J. theor. Biol. (1997) **188,** 1–10

Iterated Prisoner's Dilemma: Pay-off Variance

K. NISHIMURA† AND D. W. STEPHENS

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

(*Received on* 14 *February*, 1996, *Accepted in revised form on* 6 *March*, 1997)

The iterated Prisoner's Dilemma (IPD) is usually analysed by evaluating arithmetic mean pay-offs in an ESS analysis. We consider several points that the standard argument does not address. Finite population size and finite numbers of matches in the IPD game lead us to consider both pay-off variance and the sampling process in the evolutionary game. We provide a general form for the pay-off variance of a Markov strategist in the IPD game, and present a general analysis of the initial invasion process of an "all defection strategist" (ALLD) into a "tit-for-tat" (TFT) strategist population by considering stochastic processes. Finite population size, strategic error and the variances of pay-offs alter the prediction concerning the initial invasion of ALLD compared with the standard Evolutionarily Stable Strategy (ESS) analysis. Even though TFT gets the larger arithmetic mean, the variance of its pay-off is also larger when the expected iterations of the game are sufficiently large. Therefore, the boundary of the parameter of the probability of game continuity, w, above which ALLD does not have advantage to invade into the TFT population, becomes a bit larger than predicted by the deterministic model.

Introduction

The Prisoner's Dilemma attracts attention as a metaphor for the problems of non-kin cooperation. The Prisoner's Dilemma (PD) is a two-player game: each player can opt for one of the two strategies C (to cooperate) and D (to defect). If both players cooperate, their pay-off R is higher than the pay-off P for joint defection. But a player defecting unilaterally obtains a pay-off T, which is larger than R, while the opponent ends up with a pay-off S, which is smaller than P. In addition to this rank ordering, we assume that $2R > S + T$.

The rational decision in this game is to play D, since this yields the higher pay-off no matter whether the opponent uses C or D. As a result, both players defect and earn P instead of the larger reward R for joint cooperation. However, if the same players interact repeatedly with uncertain termination, there need no longer be a single best strategy in what is called the Iterated Prisoner's Dilemma (IPD) (Boyd & Lorberbaum, 1987; Farrell & Ware, 1989; Nowak & Sigmund, 1992, 1993).

The IPD is often played to compare elements of an arbitrary, finite set of strategists. Axelrod (1984) conducted a tournament in which a variety of strategies submitted by evolutionary biologists, computer scientists, and political scientists were pitted against each other in a round-robin competition. The tit-for-tat (TFT) strategy (start with a C, and then use the co-player's previous move) was a conspicuous success. Axelrod & Hamilton (1981) conceptualized the possible advantage for TFT in evolutionary competition by three properties: robustness, stability and initial viability. Among the three, a criterion of stability of a TFT population against invasion by other strategists, inequality (1) in Axelrod & Hamilton (1981, p. 1393), has been widely accepted. However, recent studies (Boyd & Lorberbaum, 1987;

† Present address (corresponding author): Faculty of Fisheries, Hokkaido University, Hakodate 041, Japan.
E-mail: strixkn@fish.fish.hokudai.ac.jp.

0022–5193/97/170001 + 10 $25.00/0/jt970439

Farrell & Ware, 1989) suggest there is no single evolutionarily stable pure strategy in an IPD, when mixtures of pure strategies invade.

However, Boyd (1989) presented that if individuals make mistakes (e.g. C for D and D for C), there exists evolutionarily stable strategy against any mix of invading pure strategies. Other than the rigorous ESS analysis of strategies in IPD, reciprocity is regarded as the underlying mechanism or cooperation. TFT still survive as a simple metaphorical example of reciprocal strategy (Dugatkin, 1991, 1994; Mesterton-Gibbon & Dugatkin, 1992).

Most evolutionary analyses of the IPD game have evaluated arithmetic mean payoffs of competing strategists (Feldman & Thomas, 1987; Boyd, 1989; Mesterton-Gibbon & Dugatkin, 1992; Axelrod & Hamilton, 1981). For example, the widely accepted ESS conditions for stability of TFT against all defection strategists (ALLD) (or) against a player that alternates D and C are based on comparison of the arithmetic means (Axelrod & Hamilton, 1981, page 1393). That is, the argument evaluates the threshold value of w (the probability of repeating play between the same two individuals) implying stability of TFT with respect to a pay-off set (T, R, P, S), on the basis of arithmetic mean pay-offs (Axelrod & Hamilton, 1981; Mesterton-Gibbon & Dugatkin, 1992).

However, this assumption may provide an inadequate general simplification. When an individual experiences infinitely many IPD games with players of a same type of strategy in one generation, it requires infinite population size and infinitely long longevity. When population size is finite and lifetime is short, the possibility that all players meet a new partner in every matching is quite a few, which is contrary to the usual assumptions.

Thus, we may choose to consider that the number of games played by each individual is finite and few, if we assume that the population is not too large; the most extreme case assumes that population size as finite and each individual plays only one IPD game during their lifetime. In this case several stochastic factors will influence the fitness of individual players. In a small population, mutant invasion will occur at most by only a single strategist and a no mixing effect of several strategists might exist. In turn drift may influence the frequency of strategists in the next generation. A small population itself is a cause of drift, irrespective of pay-offs to different strategists (Kimura, 1983). However, in this paper we focus on the stochasticity of pay-offs in an IPD game. When only one IPD game is played over the lifespan of players, because of small population size, stochasticity of pay-off may influence the evolutionary dynamics of strategies.

Stochastic effect in IPD plays an important role in stability of strategies and it has been studied analytically and numerically in a restricted sense (Boyd, 1989; Nowak, 1990; Nowak & Sigmund, 1992, 1993; Stephens *et al*., 1995). These studies treated the expected payoff of ergodic Markovian strategists in order to analyze the stability or the dominance under the assumption of an infinite large population.

We can identify the following factors that affect stochasticity of payoffs in the IPD game. Stochasticity of game repetition should be important. Although repetitions of the game can be assumed an arbitrary integer-valued random variable, we adopt a geometric probability distribution to express stochasticity of game repetition. This is the traditional simplification (Axelrod & Hamilton, 1981; Mesterton-Gibbon, 1992). When repetition of a game depends on a geometric probability distribution with a parameter w, a constant probability of continuity, the expected number of iterations if $1/(1-w)$ and the variance of game length is $w/(1-w)^2$. Thus, a longer expected game length implies a larger variance.

Variance of pay-off in the IPD is also produced by another source. If both players are Markovian strategists (see the next section), the pay-off to each player varies through the game according to the two probabilistic strategies. Since the pay-off in every move of the game varies randomly, we have variance over the total IPD game.

In this paper we analyze first-order Markovian strategies that have been recently considered in several studies of the IPD (Nowak, 1990; Nowak & Sigmund, 1993; Stephens *et al*., 1995) in order to express commitment of errors in typical strategists (TFT and ALLD). Next we introduce a formalization of the IPD game between two Markovian strategists (Stephens *et al*., 1995). We formalize the pay-off variance for each strategist according to stochastic strategies and stochastic game termination, and finally discuss the effect of pay-off variance on evolutionary dynamics given by the IPD game.

Markovian Strategy

Even though there are infinite numbers of possible strategies in the IPD, we can reasonably consider a restricted set of rational strategies (Nowak, 1990; Nowak & Sigmund, 1992, 1993). Following Nowak & Sigmund (1993), we consider the class of strategies represented by a four-tuple of probabilities of cooperating after receiving each of the four possible payoffs in the PID: $\mathscr{S} = (t, r, p, s) = (\Pr(C|T),$

$\Pr(C|R)$, $\Pr(C|P)$, $\Pr(C|S)$). Many strategies can be represented this way. Suppose that player 1 adopts strategy $\mathscr{S}_1 = (t_1, r_1, p_1, s_1)$, and player 2 adopts strategy $\mathscr{S}_2 = (t_2, r_2, p_2, s_2)$, then rewards gained by player 1 are determined by the game transition matrix, $\mathbf{M}_{1,2} =$

$$\begin{bmatrix} s_1(1-t_1) & (1-r_1)r_2 & (1-p_1)p_2 & (1-s_1)t_2 \\ s_2 t_1 & r_1 r_2 & p_1 p_2 & s_1 t_2 \\ (1-s_2)(1-t_1) & (1-r_1)(1-r_2) & (1-p_1)(1-p_2) & (1-s_1)(1-t_2) \\ (1-s_2)t_1 & r_1(1-r_2) & p_1(1-p_2) & s_1(1-t_2) \end{bmatrix}$$

where the entries in the first column represent the conditional probabilities of receiving the pay-offs 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 specify a strategy one must specify not only the strategy vector $\mathscr{S}$, but also initial behavior. We present the probability of initial cooperation as c. So, we can completely specify a strategy by specifying the strategy vector $\mathscr{S}$ and the scalar c, say $\Lambda_i = \{\mathscr{S}_i, c_i\}$. Hence any two interacting strategies specify a matrix $\mathbf{M}_{1,2}$, and initial state probability of pay-offs

$$\vec{\mathbf{y}}_{1,2} = \begin{bmatrix} (1-c_1)c_2 \\ c_1 c_2 \\ (1-c_1)(1-c_2) \\ c_1(1-c_2) \end{bmatrix}.$$

This generalization can express several strategies that have been nominated in the study of IPD such as ALLD, TFT, generous-TFT and so on. Further, the generalization can express the set of all matches of first order Markovian strategists.

When repetition of the game follows a geometric probability distribution, and the protagonist player is denoted 1 and the opponent is 2, the game transition matrix $\mathbf{M}_{1,2}$, initial pay-off probability vector $\vec{\mathbf{y}}_{1,2}$, pay-off vector $\vec{\mathbf{v}} = (T, R, P, S)$, and the game repetition parameter w completely characterize the pay-offs in the IPD. The expected pay-off, or arithmetic mean of player 1 in the IPD is $\mu_{1,2} = \vec{\mathbf{v}} \cdot (\mathbf{I} - w\mathbf{M}_{1,2})^{-1} \cdot \vec{\mathbf{y}}_{1,2}$, where $\mathbf{I}$ is an identity matrix (Stephens *et al*., 1995).

Formalization of Pay-off Variance

In the following we formalize the pay-off variance of a strategist in a single IPD game. Consider a sequence of pay-offs, $\{x_1, x_2, \ldots, x_1, \ldots\}$, of a strategist, where x_i is a pay-off at the i-th move. The pay-off x_i is a stochastic variable, and the length of game is also a stochastic variable. x_i depends on the strategy vectors $\mathscr{S}_i$ and the initial state probabilities c_i of the protagonist and the opponent. In the following we omit subscripts that express the strategists for convenience. The pay-off variance is defined as

$$\sigma^2 = E\left[\left(\sum x_i\right)^2\right] - E\left[\sum x_i\right]^2. \quad (1)$$

$E[\Sigma x_i]$ is given as $\vec{\mathbf{v}} \cdot (\mathbf{I} - w\mathbf{M})^{-1} \cdot \vec{\mathbf{y}}$ (Stephens *et al*., 1995). Focus on the first term of eqn (1): the expectation of a squared sum. We can express this as

$$E\left[\left(\sum x_i\right)^2\right] = E\left[E\left[\left(\sum_i^N x_i\right)^2 \middle| N\right]\right]$$

$$= \sum_N^\infty Pr(N) E\left[\left(\sum_i^N x_i\right)^2 \middle| N\right], \quad (2)$$

where N is the length of the game and $E[(\Sigma_i^N x_i)^2 | N]$ is the expected squared pay-off sum conditioned on any particular number of interactions N.

Now, given that $N = n, (\Sigma_i^n x_i)^2$ is distributed as the square of the sum of n random variables. Summing up all game lengths N, weighed by $Pr(N)$, we get the expectation of the squared sum. Expanding the square, $(\Sigma x_i)^2$, for each N yields:

$$\sum_N^\infty \Pr(N) E\left[\sum_i^N x_i^2 + \sum_i^N \sum_{j \neq i}^N x_i x_j \middle| N\right]$$

$$= \sum_N^\infty \Pr(N) \left(\sum_i^N E[x_i^2 | N] + \sum_i^N \sum_{j \neq i}^N E[x_i x_j | N]\right),$$

where $E[x_i^2 | N]$ and $E[x_i x_j | N]$ are the conditional expected square and co-product given each game length, N. Since we assume that $Pr(N)$ follows a geometric distribution with parameter w, we can rearrange the unconditional expected square of the pay-off sum as

$$\sum_{i=1}^\infty w^{i-1} E[x_i^2] + 2\left(\sum_{i=1}^\infty \sum_{j=i+1}^\infty w^{j-1} E[x_i x_j]\right).$$

Joint Probability Distribution of States

To get the expected co-products, $E[x_i x_j]$, we require the joint probability distribution of states for the i-th and j-th moves of the strategist. The joint probability distribution of states is generated by the matrix $\mathbf{M}$ and initial state probability vector $\vec{\mathbf{y}}$.

To begin, think about the joint probability distribution of states at the first and the next move. Let the initial distribution of pay-offs be $\vec{\mathbf{y}}_1 = (y_1, y_2, y_3, y_4)^T$, and let the subscript of the vector $\vec{\mathbf{y}}$ be the number of moves. The joint probability mass function of pay-offs between the first and the next move is

$$\mathbf{P}_{12} = \begin{pmatrix} m_{11}y_1 & m_{12}y_2 & m_{13}y_3 & m_{14}y_4 \\ m_{21}y_1 & m_{22}y_2 & m_{23}y_3 & m_{24}y_4 \\ m_{31}y_1 & m_{32}y_2 & m_{33}y_3 & m_{34}y_4 \\ m_{41}y_1 & m_{42}y_2 & m_{43}y_3 & m_{44}y_4 \end{pmatrix}, \quad (4)$$

where m_{ij} $(i, j = 1 \ldots 4)$ is an element of $\mathbf{M}$. Notice that an entry of $\mathbf{P}$, p_{ij}, shows the probability of pay-off i in the next move given pay-off j in the initial play (i and j follow with sequence of pay-offs (T, R, P, S)).

We denote $\mathbf{P}_{12}$ by the product $[\mathbf{M}]\cdot(\vec{\mathbf{y}}_1)$ for convenience in the following formalization. It might be useful to mention that the composition theorem and commutative law hold,

$$[\mathbf{M}]\cdot(\vec{\mathbf{y}}) + [\mathbf{M}']\cdot(\vec{\mathbf{y}}) = [\mathbf{M} + \mathbf{M}']\cdot(\vec{\mathbf{y}}) \quad (5)$$

$$a[\mathbf{M}]\cdot(\vec{\mathbf{y}}) = [a\mathbf{M}]\cdot(\vec{\mathbf{y}}) = [\mathbf{M}]\cdot(a\vec{\mathbf{y}}),$$

where a is a scalar.

Similarly, the joint probability distribution of states in the initial and third moves is expressed as $\mathbf{P}_{13} = [\mathbf{M}^2]\cdot(\vec{\mathbf{y}}_1)$. Thus, the joint probability distribution for the states in the initial move and the i-th move is given as $\mathbf{P}_{1i} = [\mathbf{M}^{i-1}]\cdot(\vec{\mathbf{y}}_1)$. Generally, the joint probability distribution of states between the i-th and j-th $(i < j)$ moves is given by

$$\mathbf{P}_{ij} = [\mathbf{M}^{j-i}]\cdot(\vec{\mathbf{y}}_i) = [\mathbf{M}^{j-i}]\cdot(\mathbf{M}^{i-1}\cdot\vec{\mathbf{y}}_1), \quad (7)$$

where $\vec{\mathbf{y}}_i = \mathbf{M}^{i-1}\cdot\vec{\mathbf{y}}_1$ (see Stephens *et al*., 1995). Notice that the joint probability distribution, $\mathbf{P}_{ii}$, (the diagonal elements are the probabilities of getting T, R, P, S in the i-th move, and the off-diagonal elements are all zero) is

$$\mathbf{P}_{ii} = [\mathbf{M}^{i-i}]\cdot(\vec{\mathbf{y}}_i) = [\mathbf{M}^0]\cdot(\vec{\mathbf{y}}_i) = [\mathbf{I}]\cdot(\vec{\mathbf{y}}_i) = [\mathbf{I}]\cdot(\mathbf{M}^{i-1}\cdot\vec{\mathbf{y}}_1). \quad (8)$$

Expectation of Pay-off Products and Co-products

We define the matrix of pay-off products and co-products as

$$\mathbf{V} = \begin{pmatrix} T^2 & TR & TP & TS \\ RT & R^2 & RP & RS \\ PT & PR & P^2 & PS \\ ST & SR & SP & S^2 \end{pmatrix} \quad (9)$$

The expected pay-off product, $E[x_i^2]$ is given as

$$E[x_i^2] = \sum_k^4 \sum_l^4 v_{kl} p_{(ii)kl},$$

where v_{kl} and $p_{(ii)kl}$ are the k-th row and l-th column element of matrix $\mathbf{V}$ and $\mathbf{P}_{ii}$, respectively. Similarly, the expected pay-off co-product, $E[x_i x_j]$, $(i < j)$ is given as

$$E[x_i x_j] = \sum_k^4 \sum_l^4 v_{kl} p_{(ij)kl}.$$

Notice that for the operator $[\mathbf{V}]\cdot[\mathbf{P}_{ij}] = E[x_i x_j]$, the composition theorem and commutative law hold,

$$[\mathbf{V}]\cdot[\mathbf{P}] + [\mathbf{V}]\cdot[\mathbf{P}'] = [\mathbf{V}]\cdot[\mathbf{P} + \mathbf{P}'] \quad (10)$$

$$a[\mathbf{V}]\cdot[\mathbf{P}] = [a\mathbf{V}]\cdot[\mathbf{P}] = [\mathbf{V}]\cdot[a\mathbf{P}], \quad (11)$$

where a is a scalar value.

Then, the expected pay-off product at the i-th move is

$$E[x_i^2] = [\mathbf{V}]\cdot[\mathbf{P}_{ii}] = [\mathbf{V}]\cdot[[\mathbf{I}]\cdot(\mathbf{M}^{i-1}\cdot\vec{\mathbf{y}}_1)], \quad (12)$$

and the expected pay-off co-product, $E[x_i x_j]$, $(i < j)$, for the i-th and the j-th moves is

$$E[x_i x_j] = [\mathbf{V}]\cdot[\mathbf{P}_{ij}] = [\mathbf{V}]\cdot[[\mathbf{M}^{j-i}]\cdot(\mathbf{M}^{i-1}\cdot\vec{\mathbf{y}}_1)]. \quad (13)$$

Pay-off Variance

The first term of eqn (3) (parts of variance) is symbolically expressed as

$$\sum_{i=1}^{\infty} w^{i-1} E[x_i^2] = \sum_{i=1}^{\infty} w^{i-1} [\mathbf{V}]\cdot[[\mathbf{I}]\cdot(\mathbf{M}^{i-1}\cdot\vec{\mathbf{y}}_1)] \quad (14)$$

by substituting eqn (12). Then, eqn (14) can be rearranged as

$$\begin{aligned} &\sum_{i=1}^{\infty} w^{i-1}[\mathbf{V}]\cdot[[\mathbf{I}]\cdot(\mathbf{M}^{i-1}\cdot\vec{\mathbf{y}}_1)] \\ &= [\mathbf{V}]\cdot\left[\sum_{i=1}^{\infty} w^{i-1}[\mathbf{I}]\cdot(\mathbf{M}^{i-1}\cdot\vec{\mathbf{y}}_1)\right] \\ &= [\mathbf{V}]\cdot\left[[\mathbf{I}]\cdot\left(\left(\sum_{i=1}^{\infty} w^{i-1}\mathbf{M}^{i-1}\right)\cdot\vec{\mathbf{y}}_1\right)\right] \\ &= [\mathbf{V}]\cdot[[\mathbf{I}]\cdot((\mathbf{I} - w\mathbf{M})^{-1}\cdot\vec{\mathbf{y}}_1)]. \quad (15) \end{aligned}$$

Next, the term in the parentheses of the second term of eqn (3) (parts of covariance) is rearranged as

$$\sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty} w^{j-1}E[x_i x_j] = \sum_{i=1}^{\infty}\sum_{i<j}^{\infty} w^{j-1}E[x_i x_j]$$

$$= \sum_{j=2}^{\infty} w^{j-1}E[x_1 x_j] + \sum_{j=3}^{\infty} w^{j-1}E[x_2 x_j] + \cdots. \quad (16)$$

Substitution of eqn (13) into eqn (16) yields,

$$\sum_{j=2}^{\infty} w^{j-1}[\mathbf{V}]\cdot[[\mathbf{M}^{j-1}]\cdot(\mathbf{I}\cdot\hat{\mathbf{y}}_1)]$$
$$+ \sum_{j=3}^{\infty} w^{j-1}[\mathbf{V}]\cdot[[\mathbf{M}^{j-2}]\cdot(\mathbf{M}\cdot\hat{\mathbf{y}}_1)] + \cdots \quad (17)$$

Using the composition theorem and commutative law, the above is

$$[\mathbf{V}]\cdot\left[\left[\sum_{j=2}^{\infty} w^{j-1}\mathbf{M}^{j-1}\right]\cdot(\mathbf{I}\cdot\hat{\mathbf{y}}_1)\right]$$
$$+ [\mathbf{V}]\cdot\left[\left[\sum_{j=3}^{\infty} w^{j-1}\mathbf{M}^{j-2}\right]\cdot(\mathbf{M}\cdot\hat{\mathbf{y}}_1)\right] + \cdots. \quad (18)$$

The summations, $\Sigma_{j=2}^{\infty} w^{j-1}\mathbf{M}^{j-1}$, $\Sigma_{j=3}^{\infty} w^{j-1}\mathbf{M}^{j-2}, \ldots,$ can be written as $w\mathbf{M}\cdot(\mathbf{I}-w\mathbf{M})^{-1}$, $w^2\mathbf{M}\cdot(\mathbf{I}-w\mathbf{M})^{-1}, \ldots$ (see Appendix of Stephens *et al.*, 1995). Thus, eqn (17) is

$$[\mathbf{V}]\cdot[[w\mathbf{M}\cdot(\mathbf{I}-w\mathbf{M})^{-1}]\cdot(\mathbf{I}\cdot\hat{\mathbf{y}}_1)]$$
$$+[\mathbf{V}]\cdot[[w^2\mathbf{M}\cdot(\mathbf{I}-w\mathbf{M})^{-1}]\cdot(\mathbf{M}\cdot\hat{\mathbf{y}}_1)]$$
$$+[\mathbf{V}]\cdot[[w^3\mathbf{M}\cdot(\mathbf{I}-w\mathbf{M})^{-1}]\cdot(\mathbf{M}^2\cdot\hat{\mathbf{y}}_1)] + \cdots \quad (19)$$

and further rearrangement shows

$$w[\mathbf{V}]\cdot[[\mathbf{M}\cdot(\mathbf{I}-w\mathbf{M})^{-1}]\cdot((\mathbf{I}+w\mathbf{M}+w^2\mathbf{M}^2+\cdots)\cdot\hat{\mathbf{y}}_1)]$$
$$= w[\mathbf{V}]\cdot[[\mathbf{M}\cdot(\mathbf{I}-w\mathbf{M})^{-1}]\cdot((\mathbf{I}-w\mathbf{M})^{-1})\cdot\hat{\mathbf{y}}_1)]. \quad (20)$$

Substituting eqns (15) (the parts of variance) and (20) (the parts of covariance) into eqn (3), we get the unconditional expectation of the pay-off sum squared, $E[(\Sigma x_i)^2]$,

$$[\mathbf{V}]\cdot[[\mathbf{I}]\cdot((\mathbf{I}-w\mathbf{M})\cdot\hat{\mathbf{y}}_1)]$$
$$+ 2w[\mathbf{V}]\cdot[[\mathbf{M}\cdot(\mathbf{I}-w\mathbf{M})^{-1}]\cdot((\mathbf{I}-w\mathbf{M}^{-1})\cdot\hat{\mathbf{y}}_1)]$$
$$= [\mathbf{V}]\cdot[[\mathbf{I}+2w(\mathbf{M}\cdot(\mathbf{I}-w\mathbf{M})^{-1})]\cdot((\mathbf{I}-w\mathbf{M})^{-1}\cdot\hat{\mathbf{y}}_1)]. \quad (21)$$

Substituting eqn (21) into $E[(\Sigma x_i)^2]$, and $\vec{\mathbf{v}}\cdot(\mathbf{I}-w\mathbf{M})\cdot\hat{\mathbf{y}}$ into $E[(\Sigma x_i)]$ of (1), the pay-off variance in the IPD can be given as

$$\sigma^2 = [\mathbf{V}]\cdot[[\mathbf{I}+2w(\mathbf{M}\cdot(\mathbf{I}-w\mathbf{M})^{-1})]\cdot((\mathbf{I}-w\mathbf{M})^{-1}\cdot\hat{\mathbf{y}}_1)]$$
$$-E[\Sigma_i x_i]^2 = [\mathbf{V}]\cdot[[\mathbf{I}+2w(\mathbf{M}\cdot(\mathbf{I}-w\mathbf{M})^{-1})]$$
$$\cdot((\mathbf{I}-w\mathbf{M})^{-1}\cdot\hat{\mathbf{y}}_1)] - (\vec{\mathbf{v}}\cdot(\mathbf{I}-w\mathbf{M})^{-1}\cdot\hat{\mathbf{y}}_1)^2. \quad (22)$$

$\mathbf{M}$ is a non-negative matrix. Therefore, the Frobenius theorem shows that the dominant latent root of matrix $\mathbf{M}$ is $0 \leqslant \lambda_0 \leqslant 1$ (Alexander, 1983). Then, there exists a $(\mathbf{I}-w\mathbf{M})^{-1}$ for any w such that $0 \leqslant w < 1(\leqslant 1/\lambda_0)$. Equation (22) is a general pay-off variance in the IPD played by Markovian strategists under stochastic termination of the game according to a geometric probability distribution.

Strategy, Error, Mean and Variance

Stochastic processes have not been much considered in the study of the IPD (Axelrod & Hamilton, 1981; Mesterton-Gibbon & Dugatkin, 1992). A stochastic strategies are modeled as Markov process (Nowak & Sigmund, 1992, 1993; Stephens *et al.*, 1995). The pay-off variance of a particular strategist depends on the strategy of its own and that of the opponent, and the value of the probability of continued, w. As we discuss later, the effect of pay-off variance plays an important effect in a small population.

Since the underlying mechanism for cooperation would be reciprocity, we here consider TFT as a reciprocal strategist against ALLD. ALLD is unequivocally the focal strategy to analyse invasion condition into population consisting of a certain cooperation oriented strategists in the context of the IPD argument.

Error free ALLD can be defined as $\Lambda_D = \{\mathscr{S}_D, c_D\} = \{(0, 0, 0, 0), 0\}$ and error free TFT can be defined as $\Lambda_T = \{\mathscr{S}_T, c_T\} = \{(1, 1, 0, 0), 1\}$. TFT's resistance to invasion by ALLD has been studied by considering arithmetic mean of pay-offs to the strategists (Axelrod & Hamilton, 1981; Mesterton-Gibbon & Dugatkin, 1992). However, we will show how the variance of pay-off of each strategist affects the stability of TFT population against an intruder of ALLD in the next section. When error exists in the behavior of the strategists, the ALLD strategist with error can be defined as $\Lambda_D = \{\mathscr{S}_D, c_D\} = \{(\epsilon, \epsilon, \epsilon, \epsilon), \epsilon\}$ and the TFT strategist can be defined as $\Lambda_T = \{\mathscr{S}_T, c_T\} = \{(1-\epsilon,\ 1-\epsilon, \epsilon, \epsilon), 1-\epsilon\}$.

Before evaluating the pay-off variances, we show the effect of error to the mean pay-offs. Means are differently affected by strategic error in the different matches. We evaluate not only $\mu_{T,T}$ and $\mu_{D,T}$ but also $\mu_{T,D}$, in order to consider invasion of a single ALLD

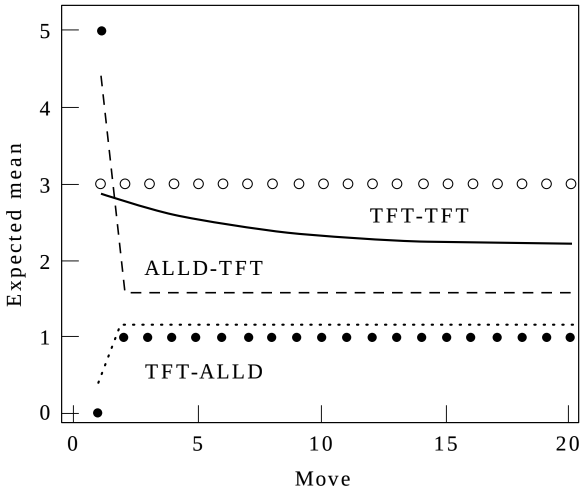

FIG. 1. When there is no error, pay-offs are $R = 3$, $R, \ldots$ for TFT against TFT (open circles), $S = 0$, $P = 1$, $P, \ldots$ for TFT against ALLD (solid circles), and $T = 5$, $P = 1$, $P, \ldots$ for ALLD against TFT (solid circles), respectively. The solid line is the pay-off mean of each move for TFT against TFT, the dotted line is the pay-off mean of each move for TFT against ALLD, and the dashed line is the pay-off mean for ALLD against TFT in the case of $\epsilon = 0.1$, respectively.

individual into TFT small population. We set $\vec{\mathbf{v}} = (T = 5, R = 3, P = 1, S = 0)$ and $\epsilon = 0.1$.

When there exists the behavioral error in a game of TFT against TFT, the probability distribution of pay-offs in each move converges to asymptotic state distribution $(T, R, P, S) = (1/4, 1/4, 1/4, 1/4)$, as iteration of game proceeds (Nowak, 1990; Nowak & Sigmund, 1992, 1993). The expected pay-off per single move taken by the error player will converge to $\frac{1}{4}(T + R + P + S)$, which is smaller than the pay-off of error free player, R (Fig. 1). Therefore, error decreases the mean pay-off of the TFT strategist (Fig. 2).

A TFT against ALLD gets $S = 0$ in the first move and $P = 1$ in the following moves, when there exists no error. When there exists the error, the expected pay-off in the first move is 0.41 instead of $S = 0$ (see Fig. 1). The second largest eigenvalue of the pay-off transition matrix $\mathbf{M}_{T,D}$ is 9×10^{-9}. The probability distribution of pay-offs quickly converges to an asymptotic one. The expected pay-off to the asymptotic distribution is 1.202 instead of $P = 1$ (see Fig. 1). Since the expected pay-off in every move of IPD with error in every move is greater than that without error, error increases the mean pay-off of the TFT strategist (see Fig. 2).

An ALLD against TFT gets $T = 5$ in the first move and $P = 1$ in the following moves, when no error exists. When error exists, the expected pay-off of the first move is 4.41 instead of $T = 5$ (see Fig. 1). As w is small, the decrement of expected pay-off in the first move works to decrease the mean pay-off (Fig. 2). The second largest eigenvalue of the pay-off transition matrix $\mathbf{M}_{T,D}$ is 9×10^{-9}, and the convergence to an asymptotic state, is very fast. The expected pay-off to the asymptotic distribution is 1.602 instead of $P = 1$ (see Fig. 1). Therefore, as w is large, the increment of expected pay-off in the latter moves is effective in increasing the mean pay-off (Fig. 2).

When there is no error in any move of the strategists, the relative magnitude of the pay-off variances is obvious. The sequence of pay-offs of a TFT that plays with TFT is "R, $R, \ldots$". The

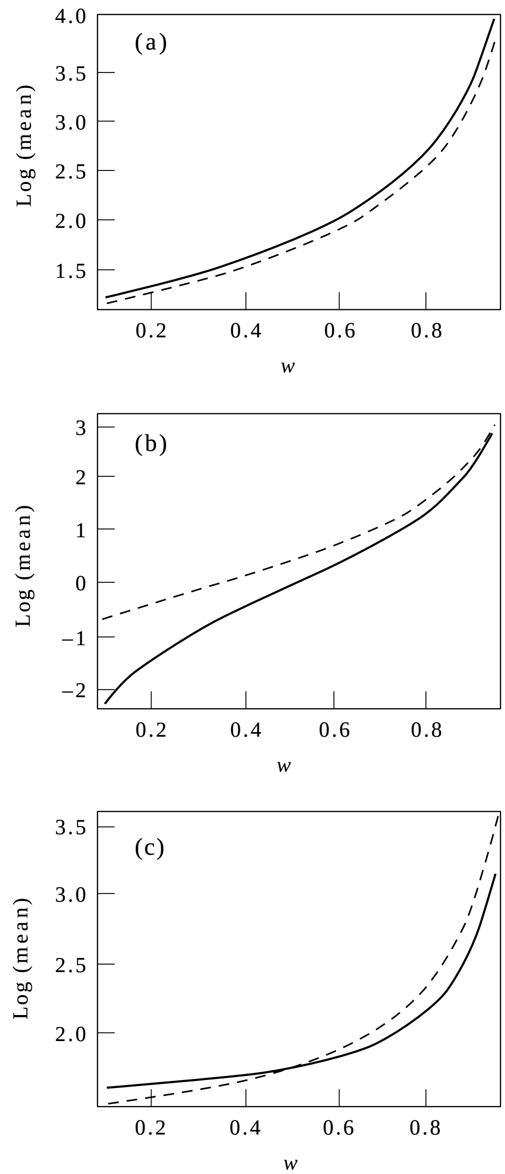

FIG. 2. The pay-off mean of TFT and ALLD transformed by a logarithm. The pay-offs in a single game are $T = 5$, $R = 3$, $P = 1$, $S = 0$. In each figure, the protagonist player and the opponent are shown as "protagonist-opponent". (a) TFT–TFT, (b) TFT–ALLD, (c) ALLD–TFT. *Key*: — $\epsilon = 0$, - - $\epsilon = 0.1$.

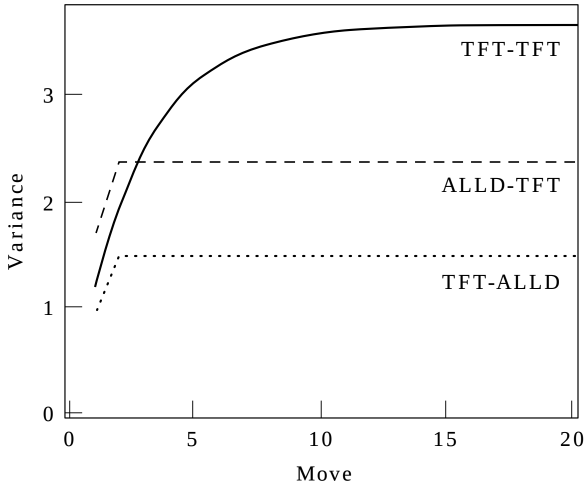

FIG. 3. The solid line is the pay-off variance of each move for TFT against TFT, the dotted line is the pay-off variance of each move for TFT against ALLD, and the dashed line is the pay-off variance for ALLD against TFT in the case that $\epsilon = 0.1$, respectively.

sequence of pay-offs of TFT that plays with ALLD is "S, P, P, . . . ". And sequence of pay-offs of ALLD that plays with TFT is "T, P, P, . . .". The first payoffs of the three sequences are given with a probability of one and the following ones are given by a probability that is an identical geometric probability distribution with parameter, w. The pay-off variance of TFT against TFT is $R^2w/[(1-w)^2]$, and that of TFT against ALLD and of ALLD against TFT are both $P^2w/[(1-w)^2]$. Therefore, the relative magnitude of pay-off variance of TFT that plays with TFT is $(R/P)^2 > 1$ times the variance of the others. This suggests that TFT would owe a disadvantage because of the large variance.

By including error in the behavior, the relative magnitudes of the pay-off variance of the strategists are not obvious. Pay-off variance is composed of variance of game length, pay-off variance within a move, and the pay-off covariance between moves. We can understand heuristically the difference of the total pay-off variances with and without error considering the variance of game length and the pay-off variance within a move. Pay-off variances in a single move quickly increase and converge to the asymptotic values after the second move for TFT against ALLD (1.505), and for ALLD against TFT (2.384) (Fig. 3). On the other hand, pay-off variance in a single move for TFT against itself gradually increases and converges to the asymptotic value (3.687) (Fig. 3).

We have shown the pay-off variances of IPD in Fig. 4. The non-zero pay-off variances due to error in a single move increases the variances of total pay-off, when the repetition is a few (i.e. small w) (see Fig. 3 and Fig. 4). On the contrary, when w is large, increment or decrement of expected pay-off in a single move plays an important role in determining the total pay-off variance (see Fig. 3 and Fig. 4), because the variance of game length times the square of the expected pay-off of a single move in an asymptotic state, dominates in the total variance.

Pay-off Variance, Fitness and Stability

Suppose a situation arises where a single ALLD tries to invade a $N-1$ TFT strategists population.

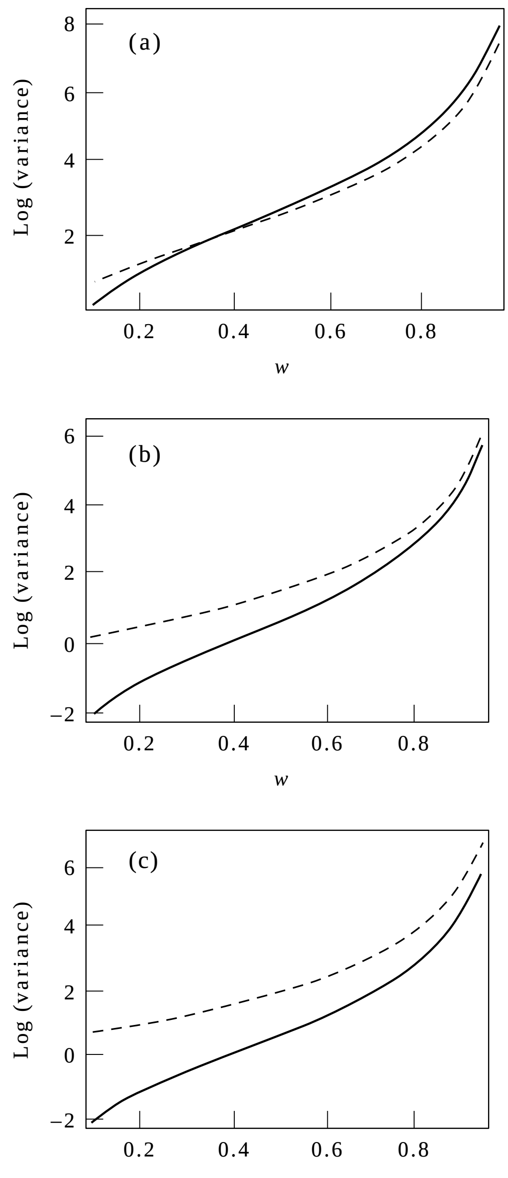

FIG. 4. The pay-off variance of TFT and ALLD transformed by a logarithm. The pay-offs in a single game are $T = 5$, $R = 3$, $P = 1$, $S = 0$. In each figure, the protagonist player and the opponent are shown as "protagonist-opponent". (a) TFT–TFT, (b) TFT–ALLD, (c) ALLD–TFT. *Key*: — $\epsilon = 0$, - - $\epsilon = 0.1$.

The relative number of interaction for an arbitrary chosen TFT to TFT or ALLD is proportional to $N-2$ and 1, respectively, and the relative number of interaction for the ALLD to TFT is proportional to $N-1$.

The pay-off of the ALLD player who plays against TFT distributes with a mean of $\mu_{D,T}$ and a variance of $\sigma^2_{D,T}$; the pay-off of the most TFT player who plays against TFT distributes with a mean of $\mu_{T,T}$ and a variance of $\sigma^2_{T,T}$; and the pay-off of a TFT who plays against ALLD distributes with a mean of $\mu_{T,D}$ and a variance of $\sigma^2_{T,D}$.

The pay-off mean and variance of TFT is $[(N-2)/(N-1)]\mu_{T,T}+[1/(N-1)]\mu_{T,D}$ and $[(N-2)/(N-1)]\sigma^2_{T,T}+[1/(N-1)]\sigma^2_{T,D}$, respectively, if matching is conducted only once. A difference of the pay-offs between ALLD and TFT is distributed with the mean,

$$\mu_{D,T}-\left(\frac{N-2}{N-1}\mu_{T,T}+\frac{1}{N-1}\mu_{T,D}\right),$$

and variance

$$\sigma^2_{D,T}+\frac{N-2}{N-1}\sigma^2_{T,T}+\frac{1}{N-1}\sigma^2_{T,D}.$$

When the number of IPD played by all players is n, the difference distributes with mean,

$$n\left(\mu_{D,T}-\left(\frac{N-2}{N-1}\mu_{T,T}+\frac{1}{N-1}\mu_{T,D}\right)\right)$$

and variance,

$$n\left(\sigma^2_{D,T}+\frac{N-2}{N-1}\sigma^2_{T,T}+\frac{1}{N-1}\sigma^2_{T,D}\right).$$

The ration of standard deviation to mean (coefficient of variation) is

$$\frac{\sqrt{\sigma^2_{D,T}+\frac{N-2}{N-1}\sigma^2_{T,T}+\frac{1}{N-1}\sigma^2_{T,D}}}{\sqrt{n}\left(\mu_{D,T}-\left(\frac{N-2}{N-1}\mu_{T,T}+\frac{1}{N-1}\mu_{T,D}\right)\right)}.$$

As n becomes large, the coefficient of variation will vanish. This is the reason why most studies did not consider the pay-off variance and only considered the arithmetic means. Furthermore, if the population size N is large enough, the mean and variance of a TFT who plays against ALLD becomes less weighted, and this is the reason why the ESS analysis does not consider $\mu_{T,D}$.

The number of IPD played by players in the population depends on the population size and the assumption of how matchings occur. The number of IPD played by the players tends to be small, when the population is small. Hence, to clarify the effect of pay-off variance, we consider the most extreme case, where each member plays one IPD in their lifetime. Then the variance of pay-off plays a considerably more important role in the evolutionary process.

In small populations, not only pay-off variance, but also small population size itself, generates stochastic fluctuation in evolutionary dynamics. In finite population dynamics, there should exist a strong stochastic fluctuation in the relative frequency of strategists between generations (Kimura, 1983).

The prediction of evolutionary dynamics of the IPD game considers that these factors are of a different form than those when only the arithmetic means of pay-offs are considered (standard stability analysis). However, as is often the case with some other stochastic modeling, consideration of several stochastic factors makes the evolutionary process intractable analytically (Mangel & Clark, 1988).

Consider the first step of invasion by ALLD into a TFT population with finite size. Assume that the intruder is only one ALLD individual into the $N-1$ TFT strategist population. Suppose that the expected proportion of pay-offs of the invader, ALLD, among the total pay-offs of all the players in the population in the first generation. If the pay-offs can be assumed to be parallel to the fitness, the proportion will affect the proportion of the strategist in the next generation.

The proportion of pay-off of the intruder, ALLD, is expressed as

$$P_{\text{ALLD}}=\frac{\alpha}{\alpha+\beta_\alpha+\sum_{i=1}^{N-2}\beta_i}, \tag{23}$$

where α, β_α and β_i are the stochastic variables of pay-off of the intruder, ALLD, a resident strategist that plays with the intruder, TFT–ALLD, and the i-th resident strategist, that plays with the same strategist, TFT–TFT, respectively. Equation (23) is a concave upward function with respect to variable α and a concave downward with respect to variables β_α and β_i. $E[P_{\text{ALLD}}]$ decreases with increasing variance of α and increases with increasing variance of β_α and β_i (DeGroot, 1970). An approximation by the Taylor expansion presents how pay-off variance of the strategists affects expected proportion of pay-offs of the intruder, ALLD. The approximation is,

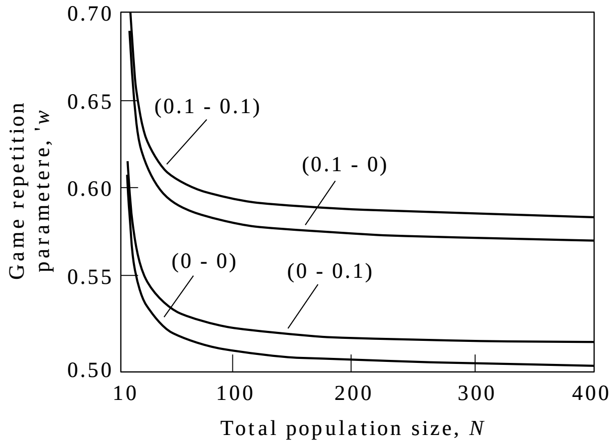

FIG. 5. The critical game repetition parameter ws with respect to the population size in consideration of pay-off variance. The numbers in parenthesis are the error of resident player, TFT, and the error of the invader, ALLD, respectively.

$$E[P_{\mathrm{ALLD}}] \approx \frac{\mu_\alpha}{\mu_\alpha + \mu_{\beta_\alpha} + (N-2)\mu_\beta} + \frac{\mu_\alpha(\sigma^2_{\beta_\alpha} + (N-2)\sigma^2_\beta) - (\mu_{\beta_\alpha} + (N-2)\mu_\beta)\sigma^2_\alpha}{(\mu_\alpha + \mu_{\beta_\alpha} + (N-2)\mu_\beta)^3}, \quad (24)$$

where μ and σ^2 are the pay-off mean and variance of players. Equation (24) indicates that a smaller (or larger) variance and larger (or smaller) mean of the intruder compared with that of the resident player leads to an advantage (or disadvantage) for the intruder to be selected in the next generation. Further, the magnitude of the second term of eqn (24) is $O((N-2)^{-2})$. Hence, as N goes to infinity, the variance component vanishes. On the contrary, when N is small, then variances in the numerator of the second term may influence the expected proportion of the ALLD strategist in the next generation. In order to feel the effect of pay-off variance, not only the means, but also the variances should be calculated.

The pay-off means and variances depend on the strategy of both players and game repetition parameter w. Hence, error would affect the limit of the repetition parameter above which ALLD cannot invade in a given population size. Equation (24) tells us the fitness consequence of an ALLD player within the first generation of invasion in a finite population. If the value of eqn (24) is less than $1/N$, the population statistically tends to be protected by invasion of the ALLD. We calculated the boundary condition of game repetition parameter, w, with respect to population size, N using eqn (24).

In Fig. 5 we plotted the curves above which an invading ALLD cannot expect equal or larger fitness than a resident of the population. Even though TFT gets the larger arithmetic mean, the variance of its pay-off is also larger, when the expected iterations of the game are sufficiently large. Above the boundary of the parameter, w, ALLD does not have an advantage to invade into the TFT population. The boundary becomes a bit larger than predicted by the deterministic model. In a small population more repetition is required in order to prevent invasion of ALLD. Thus, the ALLD player can expectedly invade in a small population. The critical repetition parameter value, w increases with error, and error takes a side with an invading ALLD.

Surprisingly, if the error is not symmetrical between the two strategists, the critical w increases with the error of the invader, ALLD. This means that an ALLD with an error is a better invader than an ALLD that has no error. If error of each strategist is caused by independent random mutation, selection would act upon each strategist such that error is favored by ALLD and aversed by TFT.

Conclusions

Stochastic strategies and their stability in IPD have been considered by Boyd & Lorberbaum (1987) and Farrell & Ware (1989). The main theoretical consequences are that there exists no pure single strategy in an infinite population, when strategies are deterministic; but if a stochastic factor is incorporated in strategies, there exists an evolutionarily stable strategy (Boyd, 1989). Simulation studies suggested some robust strategies in a huge set of strategies in IPD (Nowak & Sigmund, 1992, 1993).

Infinite population is a common assumption of these studies according to the tradition of evolutionarily stable analysis (Maynard Smith, 1982). However, an infinite large population is not a unique reality in a natural situation. In a small population, the consequences of stochastic factors play an important role and it is important to survey the stochastic factors. Even though the total effects of stochasticity in the evolutionary process of the IPD game are not obvious, but the general analysis presented above that the survey is only the first step in the invasion of an ALLD, we have shown the effect of pay-off variances of both resident individuals and an invader, and how behavioral errors affect the invasion possibility of an ALLD. We have stressed the effect of behavioral error and finite population size. This destabilizes the general arguments of the evolutionary process (Axelrod & Hamilton, 1981; Mesterton-Gibbon & Dugatkin, 1992; Nowak & Sigmund, 1992, 1993).

We thank Tom Caraco and an anonymous referee for helpful suggestions that markedly improved the manuscript.

REFERENCES

ALEXANDER, B. (1983). *Applied Matrix Algebra in the Statistical Sciences*. New York: North-Holland.

AXELROD, R. (1984). *The Evolution of Cooperation*. New York: Basic Books Inc.

AXELROD, R. & HAMILTON, W. (1981). The evolution of cooperation. *Science*, **211,** 1390–1396.

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.

DEGROOT, M. H. (1970). *Optimal Statistical Decisions*. New York: McGraw-Hill Book Company.

DUGATKIN, L. A. (1991). Dynamics of the tit for tat strategy during predator inspection in the guppy (*Poeciliareticulata*). *Behav. Ecol. Sociobiol.* **29,** 127–132.

DUGATKIN, L. A. (1994). Altruism, tit for tat and ''outlaw'' genes. *Evol. Ecol.* **8,** 431–437.

FARRELL, J. & WARE, R. (1989). Evolutionary stability in the repeated prisoner's dilemma. *Theor. Pop. Biol.* **36,** 161–166.

FELDMAN, M. W. & THOMAS, E. A. C. (1987). Behavior-dependent contexts for repeated plays of the prisoner's dilemma. ii: Dynamical aspects of the evolution of cooperation. *J. theor. Biol.* **128,** 297–315.

KIMURA, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press.

MANGEL, M. & CLARK, C. W. (1988). *Dynamic Modeling in Behavioural Ecology*. Princeton: Princeton University Press.

MAYNARD SMITH, J. (1982). *Evolution and the Theory of Games*. London: Cambridge University Press.

MESTERTON-GIBBON, M. (1992). On the iterated prisoner's dilemma in a finite population. *Bull. of Math. Biol.* **54,** 423–443.

MESTERTON-GIBBON, M. & DUGATKIN, L. A. (1992). Cooperation among unrelated individuals: evolutionary factors. *Quart. Rev. of Biol.* **67,** 267–281.

NOWAK, M. (1990). Stochastic strategies in the prisoner's dilemma. *Theor. Pop. Biol.* **38,** 93–112.

NOWAK, M. & SIGMUND, K. (1992). Tit for tat in heterogenous populations. *Nature* **355,** 250–253.

NOWAK, M. & SIGMUND, K. (1993). A strategy of win–stay, lose–shift that outperforms tit-for-tat in the prisoner's dilemma game. *Nature* **364,** 56–58.

STEPHENS, D. W., NISHIMURA, K. & TOYER, K. B. (1995). Error and discounting in the iterated prisoner's dilemma. *J. theor. Biol.* **176,** 457–469.