The iterated continuous prisoner’s dilemma game cannot explain the evolution of interspecific mutualism in unstructured populations

István Scheuring*

Department of Plant Taxonomy and Ecology, Research Group of Ecology and Theoretical Biology, Hungarian Academy of Sciences and Eötvös University Budapest, Pázmány P. sétány 1/c, 1117 Budapest, Hungary

Received 14 May 2004; accepted 29 July 2004
Available online 18 September 2004

Abstract

The evolutionary origin of inter- and intra-specific cooperation among non-related individuals has been a great challenge for biologists for decades. Recently, the continuous prisoner’s dilemma game has been introduced to study this problem. In function of previous payoffs, individuals can change their cooperative investment iteratively in this model system. Killingback and Doebeli (Am. Nat. 160 (2002) 421–438) have shown analytically that intra-specific cooperation can emerge in this model system from originally non-cooperating individuals living in a non-structured population. However, it is also known from an earlier numerical work that inter-specific cooperation (mutualism) cannot evolve in a very similar model. The only difference here is that cooperation occurs among individuals of different species. Based on the model framework used by Killingback and Doebeli (2002), this Note proves analytically that mutualism indeed cannot emerge in this model system. Since numerical results have revealed that mutualism can evolve in this model system if individuals interact in a spatially structured manner, our work emphasizes indirectly the role of spatial structure of populations in the origin of mutualism.

Keywords: Adaptive dynamics; Coevolution; Evolutionary stability; Reciprocal altruism; Variable investment

Mutualistic interactions between members of different species are widespread and play a central role in ecosystems (Boucher et al., 1982; Bronstein, 2001a, b). However, the evolution of mutualism has been a great challenge for theoreticians for decades. How can mutualist individuals emerge in populations where all the others are non-mutualists, and how can mutualists prevent cheaters from spreading in the population?

The classical theoretical framework for studying cooperation of unrelated individuals within a species is the Prisoner’s Dilemma Game (Trivers, 1971), in which partners can choose either a defective (cheating) or a cooperative strategy. If both partners defect, they get a smaller fitness than if both cooperate, but a defector has an even higher fitness value if its opponent cooperates. However, the cheated cooperator receives the smallest fitness if its opponent is a defector. It is easy to see that defection is the only evolutionary stable state in this model, and cooperators cannot spread in a defecting population. On the other hand, defectors can invade and destroy cooperation in a cooperative population (Trivers, 1971; Axelrod and Hamilton, 1981). Cooperative strategies emerge and are stable against the invasion of defective ones if individuals can interact with each others repeatedly (Axelrod and Hamilton, 1981; Nowak and Sigmund, 1992, 1993). The general conclusions of intensive work in this field are that the successful strategies are either those that punish defector and reward cooperation in repeated encounters (Tit-for-Tat, Generous Tit-for-Tat) or those that retain their previous successful strategy (Pavlov).

Doebeli and Knowlton (1998) realized that there is no way for individuals to vary the degree of cooperation in
these model systems, which is an unrealistic assumption in many biological situations. Therefore, they introduced an interspecific prisoner’s dilemma game where the payoffs can change according to the investments made by the partners. The level of investment $I_h$ by a host individual involves a cost $C(I_h)$ to the host and gives a benefit $B(I_h)$ to its symbiont partner. (Naturally, if the investment is zero then the cost and the benefit are zero too.) Thus if the symbiont partner invests $I_s$ then the payoff for the host is $S(I_h, I_s) = B(I_h) - C(I_h)$ and similarly, the payoff of the symbiont is $S(I_s, I_h) = B(I_h) - C(I_s)$. (Species are called “host” and “symbiont” only for convenience, but they have a completely symmetrical role in the mutualistic interaction.) In the special case when one of them makes no investment to the interaction we retrieve the classical prisoner’s dilemma game again (Table 1).

Mutualistic connections were considered by Doebeli and Knowlton (1998) as a series of prisoner’s dilemma games with variable investment, where the investment of a partner depends on its payoff received in the previous round. Interaction is a general phrase covering different kinds of exchange of commodities (Bronstein, 2001a, b). The investment decision is determined by two parameters: $\alpha$, the initial offer, and $\beta$, the reward rate or the rate of increase of investment depending on the payoff $S$ in the preceding round. So the investment of a host in round $k + 1$ is

$$I_{h}^{(k+1)} = \alpha_h + \beta_h S(I_h^{(k)}, I_s^{(k)}),$$

(1)

where $S(I_h^{(k)}, I_s^{(k)})$ is the payoff of the host individual in round $k$, if it invested $I_h^{(k)}$ and its symbiont opponent invested $I_s^{(k)}$ in the previous interaction. To avoid negative investment $I_{h}^{(k+1)}$ is set to zero if $S(I_h^{(k)}, I_s^{(k)}) < -\alpha_h/\beta_h$. Similarly, the opponent’s investment in the $k + 1$-th round is

$$I_{s}^{(k+1)} = \alpha_s + \beta_s S(I_s^{(k)}, I_h^{(k)}),$$

(2)

where $\alpha_s$ and $\beta_s$ are initial offer and reward rate of the opponent. Since benefit generally increases less at higher investment (see e.g. Altman, 1979; Schulman and Rubenstein, 1983; Doebeli and Knowlton, 1998) used a concave function for $B(I)$, namely $B(I) = B_0[1 - \exp(-B_1I)]$. They assumed that cost increases linearly, that is $C(I) = C_0I$. The parameters $B_0$, $B_1$ and $C_0$ are positive constants describing the cost–benefit relations for all individuals. It is worth investing in the interaction only if $B_0 > C_0$, and therefore this relation is assumed in the following. There are a fixed number of mutualistic interactions between two generations, and the fitness of an individual is the sum of payoffs collected in each round.

The partners, independent of which species they belong to, have an initial phenotype ($\alpha$, $\beta$), so initially all of them have the same fitness. However, slightly different mutant phenotypes can emerge by chance in both populations in every generation. If a mutant has a higher fitness than the resident type, the latter is replaced by the mutant phenotype. Invasion is thus assumed to imply fixation (Doebeli and Knowlton, 1998). Mutualism emerges if the originally very small, but positive $\alpha$ and $\beta$ phenotypic traits evolve towards higher positive values.

However, according to numerical simulations mutualism cannot evolve if the partners live in “well-mixed” populations without spatial structure (Doebeli and Knowlton, 1998). Well-mixedness means that the probability of interactions between different phenotypes is equal to the product of their relative frequencies, and is assumed here for both intraspecific competition and interspecific mutualism. This assumption makes the model more tractable, but neglects the spatial structure present in most populations. It has been known for some time that spatially structured evolutionary games behave differently and in a more complex manner than “well-mixed” models (Nowak and May, 1992). It is thus not surprising that Doebeli and Knowlton (1998) placed the individuals of interacting species on the grid points of separate lattices. An individual in the first population can interact with the individual on the same grid point of the other lattice. The fitness of each individual is given by the sum of payoffs received in the interspecific interactions, which determines the competitive success with their local neighbors within the species. The phenotype which has the highest fitness among the neighbors (including the focal individual itself) will enter the next generation at the chosen site. It has been shown numerically that mutualism can emerge in this spatially explicit system, particularly if dispersion is absent or limited, and some stochasticity is present in the competition or selection processes (Doebeli and Knowlton, 1998). One of the crucial difference between the “well-mixed” and lattice models that the successful strategies spread by growing patches in the latter case. Thus the similar (or identical) strategies interact with each other with a high probability in the spatially structured models, while to meet a similar (e.g. cooperative) strategy depends on their relative frequencies in the well-mixed systems. This “viscosity” of spatial models enhance the benefit of cooperative

---

**Table 1**

Payoff matrix when the cooperative (C) strategy invests $I > 0$ and the defective (D) does not invest into the interaction

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>C</strong></td>
<td>$R = B(I) - C(I)$</td>
<td>$S = -C(I)$</td>
</tr>
<tr>
<td><strong>D</strong></td>
<td>$T = B(I)$</td>
<td>$P = 0$</td>
</tr>
</tbody>
</table>

The matrix describes the classical prisoner’s dilemma game if $T > R > P > S$ and $2R > (T + S)$, which is valid if $B(I) > C(I)$. 

---
strategies in a similar way as kin-selection acts on the evolution of cooperation in well-mixed models (van Baalen and Rand, 1998).

Variable investment is also a fruitful concept in studying cooperation among non-related individuals that belong to the same species. The model of this problem is very similar to that introduced above, but now the partners are members of the same population. Killingback et al. (1999) considered a population where individuals live on a two-dimensional lattice and play a non-iterated prisoner’s dilemma game with variable investment. Partners interact with their eight nearest neighbors. The fitness of each individual is given by the sum of payoffs received in the interactions with all neighbors. Similar to the two species case, the phenotype which has the highest fitness among the neighbors will enter the next generation. If phenotypes are allowed to change their investment by mutations, the population evolves to cooperation in this spatially non-structured version of the Doebeli and Knowlton’s (1998) mutualism model is essentially impossible in spatially non-structured models (van Baalen and Rand, 1998). The lower the threshold investment, except at very special and biologically analagous situations, is the case that this difference indeed implies that neither the small initial offer nor the reward rate can increase by variable investment, except at very special and biologically irrelevant combinations of parameters. Evolution of mutualism is thus essentially impossible in spatially non-structured version of the Doebeli and Knowlton’s (1998) model.

We are interested in whether the initially small and positive parameters \( \alpha_i \) and \( \beta_i \) (\( i = h, s \)) can increase in the iterated investment game model of mutualism. For small values of these parameters the investment is determined by Eqs. (1), (2) are also small. Therefore it is enough to consider the linear approximations of \( C(I) \) and \( B(I) \), that is \( C(I) = C_0 I \) and \( B(I) = B_0 I \), where \( C_0 = C'(0) \) and \( B_0 = B'(0) \) are the derivatives of the functions at \( I = 0 \) (Killingback and Doebeli, 2002).

Let us assume that a small \( e \) fraction of individuals mutate to \( x'_i, \beta'_i \) phenotypes in both species. Using Eqs. (1), (2) and the linear approximations of the cost and benefit functions, we obtain a recursive equation for the payoffs of resident and mutant phenotypes for both species

\[
S_h^{(k+1)} = B_0[(1 - e_h)(x_h + \beta_h S_s^{(k)}) + e_h(x'_h + \beta'_h S'_s^{(k)})]
- C_0(x_h + \beta_h S_s^{(k)}) + e_h(x'_h + \beta'_h S'_h^{(k)})]
- C_0(x'_h + \beta'_h S'_h^{(k)})
S_s^{(k+1)} = B_0[(1 - e_s)(x_s + \beta_s S_s^{(k)}) + e_s(x'_s + \beta'_s S'_s^{(k)})]
- C_0(x_s + \beta_s S_s^{(k)}) + e_s(x'_s + \beta'_s S'_s^{(k)})]
- C_0(x'_s + \beta'_s S'_s^{(k)}).
\]

Here we assumed that an \( (x_i, \beta_i) \) strategy obtains an average benefit from its mutualistic partner proportional to the frequency of the strategies present in that population, that is, there is a “playing the field” situation in well-mixed populations (Maynard Smith, 1982). Killingback and Doebeli (2002) restricted their attention to the infinitely large population limit \( (e_i = 0) \), thus their model is generalized here to the case \( e_h, e_s \neq 0 \). If there are high numbers of interactions between two generations \( (k \gg 1) \), then \( S_h^{(k)} \) and \( S_s^{(k)} \) can be considered as the \( S_h, S_s \) fixed points of Eqs. (3). Thus, the fitness of a strategy \( (x_i, \beta_i) \) is approximately \( k S_i \). Since the fitnesses of both the resident and the mutant strategies are multiplied with the same constant \( k \), it is enough to consider the \( S_h, S_s \) fixed points in the following (Killingback and Doebeli, 2002). The fixed points can be computed from the linear recursion system (3), but they are too complex to display here (but see Appendix A). The relevant points here are that the fixed points exist, and they are asymptotically stable if \( \beta_i \) and \( \beta'_i \) are sufficiently small (Appendix B). To make the stability analysis tractable we assume here and in the following that \( e_h = e_s = e \).

Since the initial offer \( (x_i) \) and reward rate \( (\beta_i) \) are continuous variables, it is convenient to use adaptive dynamics to investigate the evolution of mutualism (Hofbauer and Sigmund, 1990; Geritz et al., 1998; Metz et al., 1992; Killingback and Doebeli, 2002). According to this framework, the dynamics of
phenotypic traits are

\[
\dot{x}_i = \frac{\partial S'_i}{\partial x_i} \bigg|_{(x_j, y_j) = (\delta x_i, \delta y_i)} , \\
\dot{y}_i = \frac{\partial S'_i}{\partial y_i} \bigg|_{(x_j, y_j) = (\delta x_i, \delta y_i)},
\]

where \(i\) denotes either the host (\(h\)) or the symbiont (\(s\)) species. Since our analysis is restricted to the invasion of rare mutants we can assume that \(\varepsilon \ll 1\), thus we can consider the linear approximation of \(S'_i\) in \(\varepsilon\) (see Appendix A). Thus we obtain a relatively simple form from Eq. (4)

\[
\dot{x}_i = \frac{1}{1 + C_0 y_i} \left( -C_0 + \frac{B_0^2 y_i}{\Gamma \varepsilon} \right),
\]

\[
\dot{y}_i = \frac{A_i}{(1 + C_0 y_i) \Gamma} \left( -C_0 + \frac{B_0^2 y_i}{\varepsilon} \right),
\]

where \(\Gamma = 1 + C_0 (\beta_s + \beta_h) + (C_h^2 - B_0^2) \beta_s \beta_h\), \(A_i = x_i [-C_0 + (B_0^2 - C_h^2) y_i] + B_0 x_i\), and \(i, j\) indexes denote \(h\) or \(s\) respectively. Mutualism emerges if the originally small \(x_i\) and \(y_i\) increase, that is if the right-hand sides of (5) are positive (Killingback and Doebeli, 2002).

It can be seen from Eqs. (5) that the originally small \(x_i\) decreases further at the infinite population size limit (\(\varepsilon = 0\)). The parameters \(\beta_i\) can increase when \(A_i / \Gamma\) is negative. Since \(\beta_i\) is small for both partners, \(\Gamma\) can be considered a positive number in this limit. However, if \(\beta_s < \frac{C_h}{B_0 - C_h} = \beta_c\) and \(x_i\) is sufficiently smaller than \(\beta_c\) then \(A_i\) can be negative. In this case \(\beta_i\) increases, until it reaches a limit which is definitely smaller than \(\beta_c\). Consequently, the initial offer \((x_i)\) decreases, the reward rates \((\beta_i)\) either increases or decreases initially, but cannot exceed the \(\beta_c\), thus there is no way for evolution of mutualism in the infinite population size limit.

Now let us analyse the case when the population is finite and mutants are rare (\(0 < \varepsilon \ll 1\)). Remember that \(\beta_h\) could not be arbitrarily close to zero to increase even in the single species cooperative model, so it is possible that reasonable small thresholds exist here as well, above which \(x_i\) and \(\beta_i\) will increase. Observe, that if \(\Gamma\) remains positive, but it is close to zero then the expressions in the parentheses of Eqs. (5) can be positive. Since we focus on the emergence of mutualistic interactions from a non-mutualistic state we assume that \(x_i\) and \(\beta_i\) are close to zero initially. Further, we assume that the initial propensity to the mutualistic interaction are roughly the same for both species \((x_0 = x_i\) and \(\beta_0 = \beta_i)\) and for the phenotypic traits \((x_i = \beta_i)\) too. Naturally, this is not sufficiently the case, but without these assumptions the analysis become hopelessly complex. To make the estimation more tractable, let us assume in the future that initially \(x_i = \beta_i = \beta\). The expressions in the parentheses of Eqs. (5) can be positive if \(0 < \Gamma \ll 1\) which relations are valid if \(\beta\) is close to but still smaller than \(\beta_{tr} = 1/(B_0 - C_0)\). If \(\beta > 1/(B_0 - C_0)\) then \(\Gamma < 0\), and thus \(x_i\) will decrease if \(\Gamma\) is a large positive number then the second term can be neglected in the parentheses of Eqs. (5), and this \(x_i\) will decrease again. To satisfy this condition we assume that \(\beta = (1 - \delta) \beta_{tr}\), where \(0 < \delta < 1\) measures the deviation from the threshold \(\beta_{tr}\). Using this notation and assuming that \(\delta^2 \ll \delta\) we conclude that \(\Gamma \approx \frac{2 \delta B_0}{B_0 - C_0 - C_h}\) in the \(\delta\) neighborhood of \(\beta_{tr}\). Substituting these values of \(\Gamma\) and \(\beta\) into the first equation of Eqs. (5) and rearranging the expression in the parentheses, we conclude that \(x_i\) increases if

\[
\delta < \frac{B_0 \varepsilon}{B_0 \varepsilon + 2 C_0} = \delta_{tr}.
\]

The same condition for \(\delta\) guarantees that the expression in the parentheses in the second equation of Eqs. (5) is positive as well, but \(y_i\) will be positive only if \(A_i > 0\) at the same time. Substituting \((1 - \delta) \beta_{tr}\) for \(\beta\) yields \(A_i = 1 - \delta + \frac{B_0 + C_h}{B_0 - C_0} (1 - 2 \delta)\), which is positive, at least if \(\delta\) is not too close to one. We thus conclude that the parameters \(x_i\) and \(y_i\) will increase if

\[
(1 - \delta) \beta_{tr} < \beta < \beta_{tr}.
\]

The evolution of mutualism is possible if both \(\beta_{tr}\) and \(\delta_{tr}\) are not very close to zero. In this case, the initially small \(x_i\) and \(\beta_i\) trait values will increase in a relatively wide interval of \(x\) and \(\beta\) (see Eqs. (6) and (7)). The parameter \(\beta_{tr}\) is not too small if \(B_0\) is not much higher than \(C_0\). However, since \(\varepsilon\) is a small number, thus \(\delta_{tr}\) is close to zero in this case. That is, \(x_i\) and \(\beta_i\) can increase only in a restricted interval of \(\beta\) around a relatively high value (see Eq. (7)). According to Eq. (6) the parameter \(\delta_{tr}\) can be relatively large if \(B_0 \gg C_0\), but then \(\beta_{tr} \ll 1\). Consequently, this analysis confirms the numerical results of Doebeli and Knowlton (Doebeli and Knowlton, 1998): that it is highly improbable that mutualistic interaction would emerge in this model system. We emphasize here that our analysis is restricted only to the large population limit (\(\varepsilon < 1\)), which could not exclude the evolution of mutualism in small populations. Similarly, our conclusion does not exclude the possibility of the emergence of mutualism in cases when the initial \(x_i\) and \(\beta_i\) values differ to each other meaningfully (e.g. \(x_0 / x_i \ll 1, \beta_0 / \beta_i \ll 1\), etc.), although these conditions seems to be biologically less relevant. On the other side, this Letter emphasizes indirectly the important role of spatial structures in the evolution of mutualism.

Acknowledgements

The author thank V. Müller and T. Killingback and the referees for helpful comments. The work is supported by the Hungarian National Science Foundation (OTKA, Grant no. T037726).
Appendix A. The fixed point

The fixed points of Eq. (3) are
\[ S_i = \frac{1}{1 + C_0 \beta_i} \frac{C_0 (x_i - x_i) + \Phi_i}{\Theta_i} \],
\[ S'_i = \frac{B_0 C_0 (1 - e)(x_i - x_i) \beta_i [e (\beta_j - \beta_j') - \beta_j (1 + C_0 \beta_j')] + \Phi_i}{\Theta_i} \]
\[
\text{(A.1)}
\]
where \( i \) denotes either the host or the symbiont. \( \Phi_i \) and \( \Theta_i \) are
\[ \Phi_i = B_0 C_0 (1 - e)(x_i - x_i) [e (\beta_j - \beta_j') - \beta_j - C_0 \beta_j'] + (1 + C_0 \beta_j') \beta_j B_0 (e - 1) \left[ C_0 x_j - B_0 x_j ((1 - e) + e x_j) \right] + (1 + C_0 \beta_j') \beta_j B_0 (e - 1) - B_0 x_j [e x_j (1 + C_0 \beta_j')]
\]
\[ \Theta_i = (1 + C_0 \beta_j') \beta_j B_0 (1 - e) e \beta_i + \beta_i (1 - e) + C_0 \beta_i' + (1 + C_0 \beta_i) B_0 C_0 (1 - e) e \beta_i + \beta_i (1 - e) + C_0 \beta_i' + (1 + C_0 \beta_i) B_0 C_0 (1 - e) e \beta_i + \beta_i (1 - e) + C_0 \beta_i') \]
\[
\text{(A.2)}
\]
where the \( i,j \) indexes denote different species.

To determine the adaptive dynamics of \( x_i \) and \( \beta_i \) we have to differentiate the fixed points within respect to these variables. This operation yields another complex expression, which can be simplified by using that \( e \ll 1 \). Considering the linear approximation of the derivatives in the function of \( e \) we arrive at Eq. (4). The calculations have been performed by the software Mathematica 3.0.

Appendix B. The stability of the fixed point

The analysis of the stability of the fixed points in the two-species case is not as simple as in the one species case (Killingback and Doebeli, 2002), but is still tractable. \( A_c \), the Jacobian matrix of (3) can be considered as \( A + \varepsilon B \), where
\[
A = \begin{pmatrix}
-C_0 \beta_h & 0 & B_0 \beta_s \\
0 & -C_0 \beta_h & B_0 \beta_s \\
B_0 \beta_h & 0 & -C_0 \beta_s \\
B_0 \beta_h & 0 & 0 & -C_0 \beta_s' \\
0 & 0 & -B_0 \beta_s & B_0 \beta_s' \\
0 & 0 & -B_0 \beta_s & B_0 \beta_s' \\
-B_0 \beta_h & B_0 \beta_h & 0 & 0 \\
-B_0 \beta_h & B_0 \beta_h & 0 & 0
\end{pmatrix}
\]
\[ B = \begin{pmatrix}
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{pmatrix}
\]

To determine the eigenvalues of the Jacobian matrix is a hopeless task, but since \( \varepsilon \ll 1 \) it can be considered as a perturbation of matrix \( A \). The eigenvalues of \( A \) are
\[
-\frac{C_0 \beta_h}{2} \frac{-C_0 (\beta_h + \beta_s) - \sqrt{4 (B_0^2 - C_0^2) + C_0^2 (\beta_h + \beta_s)^2}}{2} - C_0 \beta_s',
\]
\[
-\frac{C_0 (\beta_h + \beta_s) + \sqrt{4 (B_0^2 - C_0^2) + C_0^2 (\beta_h + \beta_s)^2}}{2} - C_0 \beta_s'.
\]

It is known that if \( A_c = A + \varepsilon B \) where \( \varepsilon \) is a small number then \( | \lambda_i(A) - \lambda_i(A_c) | < \text{cond}(S) \|B\| \varepsilon \) (Kato, 1980). The matrix \( S \) diagonalizes \( A \), that is \( S^{-1} AS = \text{diag}(\lambda_i(A)) \), \( \text{cond}(S) = \|S^{-1}\| \|S\| \) is the condition number of \( S \). Since the eigenvectors of \( A \) are not closely parallel, the condition number of \( S \) cannot be high, and therefore the estimation is not ill-conditioned (Demmel, 1997). Using the euclidean norm for \( B \) it can be shown that
\[
| \lambda_i(A) - \lambda_i(A_c) | < \text{cond}(S) \sqrt{2} B_0 \max \left( \sqrt{\beta_h^2 + \beta_s^2}, \sqrt{\beta_h^2 + \beta_s^2} \right) \varepsilon. \quad (B.1)
\]
Thus the condition of \( | \lambda_i(A) | < 1 \) gives a good estimation for \( | \lambda_i(A_c) | \) to be smaller than one, which guarantees the stability of the fixed points. It follows from simple calculations that \( | \lambda_i(A) | < 1 \) is valid if \( \beta_h < 1/C_0 \), \( \beta_s < 1/C_0 \), and \( 2 > 1 + \beta_h \beta_s (C_0^2 - B_0^2) \beta_h \beta_s \), which is true if \( \beta_h \) and \( \beta_s \) are small enough.

References


