Adaptive dynamics of cooperation may prevent the coexistence of defectors and cooperators and even cause extinction

Kalle Parvinen*
Department of Mathematics, University of Turku, Turku 20014, Finland

It has recently been demonstrated that ecological feedback mechanisms can facilitate the emergence and maintenance of cooperation in public goods interactions: the replicator dynamics of defectors and cooperators can result, for example, in the ecological coexistence of cooperators and defectors. Here we show that these results change dramatically if cooperation strategy is not fixed but instead is a continuously varying trait under natural selection. For low values of the factor with which the value of resources is multiplied before they are shared among all participants, evolution will always favour lower cooperation strategies until the population falls below an Allee threshold and goes extinct, thus evolutionary suicide occurs. For higher values of the factor, there exists a unique evolutionarily singular strategy, which is convergence stable. Because the fitness function is linear with respect to the strategy of the mutant, this singular strategy is neutral against mutant invasions. This neutrality disappears if a nonlinear functional response in receiving benefits is assumed. For strictly concave functional responses, singular strategies become invadable. Evolutionary branching, which could result in the evolutionary emergence of cooperators and defectors, can occur only with locally convex functional responses, but we illustrate that it can also result in coevolutionary extinction.

Keywords: cooperation; adaptive dynamics; evolutionarily stable strategy; evolutionary suicide; evolutionary game theory

1. INTRODUCTION
Public goods games for groups of interacting individuals have widely been used to investigate the possibility of emergence and maintenance of cooperation. In these games, individual participants can typically choose how much of their money, or resources in general, they invest into a common pool. These resources are then multiplied with the factor \( r \) and shared equally among all participants. On the population level, the best outcome is achieved when all participants invest all of their resources into the common good. This situation may, however, be vulnerable to the invasion of defectors. Let \( s \) denote the investment of a focal individual. The payoff that this focal individual receives from itself in a game with \( N \) participants is equal to \( (r/N - 1)s \). It is thus well known that in a well-mixed model, cooperation is favoured only if \( r > N \) (e.g. Hauert et al. 2006).

Several mechanisms promoting cooperation in case \( r < N \) have been proposed. If the participants of the game are genetic relatives, helping others becomes more beneficial owing to kin selection (Hamilton 1964). Even among non-related individuals, cooperation can be maintained owing to direct reciprocity if the participants are likely to meet again. Various forms of reputation and social norms (Ohtsuki & Iwasa 2006) can result in indirect reciprocity and promote cooperation. Various spatial settings without complete mixing may result in the assortment of cooperators, which can also promote cooperation.

Hauert et al. (2006) suggested that ecological feedback mechanisms can promote cooperation. In their set-up, the groups of maximum \( N \) individual participants are chosen randomly in a well-mixed population, but because the probability of finding a participant is usually less than one, the effective group size \( S \) is typically less than \( N \). If the population size is low such that the effective population size is small enough, cooperation will be favoured, also resulting in the increase in the population size, until the population size becomes so large that increasing cooperation is no longer beneficial. Hauert et al. (2006) investigated a natural extension of replicator dynamics (Hofbauer & Sigmund 1998) of defectors and cooperators with a fixed cooperative investment. Depending on parameter values and initial conditions, replicator dynamics may result in the extinction of the whole population, in the coexistence of defectors and cooperators either in a stable equilibrium or in a periodic orbit, or in the presence of cooperators without defectors (see also Hauert et al. 2008). They concluded that ecological feedback can maintain cooperation in public goods games.

In this article, we relax the assumption of fixed amounts of cooperative investment, and instead treat them as continuously varying traits under natural selection. Modelling traits as continuous variables has a long history in evolutionary biology. A real biological example of such a continuous public good is bacteria producing siderophores, which turn iron into bioavailable forms (Griffin et al. 2004; Kuemmerli et al. 2009). A mutation
has a quantitative effect on the evolving trait. If a mutant performs better than other present traits, it will become more abundant, and possibly oust the other traits. Adaptive dynamics (Metz et al. 1996; Geritz et al. 1997, 1998) provides tools for the investigation of long-term evolution of continuous strategies in realistic ecological models.

In addition to analysing continuously varying traits, we make the generalization that payoffs affect the birth rate through a functional response. In §2, we present the generalized model, together with some concepts and methods of adaptive dynamics. In §3, we investigate first the monomorphic population dynamics of the model, and then monomorphic evolutionary dynamics. As some situations cause disruptive selection, we also investigate polymorphic adaptive dynamics. The analytical proofs of various results are given in the electronic supplementary material.

2. MODEL AND METHODS

(a) Ecological public goods game

Assume that there are $k$ different strategies present in the population, and let $x_i$ be the density of individuals with cooperation strategy $s_i$, $i = 1, 2, \ldots, k$. We assume that individuals are randomly distributed in a habitat, and the population is well mixed. Furthermore, let $z$ denote the proportion of empty space, thus

$$0 \leq \sum_{i=1}^{k} x_i = 1 - z \leq 1.$$  (2.1)

Following Hauert et al. (2006), we assume that individuals are randomly chosen to form a group of at most $N$ individuals. This assumption is motivated by events where cooperation is possible occurring locally, which thus involve only a small proportion of the whole habitat. For example, warning signals about an arriving predator are observed (and also useful) only locally. The chance that an individual finds itself in a group of size $S \leq N$ is given by the chance to find $S - 1$ interaction partners, where $S - 1$ is binomially distributed, $S \sim \text{Bin}(N - 1, 1 - z)$

$$p_S(N, z) = \frac{N - 1}{S - 1} (1 - z)^{S - 1} z^{N - S}.$$  (2.2)

Provided that $S - 1$ interaction partners have been found, the numbers of partners $X_i$ with strategy $s_i$ are multinomially distributed, i.e. for $\alpha_1 + \alpha_2 + \cdots + \alpha_k = S - 1$, we have

$$P(X_1 = \alpha_1, X_2 = \alpha_2, \ldots, X_k = \alpha_k) = \frac{(S - 1)!}{\alpha_1!\alpha_2!\cdots\alpha_k!} \frac{x_1^{\alpha_1}}{1-z^{\alpha_1}} \frac{x_2^{\alpha_2}}{1-z^{\alpha_2}} \cdots \frac{x_k^{\alpha_k}}{1-z^{\alpha_k}}.$$  (2.3)

Equation (2.3) means also that the random variables $X_i$ are binomially distributed: $X_i \sim \text{Bin}(S - 1, x_i/(1 - z))$.

Following Hauert et al. (2006), we assume that the total amount of invested resources are multiplied with the factor $\rho$ and then shared equally among all partners, and this has a positive effect on the birth rate. The cooperation strategy of an individual has also a direct negative linear effect on the birth rate of itself only. Death rate $\delta$ is assumed to be constant. Generalizing the model of Hauert et al. (2006), we assume that the positive effect to the birth rate occurs through a functional response, described by the function $h(y)$. We assume that it is a positive, continuous and strictly increasing function with $h(0) = 0$. Some biological situations, such as group foraging, could also result in stepwise benefits (Bach et al. 2006). By setting $h(y) = y$, we obtain the original model by Hauert et al. (2006). With the assumptions above, for given numbers $\alpha = (\alpha_1, \ldots, \alpha_k)$ of different types of partners $s = (s_1, \ldots, s_k)$, the birth rate of an individual with cooperation strategy $s_{\text{mut}}$ is equal to

$$P_{\text{mut}}(S, s, \alpha) = \frac{h \left( \rho \left( s_{\text{mut}} + \sum_{i=1}^{k-1} \alpha_i s_i \right) / S \right)}{\sum_{i=1}^{k} \alpha_i s_i} - s_{\text{mut}}.$$  (2.4)

The average birth rate with $S - 1$ interaction partners is obtained by summing over all possible combinations of partners

$$P_{\text{mut}}(S, s, x) = \sum_{\alpha_1 + \alpha_2 + \cdots + \alpha_k = S - 1} P(X_1 = \alpha_1, \ldots, X_k = \alpha_k) \times P_{\text{mut}}(S, s, \alpha),$$  (2.5)

where $x = (x_1, \ldots, x_k)$. In the case of a linear functional response $h(y) = y$, equation (2.5) can be written as

$$P_{\text{mut}}(S, s, x) = \frac{\rho}{S} \left( s_{\text{mut}} + \frac{S - 1}{1 - z} \sum_{i=1}^{k} x_i s_i \right) - s_{\text{mut}}.$$  (2.6)

because the birth rate function (2.4) is then linear with respect to $\alpha_i$ and $s_i$. The final average payoff is obtained, when equation (2.5) is summed over the distribution of the number of interaction partners.

$$f(s_{\text{mut}}, s, x, N, z) = \sum_{S=2}^{N} p_S(N, z) P_{\text{mut}}(S, s, x).$$  (2.7)

The threshold of requiring two individuals for public goods production (Hauert et al. 2006) is the mechanism behind the Allee effect in monomorphic population dynamics (figure 1). Replicator dynamics for other threshold values has been studied by Pacheco et al. (2009). Like Hauert et al. (2006), we assume that reproduction can only occur in empty space. Therefore, the per capita population growth rate is

$$r(s_{\text{mut}}, s, x, N, z) = f(s_{\text{mut}}, s, x, N, z) z - \delta.$$  (2.8)

(b) Adaptive dynamics

Invasion fitness (Metz et al. 1992) is a central concept in adaptive dynamics. It is defined as the long-term exponential growth rate $r(s_{\text{mut}}, E_{\text{res}})$ of a rare mutant $s_{\text{mut}}$ in an environment $E_{\text{res}}$ set by the resident. If the invasion fitness of a mutant is positive, it is able to grow in population size. Therefore, it may invade and become a new resident itself. Random small mutations are assumed to happen rarely, so that the population dynamics has settled to an attractor before the next mutation happens. These mutation–invasion events result in the change of the strategy of the individuals constituting the population. These events define a trait-substitution sequence. Each element of such a sequence is either a strategy or a set of several strategies, replacing the phenotype(s) that

Figure 1. Positive equilibrium population sizes with respect to the cooperative investment strategy $s_{\text{res}}$ in case of (a) linear functional response for different values of the multiplication factor $\rho$ Holling type II functional response, for (b) different values of the functional response parameter $\alpha$, when $\rho = 4$ and (c) for different values of $\rho$ when $\alpha = 0.1$. Arrows indicate the direction of demographic change. Stable equilibria are drawn with a thick solid curve, and unstable equilibria with a thin dotted curve. Other parameters: $N = 5$, $\delta = 0.6$. (a) $h(y) = y$; (b and c) $h(y) = y/(1 + \alpha y)$.

were previously present. Although each trait-substitution sequence is a realization of a stochastic process, there are analytical methods of adaptive dynamics that tell what will happen to all possible trait-substitution sequences in the long run.

If the invasion fitness of any mutant is non-positive in the environment set by a specific resident, then this resident strategy is uninvadable, and called an evolutionarily stable strategy (Maynard Smith 1976). In this case, the resident strategy is necessarily a (local) fitness maximum and thus the selection gradient, i.e. the derivative of invasion fitness with respect to the strategy of the mutant is zero at such points,

$$\frac{\partial}{\partial s_{\text{mut}}} r(s_{\text{mut}}, E_{\text{res}})|_{s_{\text{mut}} = s_{\text{res}}} = 0,$$

and the second derivative $(\partial^2/\partial s_{\text{mut}}^2) r(s_{\text{mut}}, E_{\text{res}})|_{s_{\text{mut}} = s_{\text{res}}} < 0$ is negative. Strategies for which the selection gradient is zero are called evolutionarily singular strategies (Geritz et al. 1998). Such a singular strategy $s^*$ is convergence stable or an evolutionary attractor if the repeated invasion of nearby mutant strategies into resident strategies will lead to the convergence of resident strategies towards $s^*$ (Christiansen 1991). If an evolutionary attractor is also evolutionarily stable, it is called a continuously stable strategy (Eshel 1983), and it is a feasible final outcome of an evolutionary process. In case the second derivative $(\partial^2/\partial s_{\text{mut}}^2) r(s_{\text{mut}}, E_{\text{res}})|_{s_{\text{mut}} = s_{\text{res}}} > 0$ is positive for a monomorphically attracting singular strategy, this singular strategy is not unbeatable, and evolutionary branching (Geritz et al. 1998) may occur because of disruptive selection. This phenomenon can result in the evolutionary emergence of cooperators and defectors in an initially monomorphic population (Doebeli et al. 2004; Brännström & Dieckmann 2005).

We can illustrate the monomorphic adaptive dynamics by drawing pairwise invasibility plots (PIPs) (Matsuda 1985; Van Tienderen & De Jong 1986), where the sign of the invasion fitness is displayed in dependence on resident and mutant strategies (figures 2 and 3a). Because the resident population is at a population-dynamical equilibrium, a mutant population with the same strategy as the resident will have a neutral growth rate. Therefore, there is a zero contour at the diagonal $s_{\text{mut}} = s_{\text{res}}$ in PIPs. Evolutionarily singular strategies lie at those points where other zero contours cross the diagonal.

3. RESULTS

Hauert et al. (2006) assumed that there are two strategies present, a defector with $s = 0$ and a cooperator with fixed $s > 0$, and studied their population dynamics and possibilities of coexistence. In this article, we assume instead that the cooperation strategy is a continuously varying trait under natural selection. We will next use adaptive dynamics (Metz et al. 1996; Geritz et al. 1997, 1998) to study the evolutionary consequences of this assumption.

(a) Monomorphic population dynamics

Let us begin with a simplifying assumption that there is only one resident phenotype ($k = 1$, $x_1 = 1 - z$). First we need to find the population-dynamical attractors of the resident population. Consider a focal individual with strategy $s_{\text{mut}} = s_{\text{res}}$. Since everybody in the population has the same strategy $s_{\text{res}}$ the amount of received benefits (2.4)

$$P_{\text{res}}(S, s_{\text{res}}, 1 - z) = h\left(\rho\left(s_{\text{res}} + (S - 1)s_{\text{res}}\right)/S\right) - s_{\text{res}}$$

$$= h(\rho s_{\text{res}}) - s_{\text{res}},$$

does not depend on the number of players in the game, provided that the game is played. The game is not played if no partners are found for the focal individual, which occurs with probability $z^{N-1}$. In this model, this assumption is the mechanism behind the Allee effect (Allee et al. 1949), which means that individuals benefit from the presence of conspecifics, and these benefits are lost as population densities decline. Based on the reasoning above, the per capita growth rate (equation 2.8) of a monomorphic resident is

$$r(s_{\text{res}}, N, z) = (h(\rho s_{\text{res}}) - s_{\text{res}})(1 - z^{N-1})z - \delta.$$  

For a given $s_{\text{res}}$ the equilibrium population sizes $x = 1 - z$ are obtained by finding solutions to $r(s_{\text{res}}, N, z) = 0$, see figure 1.

Since $r(s_{\text{res}}, N, 1) = -\delta < 0$, the extinction equilibrium $z = 1, x = 1 - z = 0$ is always stable. Using the fact that $(1 - z^{N-1})z$ reaches its maximum at $z^* = N^N/(1 - N^N)$, we can show (see the electronic supplementary
material, proposition A.1) that there exist two positive equilibria if and only if
\[ (h(p_{res}) - s_{res})(N - 1)N^{N/(1-N)} - \delta > 0. \] (3.3)

Strategies, for which the left-hand side of the inequality (3.3) is equal to zero, are boundaries of viability, denoted as \( s_{\text{inf}} \). Note that the population is not viable for \( s_{res} = 0 \). The larger equilibrium is stable, and the smaller one is unstable. In the case of a linear functional response \( h(y) = y \), condition (3.3) reduces to
\[ s_{res} > \frac{\delta}{\rho - 1} \frac{N^{N/(1-N)}}{N - 1} = s_{\text{inf}}, \] (3.4)
and the value of the stable positive equilibrium is an increasing function of the resident strategy \( s_{res} \) (figure 1a), see also proposition A.2 in the electronic supplementary material. In the case of Holling type II functional response \( h(y) = y(1 + \alpha y) \), the population is viable only with intermediate cooperation strategies \( s_{res} \). The population size at the stable equilibrium reaches its maximum with \( s_{res} = (\sqrt{\rho - 1})/\alpha \rho \) (see figure 1b).

(b) Monomorphic adaptive dynamics

Assume now that the resident population with strategy \( s_{res} \) is at the stable positive equilibrium, where \( z \) is the proportion of empty space. Then investigate how an individual with strategy \( s_{\text{mut}} \) will perform in the environment set by the resident. The invasion fitness of the mutant (equation 2.8) now becomes
\[ r(s_{\text{mut}}, s_{res}, N, z) = \sum_{S=2}^{N} p_S(N, z) \left[ h\left( \frac{\rho_{\text{mut}} + (S-1)s_{\text{res}}}{S} \right) - s_{\text{mut}} \right] - \delta. \] (3.5)

The fitness gradient \( D(s_{\text{res}}) = (\partial/\partial s_{\text{mut}}) r(s_{\text{mut}}, s_{\text{res}}, N, z)_{|s_{\text{mut}}=s_{\text{res}}} \) gives the direction into which strategies are expected to change with mutations of small size. Although singular strategies \( s^{*} \) can in principle be solved from \( D(s^{*}) = 0 \), we do not obtain a simple explicit expression for them. Some general results can, however, be obtained. We can determine whether a singular strategy \( s^{*} \) is locally uninvadable from the sign of the second derivative
\[ \frac{\partial^2}{\partial s_{\text{mut}}^2} r(s_{\text{mut}}, s_{\text{res}}, N, z)_{|s_{\text{mut}}=s_{\text{res}}} = h''(s_{\text{res}}) \sum_{S=2}^{N} p_S(N, z) \left( \frac{\rho}{S} \right)^2. \] (3.6)

Since \( p_S(N, z) > 0 \), the sign of equation (3.6) is determined by the sign of the second derivative of the functional response function \( h''(s_{\text{res}}) \). The singular strategy \( s^{*} \) is thus locally uninvadable (evolutionarily stable) if the functional response function \( h \) is locally concave, \( h''(s_{\text{res}}) < 0 \). In the case of a linear functional response \( h(y) = y \), the second derivative \( h''(y) = 0 \), and invasion fitness (equation 3.5) is a linear function of the strategy of the mutant. In case \( h \) is locally convex, \( h''(s_{\text{res}}) > 0 \), the singular strategy is not locally uninvadable, and if the singular strategy is also convergence stable, evolutionary branching will occur.

As expected based on the results above, the PIP in figure 2a illustrating a case with linear functional response has a vertical contour line, and the convergence stable singular strategy \( s^{*} \approx 0.980 \) is neutral against mutant invasions. Because for Holling type II functional response \( h''(y) < 0 \), any singular strategy is uninvadable. The PIP in figure 2b has an expected shape: there is one convergence stable singular strategy \( s^{*} \approx 0.595 \), which is uninvadable, and thus an evolutionary endpoint. Also note that these results are robust against large mutational steps.

Surprisingly, these two shapes are not the only possible shapes of a PIP in the model with a linear or concave functional response. As illustrated in figure 3a, it is possible that the fitness gradient is negative for all viable strategies, and there is no singular strategy at all. In this case, whatever the strategy of the resident is, a mutant with a lower cooperation strategy has positive fitness, and can at least initially increase in population size. Therefore, the strategy of the evolving population will decrease owing to natural selection (figure 3b) until it is near the extinction boundary. But even then, a mutant with a lower strategy value has positive fitness. This mutant is, however, not viable alone, and the invasion by this ‘kamikaze’ mutant will take the population away from its stable equilibrium.
we obtain that the initial strategy of the population is near the lower boundary of viability $s_{\text{buf}}$. Note that $N^{N/(N-1)} - N > 2$ for $N \geq 3$ and $N^{N/(N-1)} - N = 2$ for $N = 2$. For the linear functional response $h(y) = y$, condition (3.8) becomes $\rho < N^{N/(N-1)} - N$. In other words, evolutionary suicide occurs for values of $\rho$ that are lower than the boundary given by equation (3.8), but still large enough that the population is viable for some strategies (figure 4).

For some types of the functional response function, it is possible, but also quite tedious, to prove that the local evolutionary properties of the lower boundary of viability also determine global evolutionary dynamics in the model. When equation (3.8) holds and thus the lower boundary of viability is locally evolutionarily attracting, it is also globally evolutionarily attracting. Therefore, there are no singular strategies and evolutionary suicide happens with any initial condition. Furthermore, when equation (3.8) does not hold, and thus the lower boundary of viability is evolutionarily repelling, there exists a unique singular strategy, which is globally evolutionarily attracting. Thus, we obtain analytically a full classification of monomorphic adaptive dynamics. The covered types include Holling type I and II, but not Holling type III functional response (Holling 1959).

**Theorem 3.1.** Assume that the functional response function is either linear ($h(y) = y$) or concave ($h''(y) \leq 0$). Furthermore, assume that with the given parameters, the population is viable for some strategies $s_{\text{res}}$ and thus there necessarily exists a lower boundary of viability $s_{\text{buf}} > 0$. Then, we have the following classification of evolutionary dynamics:

(1) If equation (3.8) holds and thus $D(s_{\text{buf}}) < 0$, the fitness gradient is negative for all viable $s_{\text{res}}$. Therefore, evolutionary suicide occurs from any initial condition (figure 3).

(2) Otherwise, there exists a unique globally attracting singular strategy. It is uninvadable if the functional response function $h$ is strictly concave (figure 2b). If it is linear $h(y) = y$, the singular strategy is neutral against mutant invasions (figure 2a).

**Proof.** See the electronic supplementary material. \[ \square \]

How much are we expected to observe cooperation in the investigated model? To answer this question, we have plotted in figure 5 the numerical values of the evolutionarily attracting singular strategies for a range of parameters. We observe that increasing the maximal group size will decrease the singular strategy or even result in evolutionary suicide. Our intuitive explanation for this is that for a given strategy, increasing the maximal group size will increase the population size $x$ at equilibrium, as can be seen from equation (3.2). In the new situation, benefits of cooperation are shared among more individuals, which is expected to select for decreased cooperation.

How about the effect of the multiplication factor $\rho$? Intuitively, one could say that increasing $\rho$ should always select for increased cooperation, because it directly increases the benefits of cooperation. However, we can see from equation (3.2) that increasing $\rho$ will also increase the population size $x$ at equilibrium, which is expected to have a negative effect on the benefits of cooperation. Depending on the relative strengths of these two effects, cooperation may either increase or decrease. As we observe from figure 5, this is indeed the case. For lowest values of $\rho$, either the evolving population is not viable or evolutionary

**Figure 3.** Evolutionary suicide: selection for too little cooperation causes the extinction of the evolving population. 
(a) PIP where the selection gradient is always negative. 
(b,c) Illustration of an evolutionary simulation. Parameters: $\rho = 2.3$, $a = 0$, $N = 5$, $\delta = 0.6$.

When does evolutionary suicide occur in the investigated model? We have already noticed that at the extinction boundary $s_{\text{buf}}$ we have $z_0 = N^{1/(1-N)}$. By substituting this into the fitness gradient

$$D(s_{\text{res}}) = z \left[ \rho h'(s_{\text{res}}) \left( \frac{1 - z^N}{(1 - z)N} - z^{N-1} \right) + z^{N-1} - 1 \right],$$

we obtain that $D(s_{\text{buf}}) < 0$ and thus evolutionary suicide occurs if and only if

$$\rho h'(s_{\text{buf}}) < N^{N/(N-1)} - N,$$

provided that the initial strategy of the population is near the lower boundary of viability $s_{\text{buf}}$. Note that $N^{N/(N-1)} - N > 2$ for $N \geq 3$ and $N^{N/(N-1)} - N = 2$ for $N = 2$. For the linear functional response $h(y) = y$, condition (3.8) becomes $\rho < N^{N/(N-1)} - N$. In other words, evolutionary suicide occurs for values of $\rho$ that are lower than the boundary given by equation (3.8), but still large enough that the population is viable for some strategies (figure 4).

For some types of the functional response function, it is possible, but also quite tedious, to prove that the local evolutionary properties of the lower boundary of viability also determine global evolutionary dynamics in the model. When equation (3.8) holds and thus the lower boundary of viability is locally evolutionarily attracting, it is also globally evolutionarily attracting. Therefore, there are no singular strategies and evolutionary suicide happens with any initial condition. Furthermore, when equation (3.8) does not hold, and thus the lower boundary of viability is evolutionarily repelling, there exists a unique singular strategy, which is globally evolutionarily attracting. Thus, we obtain analytically a full classification of monomorphic adaptive dynamics. The covered types include Holling type I and II, but not Holling type III functional response (Holling 1959).

**Theorem 3.1.** Assume that the functional response function is either linear ($h(y) = y$) or concave ($h''(y) \leq 0$). Furthermore, assume that with the given parameters, the population is viable for some strategies $s_{\text{res}}$ and thus there necessarily exists a lower boundary of viability $s_{\text{buf}} > 0$. Then, we have the following classification of evolutionary dynamics:

(1) If equation (3.8) holds and thus $D(s_{\text{buf}}) < 0$, the fitness gradient is negative for all viable $s_{\text{res}}$. Therefore, evolutionary suicide occurs from any initial condition (figure 3).

(2) Otherwise, there exists a unique globally attracting singular strategy. It is uninvadable if the functional response function $h$ is strictly concave (figure 2b). If it is linear $h(y) = y$, the singular strategy is neutral against mutant invasions (figure 2a).

**Proof.** See the electronic supplementary material. \[ \square \]

How much are we expected to observe cooperation in the investigated model? To answer this question, we have plotted in figure 5 the numerical values of the evolutionarily attracting singular strategies for a range of parameters. We observe that increasing the maximal group size will decrease the singular strategy or even result in evolutionary suicide. Our intuitive explanation for this is that for a given strategy, increasing the maximal group size will increase the population size $x$ at equilibrium, as can be seen from equation (3.2). In the new situation, benefits of cooperation are shared among more individuals, which is expected to select for decreased cooperation.

How about the effect of the multiplication factor $\rho$? Intuitively, one could say that increasing $\rho$ should always select for increased cooperation, because it directly increases the benefits of cooperation. However, we can see from equation (3.2) that increasing $\rho$ will also increase the population size $x$ at equilibrium, which is expected to have a negative effect on the benefits of cooperation. Depending on the relative strengths of these two effects, cooperation may either increase or decrease. As we observe from figure 5, this is indeed the case. For lowest values of $\rho$, either the evolving population is not viable or evolutionary
suicide occurs. When the values of \( \rho \) increase, a singular strategy appears and it first decreases, reaches a minimum and starts to increase again. The final increase is rather steep in case of linear trade-off, but the Holling type II functional response diminishes the highest benefits, and starts to increase again. The final increase is rather small. Intuitively, one could also say that increasing the costs of cooperation, and thus cooperation should decrease. Our numerical explorations suggest that mostly this is the case, but not for all parameter values. Since the population size at equilibrium also decreases, the total effect may increase cooperation (not illustrated).

**(e) Polymorphic adaptive dynamics**

So far, we have investigated only monomorphic adaptive dynamics of cooperation, which was relatively easy to investigate because all attractors of a monomorphic resident population in the investigated model are equilibria. This is not the case with a polymorphic resident: Hauert et al. (2006) observed that a dimorphic population of cooperators and defectors can coexist ecologically in an equilibrium or in a limit cycle. Therefore, we need to extend fitness calculation to cover limit cycles as well (Metz et al. 1992). Let \( x_i(t) \) denote the density of individuals with cooperation strategy \( s_i \) at time \( t \). Now the polymorphic fitness of a mutant with strategy \( s_{mut} \) is

\[
r(s_{mut}, s, x, N, z) = \lim_{T \to \infty} \frac{1}{T} \int_{t=0}^{T} f(s_{mut}, s, x(t), N, z(t)) z(t) - \delta dr.
\]

What is then the expected evolutionary outcome? Let us first look at the case when the functional response function is strictly concave, \( h'(y) < 0 \) for all \( y \geq 0 \). In this case, the singular strategy, when it exists, is evolutionarily stable. In such a case, there is no need to investigate polymorphic adaptive dynamics in detail because the general theory of adaptive dynamics tells us that selection would cause a polymorphic population to become monomorphic at least if the initial strategies are near the singular strategy. Furthermore, by differentiating equation (3.9) twice, we actually observe that under the present assumptions the polymorphic fitness function is concave with respect to the strategy of the mutant. Therefore, the invasion fitness (equation 3.9) can be zero for at most two strategy values. This means that at most, two strategies can coexist ecologically, and only mutants with intermediate strategies can invade. Therefore, any polymorphic population will become monomorphic.

In the case of a linear functional response \( h(y) = y \), by using equations (2.6) and (2.7), polymorphic invasion fitness (equation 3.9) can be written as

\[
r(s_{mut}, s, x, N, z) = s_{mut} \lim_{T \to \infty} \frac{1}{T} \sum_{S=2}^{N} \frac{1}{S} \sum_{i=1}^{k} x_i(t) \hat{s}(N, z(t))
\]

\[
\times \left[ \frac{\rho}{S} - 1 \right] z(t) dt + \lim_{T \to \infty} \frac{1}{T} \sum_{S=2}^{N} \frac{1}{S} \sum_{i=1}^{k} x_i(t) \hat{s}(N, z(t))
\]

\[
\times \left[ \frac{\rho}{S} \left( \frac{S-1}{1-z(t)} \sum_{i=1}^{k} x_i(t) \hat{s} \right) \right] z(t) - \delta dt.
\]

Figure 4. Effect of parameters \( \rho \) and \( N \) on the evolution of cooperation for linear and Holling type II functional response. Parameter values corresponding to the PIPs in figures 2 and 3 are marked with dots. (a) Linear: \( \alpha = 0 \), any \( \delta > 0 \); (b) \( \alpha = 0.05 \), \( \delta = 0.6 \).

Figure 5. Numerical values of evolutionarily attracting singular strategies in case of (a) linear and (b) Holling type II functional response, as a function of \( \rho \) for \( N = 4, 5, \ldots, 10 \). Parameter range with smaller values of \( \rho \) than the curves shown correspond to evolutionary suicide. Values of \( \rho > N \) are not realistic. (a) Linear: \( \alpha = 0 \), any \( \delta = 0.6 \); (b) \( \alpha = 0.05 \), \( \delta = 0.6 \).
In other words, not only the monomorphic invasion fitness (equation 3.5), but also the polymorphic invasion fitness (equation 3.9) is a linear function of the strategy of the mutant $s_{\text{mut}}$ when $h(y) = y$. This fact holds with any number $k$ of resident strategies and any type of resident attractor. Note that necessarily $r(s_i, s, x, N, z) = 0$ for all $i = 1, 2, \ldots, k$. Therefore, a direct consequence of the linearity of the fitness function is that $r(s_{\text{mut}}, s, x, N, z) = 0$ for all $s_{\text{mut}}$ when $k \geq 2$. In other words, any coalition of several coexisting strategies is neutral against mutant invasions. Therefore, although a monomorphic population can become polymorphic when $\rho$ is large enough, there is no disruptive selection to cause evolutionary branching. Note also that any small change in the functional response function making it either locally convex or concave will annihilate the linearity of the fitness function. Therefore, the linear case is structurally unstable.

Let us next investigate the evolutionary dynamics of cooperation in case the functional response function is locally convex, such as the Holling type III function $h(y) = y^2(1 + ay^2)$. It is convex for $y < 1/\sqrt{3a}$ and concave otherwise. Note that it is not realistic for a functional response function to be convex for all $y$. In our numerical investigations, we observed that the effect of the multiplication factor $\rho$ (figure 6a) is rather similar to the other cases investigated above: for small values of $\rho$, evolutionary suicide occurs, and when $\rho$ is increased, a convergence stable singular strategy appears. This singular strategy is, however, not evolutionarily stable, and thus evolutionary branching is expected to happen.
We indeed observed interesting new insights and scenarios as Hauert et al. (2006) expected, to the extent that they change the interpretation of the main results by Hauert et al. (2006): although eco-evolutionary feedback can in principle stabilize cooperation at intermediate frequencies of cooperators and defectors, we observed that in the original model with linear functional response, the ecological coexistence of cooperators and defectors is only neutrally stable against invasions by mutants. Furthermore, the coexistence of cooperators and defectors will not emerge starting from an initially monomorphic population because monomorphic singular strategies are neutrally evolutionarily stable. A convergence stable singular strategy should be invadable in order to observe evolutionary branching.

An even more striking observation is that an initially viable population may evolve towards lower and lower investment strategies until the population reaches a bifurcation point where a stable population-dynamical equilibrium collides with an unstable one. Even then, a mutant with a lower cooperation strategy has positive fitness and can invade. However, this invasion pushes the population below the Allee threshold (Allee et al. 1949) to extinction.

Already Hardin (1968) in his article about tragedy of the commons argued that selfish behaviour may lead to the ruin of the common good. The observed evolutionary suicide (Ferri`ere 2000) in this mathematical model is an extreme example of this phenomenon. Not only does cooperation steadily decrease in the population, but reaching a population-dynamical bifurcation point results in the extinction of the whole population. Evolutionary suicide has been observed in a wide variety of models (Matsuda & Abrams 1994; Gyllenberg & Parvinen 2001; Gyllenberg et al. 2002; Webb 2003; Parvinen 2005, 2007; Rankin & L´opez-Sepulcre 2005); see Parvinen (2005) for a review of the phenomenon and presentation of the relevant theory. We observed that evolutionary suicide will occur if the multiplication factor $\rho$ is low enough. Note that the assumption that at least two participants are required for the cooperative game to be played is the mechanism behind the Allee effect in this model. Without this, or some other assumption resulting in a discontinuous transition to extinction, evolutionary suicide cannot occur. It has namely been shown by Gyllenberg et al. (2002) that if the resident attractor gradually approaches the extinction equilibrium, when the resident strategy approaches the boundary of viability, then evolutionary suicide is not possible. But note that it is not a sufficient condition, as can be seen from figure 2a,b.

Analysis of the extension with a nonlinear functional response revealed a structural instability of the linear case. For a strictly concave functional response, any monomorphic singular strategy is properly invadable. Furthermore, at most two strategies can coexist ecologically, and only mutants with intermediate strategies can invade. Therefore, any polymorphic population will become monomorphic.

Although a small perturbation can turn a linear function to a convex one, it is not realistic for a functional response function to be globally convex. It can be locally convex, such as Holling type III functional response. If $s^*$ is a singular strategy, and the functional response is convex at $p^*$, thus $H'(p^*) > 0$, the singular strategy is invadable by nearby mutants. If the singular strategy is
a (monomorphic) evolutionary attractor, it is an evolutionary branching point. Therefore, the strategy of a monomorphic population will initially approach $s^*$, but eventually a successful mutant will not oust the resident, and the evolving population becomes dimorphic. The evolving population will thus consist of two groups, and the strategies of these groups will at least initially evolve further away from each other. Such evolutionary branching provides a mechanism for the evolutionary emergence of cooperators and defectors (Doebeli et al. 2004; Brännström & Dieckmann 2005). However, we illustrated that also such dimorphic population can experience evolutionary suicide. The existence of an evolutionary branching point does not guarantee the evolutionarily stable coexistence of cooperators and defectors.

We conclude that the analysis of ecological models of cooperation is not complete without an evolutionary analysis. Furthermore, evolutionarily stable coexistence of cooperators and defectors require some other mechanisms than those presented by Hauert et al. (2006). Ecological dynamics as such is not sufficient, but it may be possible in an ecological model with a mechanism resulting, for example, in direct or indirect reciprocity, assortment of cooperators or kin selection. Also the risk of the population evolving to extinction should be taken into account in empirical studies.

The author wishes to thank anonymous referees for constructive comments and the Academy of Finland for financial support (grant number 128323).

REFERENCES


