4. Nowak and Sigmund's simulation results superimposed on the PD rectangle and the "no Pavlov" boundary provided by condition (4). The solid curve is the "no Pavlov" boundary, below this we predict that Pavlov can never be stable against ALLD. The "pies" connected by a dashed line represent the "percent cooperation" (all black means 100%) in Nowak and Sigmund's simulations each is placed at its appropriate position in the PD rectangle.

situation" ( $\beta$  low), and a relatively weak temptation to defect ( $\rho$  high).

**Estimated Discounting Rates**

In examining Figs 2 and 3, a single question comes to the forefront: What is the plausible range for the "subjective repetition" parameter  $z$ ? If we accept the widely held view that  $z$  is determined solely by the real

repetition of the game, then a "handful" of repetitions (0.1 to 5 as shown in Figs 2 and 3) seems like a modest requirement.

It is much less clear, however, that the requirements for cooperation are so modest if we adopt the value degradation view of  $z$ . Consider the following simple experiment conducted in our research group. We presented feeding bluejays with the choice (Fig. 5) of a large reward delayed for 20s, or a small reward

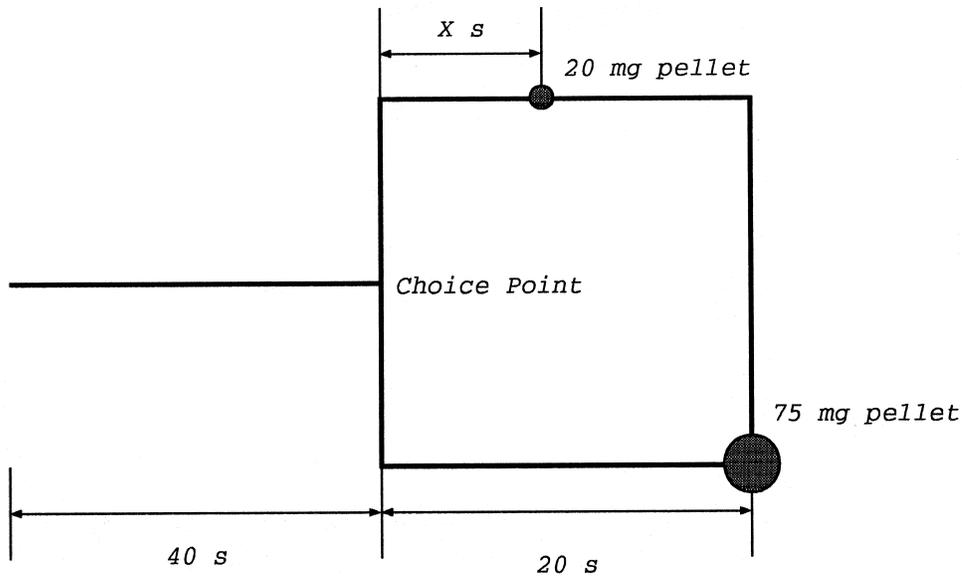


FIG. 5. Diagram of a single experimental trial. After a 40 s inter-trial interval, subject must choose between (a) a key color leading to a 75 mg food pellet after a 20 s delay, and (b) a distinct key color leading to a 20 mg pellet after an  $X$  s delay. We increased  $X$  by 0.25 s if the bluejay chose the 20 mg pellet, and decreased  $X$  by 0.25 s if the bluejay chose the 75 mg pellet.

TABLE 1

Estimated Indifference points and discounting parameters

Bird	$T_{50} \pm 95\% \text{ CL}$	Estimated $k$	Estimated $\lambda$
7	$1.28 \pm 0.17 \text{ s}$	0.181/s	0.071/s
31	$5.15 \pm 0.23 \text{ s}$	4.00/s	0.089/s

delayed for a  $Xs$ . If the subject chooses the small reward on a given trial we increment  $X$  by 0.25 s, if the subject chooses the large reward we decrement  $X$  by the same amount. To ensure that subjects have information about both options, we randomly interspersed "forced choice" trials (in which only one option is available, and no adjustments to  $X$  are made) with the free choice trials. This procedure, following earlier work by Mazur (1987), allows us to estimate an indifference point, which we can use, in turn, to estimate a discounting rate.

The two simplest models of value degradation that students of animal discounting have advocated are:

$$V(t) = V(0)e^{-\lambda t} \tag{5}$$

$$V(t) = \frac{V(0)}{1 + kt}, \tag{6}$$

where  $V(0)$  represents undelayed value and  $V(t)$  represented the perceived value given a delay of  $t$ . The first model, of course, corresponds to the simplest continuous time degradation model:

$$\frac{dV}{dt} = -\lambda V$$

while the second, or hyperbolic, model is the simplest model that admits a decreasing instantaneous rate of degradation (see Kagel *et al.*, 1986 for a discussion of non-constant discounting rates). Table 1 shows indifference point estimates for two of our bluejays, together with corresponding estimates of the  $k$  and  $\lambda$  parameters. For comparison, estimates for Mazur's (1987) four pigeons, yield  $k$  values ranging from 0.17/s to 5.6/s, and  $\lambda$  estimates ranging from 0.05/s to 0.2/s. In Clements & Stephens's (1995) experimental games, the inter-play interval varied from 200 to 400 s, and it seems reasonable to suppose that these are rather short intervals compared with natural games played by

TABLE 2  
Estimates of  $\alpha$

	Hyperbolic model		Exponential model	
	$k = 0.15/s$	$k = 6/s$	$\lambda = 0.05/s$	$\lambda = 0.2/s$
$t = 200s$	0.032	$8.33 \times 10^{-4}$	$4.54 \times 10^{-5}$	$4.25 \times 10^{-17}$
$t = 400s$	0.016	$4.16 \times 10^{-4}$	$2.06 \times 10^{-9}$	$1.80 \times 10^{-35}$

long-lived vertebrates. Table 2 presents estimates of the  $\alpha$  parameters based on these estimates; note that for these magnitudes  $z = \alpha/(1 - \alpha) \approx \alpha$ .

Several cautionary notes are in order here. First, these estimates are derived from what an operant psychologist would call "within trial" behavior, rather than the "between trial," play-to-play discounting that is important in repeated games. Second, the fact that discounting rates are probably not constant (Kagel *et al.*, 1986) suggests that the values in Table 2 may only be valid estimates of the degradation from the first to the second play. Third, it is possible that animals "discount" differently in social and non-social contexts, so that an  $\alpha$  measured in a non-social context is irrelevant to a "cooperative" game. Many students of animal discounting, however, have argued that animals will discount *more* heavily in social contexts (Kagel *et al.*, 1986; Barkan & Withiam, 1989). Fourth, operant psychologists have studied a limited range of species in limited circumstances; for example, no one has measured the discounting behavior of fish, and it is at least possible that they discount differently. Finally, all the extant non-human discounting data are based on delayed food reward, and animals may discount differently when offered different types of rewards, e.g. for mates or predator avoidance. Despite these difficulties, these data, and related data from the so-called "self-control" literature (Logue, 1988), represent the only estimates of animal discounting behavior of which we are aware. They are certainly better than guessing.

Estimated  $\alpha$  values derived from the value-degradation interpretation (roughly  $10^{-30}$  to 0.03, Table 2) are much smaller than the values that seem plausible under the game repetition view (roughly 0.1 to 0.9). Under the value-degradation view, we would expect cooperation in PDs to be restricted to a small region of the PD rectangle near the upper left corner ( $\beta = 0$ ,  $\rho = 1$ ). That is, we would expect to find animals cooperating only in situations where there is a relatively small temptation to defect ( $R \approx T$ ), and relatively large difference between the benefits of mutual cooperation and mutual defection ( $R \gg P$ ). Clearly, this interpretation is consistent with the extant evidence derived from controlled PDs which have focused on the lower left corner of the PD rectangle and especially the values  $\rho = 3/5$  and  $\beta = 1/3$  that characterize Axelrod's payoffs.

We take strong preferences for immediacy as a demonstrable fact of life, without commenting on the underlying cause of these strong preferences. There are two possibilities. First, these preferences for immediacy may confer some evolutionary advantage that, as yet, we do not fully understand. Behavioral ecologists

(Kagel *et al.*, 1986; Stephens & Krebs, 1986) have focused on the view that preference for immediacy represents an adaptation to stochastically determined “interruptions”, but there are other possibilities (e.g. rapid population growth can produce a situation analogous to monetary inflation). A diametrically opposed view is that these preferences reflect some “psychological mechanism”, that has evolved in some other context and has no particular evolutionary significance. The question that arises here is whether the tendency for animals to defect in PDs is adaptive (i.e. part of larger adaptive syndrome favoring immediate reward) or maladaptive (preferences for immediacy can't be “turned off” in experimental contexts).

## Discussion

### LIMITATIONS

Our models incorporate several assumptions that may limit their generality. We suppose, for example, that error affects both players in a repeated game equally, and that error affects each component of a given strategy (i.e. the probabilities  $t$ ,  $r$ ,  $p$ ,  $s$  and  $c$ ) in the same way. We know of no empirical justification for asymmetric error rates, but if they exist, they might change the quantitative details of our results. For example, Nowak and Sigmund's “second” Pavlov might be interpreted as a Pavlov strategist in which error effects the  $p$  component of the strategy differently than the other components. Although we have only considered a simple model of error, work by Kreps *et al.* (1982) suggests that erroneous implementations of whole strategies (e.g. an AllD strategist mistakenly playing tit-for-tat) can stabilize cooperative strategies. A second limitation is our assumption of constant discounting rates. Although this assumption follows game theoretical convention, there is compelling experimental evidence of decreasing discounting rates, and a study of the effects of this phenomenon may prove interesting. We believe, however, that this phenomenon is unlikely to change our results qualitatively, given that rewards lose so much of their value between the first and the second play. Finally, we have restricted our attention to the class of strategies that can be described by first-order Markov processes, and this may limit the generality of our approach.

### OUR PREMISE

The value of our approach hinges upon the validity of our basic premise which is that animals typically defect in IPDs. For example, this justifies our analyses of limited two strategy games in which a hypothetical

cooperative strategy is pitted against all-defection. The difficulty is that controlling payoffs (and even precisely measuring payoffs) requires experimental “artificialities,” and these artificialities may create a situation that animals simply do not recognize as “cooperative.” This problem is not unique to the Prisoner's Dilemma or even to biology. Kamil (1988) points out, for example, that almost anything an experimenter does to increase the *internal validity* of a test situation (e.g. controls for experience, order, appropriateness of model, etc.) compromises the *external validity* of the test (e.g. the extent to which the results can be generalized to other situations).

However, for non-human animals similar results from more naturalistic systems reduce the severity of this problem. For example, Packer & Rutan (1988) analyzed existing data for 26 species that hunt cooperatively, and concluded that cooperative hunters defect when there is a temptation to do so. Similarly, bowerbirds appear to defect in a naturally occurring PD (Pruett-Jones & Pruett-Jones, 1994), and ravens (Heinrich, 1988) recruit others to carrion only when recruits are needed to overcome the defenses of territory owners. Recent experimental evidence (Godin & Davis, 1995) suggests that predator approach behavior in shoaling fishes—a widely studied behavior that was previously thought to be a PD (Milinski, 1987; Dugatkin, 1988)—may not be a PD. In parallel with this, Grinnell *et al.* (1995) recently concluded that the superficially PD-like situation of group territory defense by African lions is not a PD. In light of this evidence, the most parsimonious view is that most non-kin cooperation is simple mutual selfishness (sometimes called mutualistic cooperation, or by-product mutualism: see Mesterton-Gibbons & Dugatkin, 1992). This does not mean that altruistic cooperation never occurs, only that it is rare, and when it does occur (Wilkinson, 1984) the relevance of the IPD is unclear.

Evidence for humans is less clear-cut. Colman (1982) reviewed the human experimental literature, and reports that for experimental IPDs “. . . the most striking general finding is undoubtedly the *DD lock-in effect*,” (emphasis Colman's). This suggests, in agreement with the non-human experiments that long strings of mutual defection dominate the human literature. However, Colman also reports an apparently contradictory trend in a body of experimental work that follows techniques pioneered by Rapoport & Chammah (1965) in which the relative frequency of the  $C$  choice is about 0.5 at the beginning of sequence of plays, drops for a 20–50 plays and then rebounds to asymptote near 0.6. There are several procedural details that differ between the Rapoport and

Chammah style of experiment, and experimental studies of non-human animals. First, the  $S$  and  $P$  values in these studies are negative (losses rather gains). Hence there must be some cooperation if the subjects are to gain anything; moreover, psychophysical evidence suggests that humans treat losses and gains differently; for example, they are willing to sacrifice more to avoid a loss, than to achieve an equivalent gain (Tversky & Kahneman, 1991). Second, the number of trials, although large by standards of human experimentation is small by non-human standards. Rapoport and Chammah's undergraduates played 300 times in about an hour; while Clements & Stephens's (1995) bluejays played 200 times per day for many weeks. Finally, Rapoport and Chammah's subjects were not rewarded for each play of the game. Instead, points were added or deducted from a running tally, and the subjects were paid (\$0.001 per point!) at the end of the sequence of 300 plays. This procedure encourages a relatively weak discounting of "points".

Our results suggest further empirical exploration of the PD rectangle, and especially the neighborhood of weak temptation ( $\rho \approx 1$ ) and high cooperativeness ( $\beta \approx 0$ ). We are currently conducting explorations of this type. However, we feel that the most compelling conceptual problem in the study of non-kin cooperation is the development of alternative models of altruistic cooperation. Advocates of the Prisoner's Dilemma urgently need to articulate the empirical basis for continued interest in this research program.

We are grateful to the National Science Foundation (IBN-8958228), Lincoln Telephone Company, Cliffs Charitable Foundation and Dr and Ms William Ludwick for providing financial support for this work.

## REFERENCES

- AXELROD, R. (1984). *The Evolution of Cooperation*. New York: Basic Books.
- AXELROD, R. & HAMILTON, W. (1981). The evolution of cooperation. *Science* **211**, 1390–1396.
- BARKAN, C. P. L. & WITHIAM, W. L. (1989). Profitability, rate maximization, and reward delay: a test of the simultaneous-encounter model of prey choice with *parus atricapillus* parus. *Am. Nat.* **134**, 254–272.
- BOYD, R. (1989). Mistakes allow evolutionary stability in the repeated prisoner's dilemma game. *J. theor. Biol.* **136**, 47–56.
- BOYD, R. & LORBERBAUM, J. P. (1987). No pure strategy is evolutionarily stable in the repeated prisoner's dilemma game. *Nature* **327**, 58–59.
- BURDEN, R. L. & FAIRES, J. D. (1985). *Numerical Analysis*, 3rd edn. Boston: Prindle Weber & Schmidt.
- CLEMENTS, K. C. & STEPHENS, D. W. (1995). Testing models of non-kin cooperation: mutualism and the prisoner's dilemma. *Anim. Behav.* **50**, 527–549.
- COLMAN, A. M. (1982) *Game Theory and Experimental Games: The Study of Strategic Interaction*. Oxford: Pergamon Press.
- DUGATKIN, L. (1988). Do guppies play tit for tat during predator inspection visits? *Behav. Ecol. Sociobiol.* **23**, 395–399.
- FELLER, W. (1950). *An Introduction to Probability Theory and its Applications*. New York: John Wiley and Sons.
- FLOOD, M., LENDENMANN, K. & RAPPOPORT, A. (1983)  $2 \times 2$  games played by rats: different delays of reinforcement as payoffs. *Behav. Sci.* **28**, 65–78.
- GARDNER, R. M., CORBIN, T. L., BELTRAMO, J. S. & NICKELL, G. S. (1984). The prisoner's dilemma game and cooperation in the rat. *Psychol. Rep.* **55**, 687–696.
- GODIN, J.-G. J. & DAVIS, S. A. (1995). Who dares, benefits: predator approach behaviour in the guppy (*Poecilia reticulata*) deters predator pursuit. *Proc. R. Soc. Lond. B.* **259**, 193–200.
- GRINNELL, J., PACKER, C. & PUSEY, A. E. (1995). Cooperation in male lions: kinship, reciprocity or mutualism? *Anim. Behav.* **49**, 95–105.
- HEINRICH, B. (1988). Winter foraging at carcasses by three sympatric corvids, with emphasis on recruitment by the raven, *corvus corax*. *Behav. Ecol. Sociobiol.* **23**, 141–156.
- KAGEL, J. H., GREEN, L. & CARACO, T. (1986). When foragers discount the future: constraint or adaptation? *Anim. Behav.* **34**, 271–283.
- KAMIL, A. C. (1988). Experimental design in ornithology. In: *Current Ornithology* (Johnson, R. F., ed.), pp. 313–346. New York: Plenum Press.
- KREPS, D., MILGROM, P., ROBERTS, J. & R., W. (1982). Rational cooperation in the finitely repeated prisoner's dilemma. *J. Econ. Theor.* **27**, 245–252.
- LOGUE, A. W. (1988). Research on self-control: an integrating framework. *Behav. Brain. Sci.* **11**, 665–709.
- LOMBARDO, M. (1985). Mutual restraint in tree swallows: a test of the tit for tat model of reciprocity. *Science* **227**, 1363–1365.
- MAZUR, J. E. (1987). An adjusting procedure for studying delayed reinforcement. In: *Quantitative Analyses of Behavior, Volume 5, The Effect of Delay and of Intervening Events on Reinforcement value* (Commons, M. L., Mazur, J. E., Nevin, J. A. & Rachlin, H., eds), pp. 55–73. New York: Erlbaum.
- MESTERTON-GIBBONS, M. (1992). *An Introduction to Game-theoretic Modelling*. Redwood City, CA: Addison-Wesley.
- MESTERTON-GIBBONS, M. & DUGATKIN, L. A. (1992). Cooperation among unrelated individuals: evolutionary factors. *Q. Rev. Biol.* **67**, 267–281.
- MILINSKI, M. (1987). Tit for tat and the evolution of cooperation in sticklebacks. *Nature* **325**, 433–437.
- MOLANDER, P. (1985). The optimal level of generosity in a selfish, uncertain environment. *J. Conflict Resol.* **29**, 611–618.
- NOWAK, M. (1990). Stochastic strategies in the prisoner's dilemma. *Theoret. Popul. Biol.* **38**, 93–112.
- NOWAK, M. A. & SIGMUND, K. (1992). Tit for tat in heterogeneous populations. *Nature* **355**, 250–252.
- NOWAK, M. & SIGMUND, K. (1993). A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma. *Nature* **364**, 56–58.
- PACKER, C. & RUTAN, L. (1988). The evolution of cooperative hunting. *Am. Nat.* **132**, 159–198.
- PRUETT-JONES, S. & PRUETT-JONES, M. (1994). Sexual competition and courtship disruptions: why do male bowerbirds destroy each other's bowers? *Anim. Behav.* **47**, 607–620.
- RAIFFA, H. (1968). *Decision Analysis*. Reading, MA: Addison-Wesley.
- RAPOPORT, A. (ed.) (1974). *Game Theory as a Theory of Conflict Resolution*. Dordrecht: D. Reidel Publishing Company.
- RAPOPORT, A. & CHAMMAH, A. M. (1965). *Prisoner's Dilemma a Study in Conflict and Cooperation*. Ann Arbor, MI: University of Michigan Press.
- SCODEL, A. J., MINAS, S., RATOOSH, P. & LIPETZ, M. (1959). Some descriptive aspects of two-person non-zero-sum games. *J. Conflict. Resol.* **3**, 114–119.
- STADDON, J. E. R. (1983). *Adaptive Behavior and Learning*. New York: Cambridge University Press.

STEPHENS, D. W. & KREBS, J. R. (1986). Foraging theory. *Monographs in Behavior and Ecology*. Princeton, NJ: Princeton University Press.  
 TVERSKY, A. & KAHNEMAN, D. (1991). Loss aversion in riskless choices: a reference dependent model. *Q. J. Econ.* **106**, 1039–1061.  
 WILKINSON, G. S. (1984). Reciprocal food sharing in the vampire bat. *Nature* **308**, 181–184.  
 WOLFRAM, S. (1991). *Mathematica: A System for Doing Mathematics by Computer*, 2nd edn. Redwood City, CA: Addison-Wesley.

APPENDIX

A.1. Tit-for-tat vs. AllD

NUMBER OF ROOTS

We are interested in the roots of

$$\Delta = \mathbf{v} \cdot \left( \mathbf{I} - \frac{z}{z+1} \mathbf{M}_{\mathcal{T}, \mathcal{T}} \right)^{-1} \cdot \mathbf{y}_{\mathcal{T}, \mathcal{T}} - \mathbf{v} \cdot \left( \mathbf{I} - \frac{z}{z+1} \mathbf{M}_{\mathcal{D}, \mathcal{T}} \right)^{-1} \cdot \mathbf{y}_{\mathcal{D}, \mathcal{T}}$$

which is eqn (2) of the main text, where  $\mathbf{v} = (1, \rho, \beta\rho, 0)$ . Performing the indicated matrix operations, we find that  $\Delta$  can be expressed as the rational function:

$$\Delta = \frac{A + Bz + Cz^2 + Dz^3}{(1 + 2cz)(-1 - 4cz + 4c^2z)}$$

$$A = 1 - 3\epsilon + 2\epsilon^2 - \rho + 3\epsilon\rho + \beta\epsilon\rho - 2\epsilon^2\rho - 2\beta\epsilon^2\rho$$

$$B = (-1 + 2\epsilon)(-6\epsilon + 8\epsilon^2 - 2\epsilon^3 + \rho - \beta\rho + 4\epsilon\rho + \beta\epsilon\rho - 8\epsilon^2\rho - 4\beta\epsilon^2\rho + 2\epsilon^3\rho + 2\beta\epsilon^3\rho)$$

$$C = \epsilon(-1 + 2\epsilon)(1 - 15\epsilon + 22\epsilon^2 - 8\epsilon^3 + 4\rho - 6\beta\rho + \epsilon\rho + 15\beta\epsilon\rho - 14\epsilon^2\rho - 18\beta\epsilon^2\rho + 8\epsilon^3\rho + 8\beta\epsilon^3\rho)$$

$$D = 2\epsilon^2(1 - 3\epsilon + 2\epsilon^2)(-1 + 6\epsilon - 4\epsilon^2 - \rho + 3\beta\rho - 2\epsilon\rho - 6\beta\epsilon\rho + 4\epsilon^2\rho + 4\beta\epsilon^2\rho)$$

(A.1)

First, considering the denominator of  $\Delta$ , note that it has two negative real roots, and so we know that it will remain negative for all nonnegative values  $z$ . We lose nothing, qualitatively, by considering only the negative of the numerator, say  $v$ . We note that when there is no subjective repetition ( $z=0$ ),  $v$  equals

$$-(1 - 2\epsilon)((1 - \epsilon)(1 - \rho) + \beta\rho\epsilon)$$

which is negative for all biologically plausible parameter values. As expected then, AllD can always invade tit-for-tat when there is no subjective repetition.

Further, when there is no error,  $v$  is simply linear in  $z$  and there is a single unique amount of subjective repetition

$$z = \frac{1 - \rho}{\rho(1 - \beta)}$$

above which tit-for-tat is stable against AllD. (The reader can confirm that this is equivalent to Axelrod and Hamilton’s condition  $w > (T - R)/(T - P)$ .) However, when there is error  $\epsilon \neq 0$ ,  $v$  is a cubic polynomial and the positive roots of this polynomial represent changes of sign of the difference  $\Delta$ , which in turn represent changes in the stability of tit-for-tat against AllD. We used the numerical root counting algorithm provided with Wolfram’s Mathematica software (Wolfram, 1991) (based on Sturm’s Theorem), to count the number of simple roots in the interval  $[0, \infty)$  (Fig. 2).

An asymptotic argument confirms our numerical results. Recalling that  $v$  can be written in the form

$$v = a + bz + cz^2 + dz^3$$

where  $a, b, c$  and  $d$  are the negatives of the  $A, B, C$  and  $D$  terms defined in eqn A.1. We note that for  $\epsilon$  small, the signs of  $a, b, c$  and  $d$  are the signs of  $\rho - 1, \rho(1 - \beta), 1 + (1 - 3\beta)\rho + 3(1 - \beta)\rho,$  and  $1 + (1 - 3\beta)\rho$ . So  $a < 0, b > 0$  and if  $d > 0$  then  $c > 0$ . Therefore, if  $d > 0$  then  $\partial v / \partial z > 0$  for all  $z > 0$ , and there can be only one positive root. This argument suggests that a single root is expected for  $(\beta, \rho)$  such that  $\beta \leq 1/3$ , or  $\beta > 1/3$  and  $\rho < 1/(3\beta - 1)$  in close agreement with the middle panel of Fig. 2.

A.2. Isoclines of Subjective Repetition

We can readily calculate level curves in the PD rectangle that represent conditions where a given  $z$  value is critical. This is straightforward, because despite its complexity  $v$  is a linear function of  $\rho$  and  $\beta$ , so we can readily solve for  $\rho$  as a function of  $\beta, \epsilon$ , and  $z$ . We find:

$$\rho_z = \frac{\mathcal{N}}{\mathcal{D}}$$

$$\mathcal{N} = (1 - \epsilon)(1 + (6\epsilon - 2\epsilon^2)z + (-\epsilon + 14\epsilon^2 - 8\epsilon^3)z^2 + (-2\epsilon^2 + 12\epsilon^3 - 8\epsilon^4)z^3)$$

$$\mathcal{D} = 1 - \epsilon - \beta\epsilon + (1 - \beta + 4\epsilon + \beta\epsilon - 8\epsilon^2 - 4\beta\epsilon^2 + 2\epsilon^3 + 2\beta\epsilon^3)z + (4\epsilon - 6\beta\epsilon + \epsilon^2 + 15\beta\epsilon^2 - 14\epsilon^3 - 18\beta\epsilon^3 + 8\epsilon^4 + 8\beta\epsilon^4)z^2$$

$$\begin{aligned}
 &+(2\epsilon^2 - 6\beta\epsilon^2 + 2\epsilon^3 + 18\beta\epsilon^3 - 12\epsilon^4 \\
 &- 20\beta\epsilon^4 + 8\epsilon^5 + 8\beta\epsilon^5)z^3 \tag{A.2}
 \end{aligned}$$

We use this expression to plot level curves representing points in the PD rectangle at which a given amount of repetition is critical.

BOUNDARY OF THE "ONE ROOT" REGION

We remark that the right-most boundary of the "one root" region can be found by taking the limit of eqn (A.2) as  $z$  approaches infinity yielding:

$$\rho_\infty = \frac{(-1 + \epsilon)(2\epsilon^2 - 12\epsilon^3 + 8\epsilon^4)}{2\epsilon^2 - 6\beta\epsilon^2 + 2\epsilon^3 + 18\beta\epsilon^3 - 12\epsilon^4 - 20\beta\epsilon^4 + 8\epsilon^5 + 8\beta\epsilon^5}$$

From Fig. 2 we note that there is a critical value  $\beta$  below which the "one root-sufficient repetition" result always applies, and the figure suggests that this critical  $\beta$  value occurs at  $\rho_\infty = 1$ . Solving for the  $\beta$  value that satisfies  $\rho_\infty = 1$  yields

$$\frac{2(1 - 2\epsilon)}{3 - 6\epsilon + 4\epsilon^2} \approx \frac{2}{3}$$

$$\rho_z = \frac{(\hat{z} + 1)(1 - \epsilon)}{(\hat{z} + 1)((4 - 4\beta)\epsilon^2 + (3\beta - 5)\epsilon - \beta + 2) - \epsilon(4\beta - 4)(1 - \epsilon) - (1 - \beta)} \tag{A.3}$$

which is very close to  $2/3$  for small error rates. As a crude generalization, we conclude that the sufficient repetition result of Axelrod and Hamilton holds for  $\beta < 2/3$ .

**A.3. Pavlov vs. AllD**

We are interested in the roots of

$$\begin{aligned}
 \Delta &= \mathbf{v} \cdot (\mathbf{I} - \alpha \mathbf{M}_{\mathcal{P}_\epsilon, \mathcal{P}_\epsilon})^{-1} \cdot \mathbf{y}_{\mathcal{P}_\epsilon, \mathcal{P}_\epsilon} \\
 &\quad - \mathbf{v} \cdot (\mathbf{I} - \alpha \mathbf{M}_{\mathcal{Q}_\epsilon, \mathcal{P}_\epsilon})^{-1} \cdot \mathbf{y}_{\mathcal{Q}_\epsilon, \mathcal{P}_\epsilon}
 \end{aligned}$$

Performing the indicated matrix operations, we find that this can be written as

$$\Delta = \frac{-(1 - 2\epsilon)AB}{1 + 2((1 - 2\epsilon)(1 - \epsilon) + \epsilon)z}$$

$$A = 1 + ((1 - 2\epsilon)(1 - \epsilon) + \epsilon)z$$

$$\begin{aligned}
 B &= 1 - \rho - \epsilon(1 - \rho - \beta\rho) + (1 - 2\rho + \beta\rho \\
 &\quad + \epsilon(-1 + 5\rho - 3\beta\rho) + \epsilon^2(-4\rho + 4\beta\rho))z
 \end{aligned}$$

Considering first the denominator of  $\Delta$ , we observe that it has a single negative root, and hence that its sign does not change in the feasible range of  $z \in [0, \infty]$ . We can, therefore, focus our attention on the numerator. The situation is much simpler than in tit-for-tat, because the numerator can be factored into two linear functions of  $z$ , represented as  $A$  and  $B$  in the expression above. Solving  $A = 0$  leads to a negative  $z$  value, so this solution is not biologically feasible. We find the only biologically feasible solution by solving  $B = 0$  for  $z$ , yielding,

$$\hat{z} = \frac{(1 - \rho)(1 - \epsilon) + \epsilon\rho\beta}{4\rho\epsilon^2(1 - \beta) - \epsilon(5\rho - 3\beta\rho - 1) + 2\rho - \beta\rho - 1}$$

the "critical"  $z$  value. Using this result it is straightforward to find level curves for  $\hat{z}$  in the PD rectangle:

We find the boundary of the region where Pavlov can be stable against AllD by taking the limit as  $z$  approaches infinity.

$$\rho_\infty = \lim_{z \rightarrow \infty} \rho_z = \frac{1 - \epsilon}{(1 - \beta)4\epsilon^2 + (3\beta - 5)\epsilon - \beta + 2} \tag{A.4}$$

This expression simplifies to expression (4) in the main text when  $\epsilon = 0$  as required. The last two expressions (A.3) and (A.4) are the tools we need to plot a figure like Fig. 2 for Pavlov (Fig. 3).

We calculate the relative size of the region in which AllD always invades Pavlov by letting  $\beta_0$  be the point where  $\rho_\infty$  [defined by eqn (A.4)] equals 1. Then the proportion of the PD rectangle in which Pavlov cannot be stable against AllD is given by

$$2 \left( \int_0^{\beta_0} \rho_\infty(\beta, \epsilon) d\beta - \beta_0 + \frac{1}{2} \right)$$

over the biologically feasible range of error rates  $0 < \epsilon < 0.1$ , this expression is well approximated by  $0.386 + 1.77\epsilon$ .