PROCEEDINGS OF THE ROYAL SOCIETY B

Proc. R. Soc. B (2008) **275**, 1985–1994
doi:10.1098/rspb.2007.1685
Published online 20 May 2008

Evolution of cooperation with shared costs and benefits

Joel S. Brown[1,*] **and Thomas L. Vincent**[2]

[1]*Department of Biological Sciences, University of Illinois at Chicago, Chicago, IL 60607, USA*
[2]*Department of Aerospace and Mechanical Engineering, University of Arizona, Tucson, AZ 85721-0119, USA*

The quest to determine how cooperation evolves can be based on evolutionary game theory, in spite of the fact that evolutionarily stable strategies (ESS) for most non-zero-sum games are not cooperative. We analyse the evolution of cooperation for a family of evolutionary games involving shared costs and benefits with a continuum of strategies from non-cooperation to total cooperation. This cost–benefit game allows the cooperator to share in the benefit of a cooperative act, and the recipient to be burdened with a share of the cooperator's cost. The cost–benefit game encompasses the Prisoner's Dilemma, Snowdrift game and Partial Altruism. The models produce ESS solutions of total cooperation, partial cooperation, non-cooperation and coexistence between cooperation and non-cooperation. Cooperation emerges from an interplay between the nonlinearities in the cost and benefit functions. If benefits increase at a decelerating rate and costs increase at an accelerating rate with the degree of cooperation, then the ESS has an intermediate level of cooperation. The game also exhibits non-ESS points such as unstable minima, convergent-stable minima and unstable maxima. The emergence of cooperative behaviour in this game represents enlightened self-interest, whereas non-cooperative solutions illustrate the Tragedy of the Commons. Games having either a stable maximum or a stable minimum have the property that small changes in the incentive structure (model parameter values) or culture (starting frequencies of strategies) result in correspondingly small changes in the degree of cooperation. Conversely, with unstable maxima or unstable minima, small changes in the incentive structure or culture can result in a switch from non-cooperation to total cooperation (and vice versa). These solutions identify when human or animal societies have the potential for cooperation and whether cooperation is robust or fragile.

Keywords: Darwinian dynamics; cooperation; cost–benefit game; Snowdrift game; Prisoner's Dilemma; evolutionary game theory

1. INTRODUCTION

In evolutionary ecology, game theory reveals how cooperative behaviours may emerge as adaptations, even in situations where cheating is possible (Axelrod & Hamilton 1981; Dugatkin 2006; Baalen & Jansen 2006). In economics, game theory reveals the degree to which individual stakeholders may willingly pay a cost that provides a public benefit (Bilodeau & Gravel 2004). Two advances in game theory permit us to shed novel and general insights into the evolution of cooperation and the willingness of individuals to provide public goods. The first advance imbues familiar games of cooperation and defection (such as the Prisoner's Dilemma, Game of Chicken and Tragedy of the Commons, which have strategies of complete cooperation or total defection) with costs and benefits that scale continuously with the individual's investment towards cooperation (Nowak & Sigmund 2005). Rather than being all or none, cooperation can represent a continuum of behaviours and the solutions may include some degree of cooperation (Wahl & Nowak 1999). The second advance involves the use of fitness-generating functions (*G*-functions) to represent, in a single function, all of the pay-offs resulting from all of the possible strategies and strategy combinations that may arise in a game of cooperation involving a continuum of strategies (Vincent & Brown 2005). Solutions to the game can then be investigated using both strategy dynamics and the evolutionarily stable strategies (ESS) concept. We examine two questions related to how strategies serve the individual and how they provide incidental benefits to others. First, are there convergent-stable strategies in an evolutionary sense? Second, are there strategies or sets of strategies that are resistant to invasion by rare alternative strategies? We refer to a strategy that serves the individual as well as the group as *enlightened self-interest* and the one that sacrifices group gains for gains to the individual as a *Tragedy of the Commons* (Hardin 1968). To address the above questions and outcomes, we introduce a new cost–benefit game that captures features of both enlightened self-interest and Tragedy of the Commons. A strategy determines how much cooperative effort an individual contributes. This strategy may result in costs and benefits to the individual, as well as to others. In its general form, the cost–benefit game examines several important features of animal behaviour related to altruism, cooperation, ecological engineering and niche construction. Specific forms of the cost–benefit game yield familiar games such as the Snowdrift game (Doebeli *et al.* 2004) and Prisoner's Dilemma as special cases.

2. *G*-FUNCTION

We start within the framework of a two-person symmetric matrix game with two possible strategies: 'cooperation' and 'non-cooperation' as given in the pay-off matrix.

* Author for correspondence (squirrel@uic.edu).

Received 9 December 2007
Accepted 29 April 2008

	cooperation	non-cooperation
cooperation	$(1+\alpha)b-(1+\beta)c$	$\alpha b-c$
non-cooperation	$b-\beta c$	0

This is a cost–benefit game where if neither individual cooperates, then there are no benefits and costs. However, if one person cooperates, then there are both benefits and costs distributed according to the matrix. A non-cooperative focal individual versus a cooperative opponent receives a benefit b from the opponent at a fraction $0\leq\beta\leq1$ of the opponent's cost c. A cooperating focal individual versus a non-cooperating opponent pays the full cost c and receives only a fraction $0\leq\alpha\leq1$ of the benefit b. The terms α and β introduce and scale the degree of 'selfishness' in the game (West *et al.* 2007). If both the focal individual and the opponent cooperate, then each receive their own benefit plus a fraction of the opponent's benefit at their own cost plus a fraction of the opponent's cost. To ensure that this is a game of possible cooperation, we require that $(1+\alpha)b-(1+\beta)c=(b-\beta c)+(\alpha b-c)>0$.

It is easy to show that cooperation is a global ESS so long as $\alpha b-c>0$ or $\alpha>(c/b)$, otherwise non-cooperation is the global ESS. This game does not permit an ESS with a mix of cooperators and non-cooperators. Interestingly, the conditions for cooperation to evolve correspond with Hamilton's rule (Hamilton 1963; Queller 1985) where the coefficient of relatedness must be greater than the cost–benefit ratio, and with reciprocal altruism (Trivers 1971; Axelrod & Hamilton 1981; Brown *et al.* 1982) where the coefficient of familiarity must be greater than the cost–benefit ratio. As a general result, cooperation can evolve if a sufficiently large fraction of the benefit is rebounded onto the cooperator either through the public good, non-random interactions (relatedness), or through reciprocity (Brown 2001; Nowak 2006).

We use this cost–benefit matrix game as a guide to form a continuous game of cooperation in which the individuals are able to scale the level of cooperation between total cooperation and total non-cooperation. Starting with the linear form of the cost–benefit matrix game, we add a nonlinear quadratic term to the benefits and costs. Let $0\leq u_i\leq1$ be a continuous variable describing the degree of cooperation of individual i, where $u_i=0$ represents an absence of any cooperation and $u_i=1$ represents a maximal level of cooperation. Individuals within the population are assumed to interact randomly in a pairwise fashion resulting in linear/quadratic benefit and cost functions. These assumptions lead us to the following fitness-generating function or G-function (Vincent & Brown 2005):

$$G(v,u,p)=\sum_{j=1}^{n_s}p_j\{b_2(\alpha v+u_j)^2+b_1(\alpha v+u_j)-c_2(v+\beta u_j)^2-c_1(v+\beta u_j)\},$$

where v is a virtual variable, with the property that replacing v by u_i in G results in the fitness function for a focal individual using the strategy u_i. The vector of all n_s strategies in the population is given by $\boldsymbol{u}=[u_1\ldots u_{n_s}]$, and $\boldsymbol{p}=[p_1\ldots p_{n_s}]$ is the frequency vector of players using these strategies. The fitness for an individual using strategy u_i is its expected pay-off from playing pairwise against all other players using strategies u_j including $j=i$ weighted by the probability p_j of playing someone with strategy u_j. The pay-off from an interaction has four terms. The two terms with parameters b_1 and b_2 represent the benefits bestowed from the interaction, and the two terms preceded by c_1 and c_2 represent the costs of cooperation. If the quadratic terms are removed ($b_2=0$ and $c_2=0$), then the resulting linear G-function is a continuous representation of the cost–benefit matrix game above. A method for converting a matrix game into a continuous game is given in Vincent & Brown (2005, ch. 9).

The level of cooperation by an individual's opponent using a strategy $u_j\geq0$ always contributes a benefit to the individual. We let the term $0\leq\alpha\leq1$ determine the degree to which the individual derives any direct personal benefit from its own degree of cooperation, αv. Through $b_1>0$, benefits increase linearly with the level of each individual's cooperation. Through b_2, benefits increase at either a diminishing ($b_2<0$) or an increasing ($b_2>0$) rate with the individuals' levels of cooperation. If $b_2<0$, there may be some threshold value of v above which total benefits actually decline with increasing v.

The level of cooperation by an individual, $v>0$, always incurs a cost to the individual. We let the term $0\leq\beta\leq1$ determine the degree to which the individual experiences an additional cost from its opponent's degree of cooperation, βu_j. Thus, β introduces an externality by which some of the cost of a cooperative act becomes public and unavoidable to the recipient. Through $c_1>0$ costs increase linearly with the level of each individual's cooperation. Through c_2, costs increase at either a diminishing ($c_2<0$) or an increasing ($c_2>0$) rate with the individuals' levels of cooperation. If $c_2<0$, there may be some threshold value of v above which total costs actually decline with increasing v.

While b_2 and c_2 can take on any value, they must remain sufficiently large (not too negative) to ensure that the benefits remain positive and the costs remain negative. The following constraints must be satisfied:

$$c_1>0$$
$$b_1>0$$
$$\mathcal{B}_i(v)\big|_{v=u_i}\geq0$$
$$\mathcal{C}_i(v)\big|_{v=u_i}\geq0,$$

for $i=1,\ldots,n_s$, where

$$\mathcal{B}_i(v)|_{v=u_i}=b_2(\alpha v+u_j)^2+b_1(\alpha v+u_j)$$
$$\mathcal{C}_i(v)|_{v=u_i}=c_2(v+\beta u_j)^2+c_1(v+\beta u_j).$$

and $j=1,\ldots,n_s$.

The parameters α and β determine the extent to which the benefits are public and the costs are private. Setting $\alpha=1$ and $\beta=0$ results in the Snowdrift game presented by Doebeli *et al.* (2004). This game involves two individuals clearing a road blocked by a snowdrift. Each individual benefits equally from the digging effort, but the cost to an individual depends only on one's own level of effort. In this formulation, the sum of the strategies of the two interacting individuals represents the public good and the strategy of the focal individual solely determines the private cost.

Setting $\alpha=0$ and $\beta=0$ results in a nonlinear variant of a game of altruism (Killingback & Doebeli 2002). In this case, there are no direct benefits to the focal individual from cooperating, and the costs of cooperating are entirely private. The structure of this game is similar to Prisoner's Dilemma in which an individual always benefits from the cooperation of its partner, and the individual always avoids costs by defecting. Therefore, while both individuals would be better off if their opponent cooperated rather than defected, each individual has defect (or $u=0$) as a dominating strategy. No matter what one's opponent does, one's own best strategy is to play $u=0$.

Setting $\alpha=1$ and $\beta=1$ creates a game of complete public benefits and costs. The benefits and costs to an individual are simply the sum of each individual's level of cooperation. It is this sum of cooperative behaviours that matters, not the source of the cooperative acts. Similarly, the costs of cooperation represent a complete externality. Both individuals experience the same costs regardless of which partner instigated the costs via some level of cooperative behaviour.

Setting $\alpha=0$ and $\beta=1$ represents a fourth possible combination of public versus private costs and benefits. This represents a costlier form of altruism in which an individual contributes a benefit to its partner and also imposes a public cost to oneself and the opponent.

When $\beta=0$, an individual would like an opponent to be as cooperative as possible. With $\beta=1$, there may be limits to the level of cooperation one desires from their opponent. The maximum pay-off to an opponent may occur at an intermediate level of cooperation from the cooperating individual. However, the strategy of $u=0$ still represents a dominating strategy.

All of these four games can be extended by allowing α and/or β to take on values between 0 and 1. As a last specific example, we will consider the case of $\alpha=\beta=0.5$. There is a public element to benefits and costs, but being cooperative still costs the cooperator more than the recipient, and the individual garners a smaller benefit from cooperating than the recipient.

3. CHARACTERIZING THE ESS

Maynard Smith's (Maynard-Smith & Price 1973; Maynard-Smith 1982) original ESS definition only required that a strategy, when common, be resistant to invasion by rare alternative strategies. Such a strategy need not be convergent stability and achievable through strategy dynamics. A small perturbation in mean strategy value or strategy frequency may not return an evolving system back to the unperturbed state. Likewise, a strategy that only provides convergence stability may not be resistant to invasion by rare alternative strategies. While the literature differs over whether Maynard Smith's original ESS definition should remain unchanged, we have proposed using an updated ESS definition (Vincent & Brown 1988) that requires both resistance to invasion and convergence stability (Cohen *et al.* 1999). This conforms to the definition and concept of a continuously stable strategy (CSS; Eshel 1983, 1996).

Applying these dual conditions, we examine the features of an interior, single-strategy (coalition of one) ESS for the games presented in this paper. To be resistant to invasion, the strategy u_1^* must reside on a peak of the adaptive landscape (when $p^*=1$, $v=u_1$) by satisfying the following necessary conditions:

$$\left.\frac{\partial G(v,u_1^*,p^*)}{\partial v}\right|_{v=u_1^*} = 2u_1[\alpha b_2(1+\alpha) - c_2(1+\beta)] + (\alpha b_1 - c_1) = 0, \quad (3.1)$$

$$\left.\frac{\partial^2 G(v,u_1^*,p^*)}{\partial v^2}\right|_{v=u_1^*} = 2\alpha^2 b_2 - 2c_2 < 0. \quad (3.2)$$

A necessary condition for the convergence stability of u_1 with $p^*=1$ is given by[1]

$$\left.\frac{\partial^2 G(v,u_1,p^*)}{\partial v^2} + \frac{\partial^2 G(v,u_1,p^*)}{\partial u_1\,\partial v}\right|_{v=u_1} = 2\alpha^2 b_2 - 2c_2 + 2\alpha b_2 - 2\beta c_2 < 0. \quad (3.3)$$

The candidate for an interior solution is found by solving for u_1 from equation (3.1),

$$u_1 = \frac{c_1 - \alpha b_1}{2[\alpha b_2(1+\alpha) - c_2(1+\beta)]}, \quad (3.4)$$

provided that

$$\alpha b_2(1+\alpha) - c_2(1+\beta) \neq 0$$

and that $1 > u_1 > 0$. From equation (3.4), it follows that a positive interior solution requires

$$c_1 > \alpha b_1 \quad \text{and}$$

$$\alpha b_2(1+\alpha) > c_2(1+\beta) \Rightarrow \alpha b_2 - \beta c_2 > c_2 - a^2 b_2,$$

or

$$c_1 < \alpha b_1 \quad \text{and}$$

$$\alpha b_2(1+\alpha) < c_2(1+\beta) \Rightarrow \alpha b_2 - \beta c_2 < c_2 - a^2 b_2.$$

From equation (3.2), this point is resistant to invasion (takes on a maximum) provided

$$\alpha^2 b_2 < c_2. \quad (3.5)$$

From equation (3.3), the strategy dynamics is convergent stable provided

$$\alpha b_2(1+\alpha) < c_2(1+\beta). \quad (3.6)$$

Reversing the inequalities in equations (3.5) and (3.6) provides sufficient conditions for the solution to be at a minimum point and for the solution to be unstable with respect to strategy dynamics. There are four possible outcomes for the strategy dynamics: convergent-stable minimum; convergent-stable maximum; unstable minimum; or unstable maximum. (We refer to minima or maxima that are not convergent stable as 'unstable'.) These outcomes are summarized in table 1.

Table 1. Necessary conditions for the four different types of stability at an interior solution ($S_1 = \alpha^2 b_2 - c_2$ and $S_2 = (1+\alpha)\alpha b_2 - (1+\beta)c_2$).

	maximum	minimum
convergent stable	$S_1 < 0$ $S_2 < 0$	$S_1 > 0$ $S_2 < 0$
convergent unstable	$S_1 < 0$ $S_2 > 0$	$S_1 > 0$ $S_2 > 0$

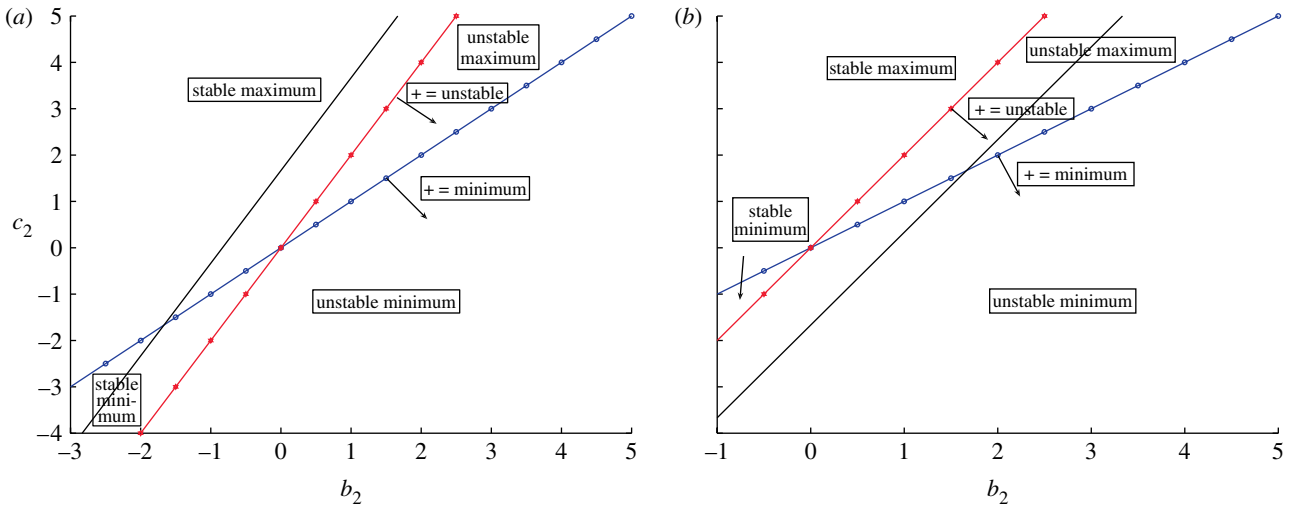

Figure 1. Nonlinearities in the cost and benefit functions affect the convergence stability and resistance to invasion of a candidate solution to the cost–benefit game. In (*a*,*b*), the *x*-axis lies on $\mathcal{C}_i(v)|_{v=u_i}=0$ and the *y*-axis lies on $\mathcal{B}_i(v)|_{v=u_i}=0$. Thus, each of these functions is positive in the regions shown, satisfying the conditions that benefits are positive and costs are negative. For a given b_2 and c_2, a candidate solution is given by equation (3.4). The line with circles ($S_1=0$) separates the space into regions where the candidate solution for u_1 is either a maximum ($S_1<0$) or a minimum ($S_1>0$) on its adaptive landscape. The line with stars ($S_2=0$) separates the space into regions where the candidate solution for u_1 is either convergent stable ($S_2<0$) or convergent unstable ($S_2>0$). Arrows indicate the direction that values for S_1 or S_2 are positive. The solid line indicates the values for b_2 and c_2 that will generate a candidate solution of $u_1=0.6$ when (i) $b_1=6$ and $c_1=4$ or (ii) $b_1=2$ and $c_1=4$. Moving from lower left to upper right along the candidate solution curve $u_1=0.6$ in (*a*), the solution shifts from being a convergent-stable minimum (not an ESS) to a convergent-stable maximum (an ESS). Moving from lower left to upper right along the candidate solution curve $u_1=0.6$ in (*b*), the solution shifts from being an unstable minimum to an unstable maximum (neither are an ESS).

4. APPLICATIONS

We use Darwinian dynamics to investigate three cases. The Darwinian dynamics in terms of *G*-functions are given by

$$\dot{p}_i = p_i\left[G(v,\boldsymbol{u},\boldsymbol{p})|_{v=u_i} - \overline{G}\,\right]$$

$$\dot{u}_i = \sigma^2 \frac{\partial G(v,\boldsymbol{u},\boldsymbol{p})}{\partial v}\bigg|_{v=u_i}$$

where

$$\overline{G} = \sum_{i=1}^{n_s} p_i G(v,\boldsymbol{u},\boldsymbol{p})|_{v=u_i},$$

n_s is the number of players, and σ^2 is a variance term associated with the spread of strategies about the mean strategies used by each of the population of players. These dynamics produce analogous changes in strategy values as the adaptive dynamics used by Doebeli & Hauert (2005) in their analysis of the continuous Snowdrift game with quadratic pay-offs. Cressman & Hofbauer (2005) used replicator dynamics to model the convergence stability of models with quadratic pay-offs. With interior solutions of $0<u<1$, they obtained the same results for evolutionary convergence as the Darwinian dynamics above.

(a) *Case 1. Snowdrift game* ($\alpha=1$ *and* $\beta=0$) *and Partial Altruism* ($\alpha=0.5$ *and* $\beta=0.5$)

The Snowdrift game and the game of Partial Altruism can produce an interior candidate for a single-strategy ESS that exhibits all the four arrangements shown in table 1. In the Snowdrift game, the benefits are public (both the cooperator and the partner benefit equally from an individual's cooperation), but the costs are private. This game analysed in Doebeli *et al.* (2004) did not analytically solve for an ESS. Here, we examine all of the possible stability outcomes identified in table 1 by letting $\alpha=1$ and $\beta=0$. In this case,

$$S_1 = b_2 - c_2,$$

$$S_2 = 2b_2 - c_2,$$

$$\mathcal{B}_i(v)|_{v=u_j} = b_2(v+u_i)^2 + b_1(v+u_i),$$

$$\mathcal{C}(v)|_{v=u_j} = c_2v^2 + c_1v$$

with an interior solution given by

$$u_1 = \frac{c_1 - b_1}{4b_2 - 2c_2}. \tag{4.1}$$

Only three of the four outcomes were identified in Doebeli *et al.* (2004). The fourth outcome is obtained by using the *G*-function to identify points that are maxima but not convergent stable.

The parameters governing the cost and benefit structure influence the outcome. In general, one must consider all the combinations of parameter values that satisfy S_1 (determines invasibility), S_2 (determines convergent stability) and the constraints $\mathcal{B}_i$ and $\mathcal{C}$. Figure 1 illustrates how these constraints divide the parameter space into different stability regions.

When the candidate solution is a convergent-stable maximum, there is just one global outcome to the game as illustrated in figure 2 (see also fig. 1*b* in Doebeli *et al.* (2004). The candidate solution is an ESS and Darwinian dynamics, starting with any number of initial strategies, will result in this ESS. This solution produces an intermediate level of cooperation manifested in all of the individuals. It is a 'glass half-full half-empty' sort of world. An observer could note success for enlightened

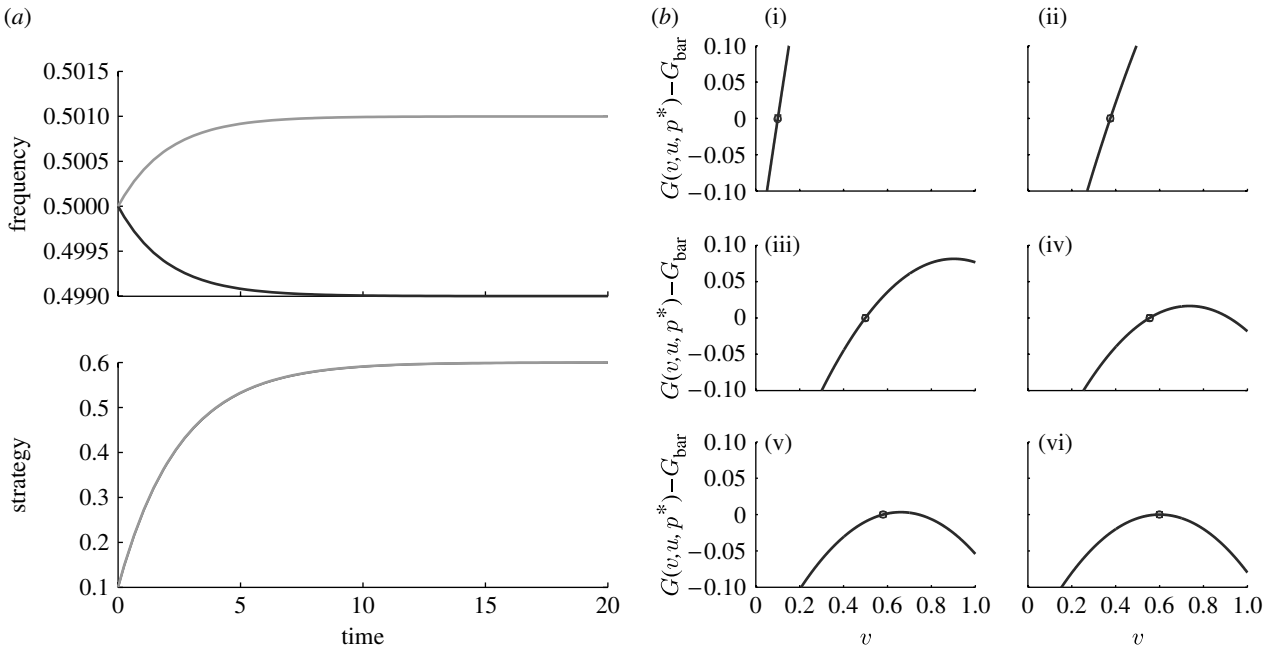

Figure 2. Depicted are the strategy dynamics and resulting ESS for a convergent-stable maximum for the Snowdrift game. (*a*) Starting with two very similar strategies ($u_1=0.1$ and $u_2=0.101$), the two strategies evolve in tandem and converge on the same value of $u_1=0.6$ and $u_2=0.6$. (*b*) Both strategies evolve towards the same maximum on the adaptive landscape. The strategy of 0.6 is an ESS, but collective fitness is not maximized at the ESS. We used $b_1=7$, $b_2=-1.5$, $c_1=4.6$ and $c_2=-1$ as in Doebeli *et al.* (2004). (i) $t=0$, (ii) $t=2$, (iii) $t=4$, (iv) $t=6$, (v) $t=8$ and (vi) $t=20$.

self-interest as each individual contributes some public goods from its selfish pursuits. Or, the observer could note elements from the Tragedy of the Commons as the game fails to produce a level of cooperation that would maximize both players' combined pay-offs.

At this ESS, by being less cooperative, one gives up more in public benefits than one gains through lower costs, and by being more cooperative one pays more in additional costs than one gains in greater benefits.

If equation (4.1) results in a convergent-stable maximum at a value of u greater than 1 or less than 0, then the actual ESS is either total cooperation ($u_1=1$) or non-cooperation ($u_1=0$), respectively (see fig. 1*d*,*e* in Doebeli *et al.* 2004). Darwinian dynamics will drive the solution to the ESS regardless of initial conditions. In summary, depending upon parameter values, a convergent-stable maximum will result in a single-strategy ESS with strategy values between 0 and 1.

If equation (4.1) results in a convergent-stable minimum, then the candidate solution is not an ESS. With just a single strategy, evolution will drive the system to a minimum of the adaptive landscape. This is an evolutionary branching point (Geritz *et al.* 1998), and the addition of another strategy, no matter how close to the first, will result in Darwinian dynamics to an ESS that has the coexistence of two distinct strategies. This outcome illustrated in figure 3 (see also fig. 1*a* in Doebeli *et al.* 2004) produces a world in which cooperators and non-cooperators both coexist. The cooperators produce public goods and the non-cooperators freeload on these public goods. The average level of cooperation in the population has similarities to the single-strategy ESS solution, but now an intermediate level of cooperation results from a mix of personality types. This solution and the game that produces it represent a form of the Producer–Scrounger game (Giraldeauand & Caraco 2000) in which some individuals produce a resource that cannot be completely defended and scroungers steal these resources from the producers.

At the ESS, both strategies experience the same pay-off, but each strategy achieves its pay-off through a different balance of costs and benefits. Those using $u_1=1$ enjoy greater average benefits but pay greater costs than those using $u_2=0$. Everyone would be better off if everyone cooperated, but such a group optimum is neither convergent stable nor resistant to invasion.

If equation (4.1) results in an evolutionarily unstable minimum, then it is not an ESS. In fact, this configuration of the game does not have a unique outcome (figure 4). Depending upon the parameters, the ESS of the game may be either total cooperation, total non-cooperation or a mix of cooperators and non-cooperators. A fixed set of parameters produces just one ESS. However, Darwinian dynamics may not necessarily drive the system to its ESS. For instance, when the ESS is an entire population of total cooperators, strategy dynamics do not necessarily result in this ESS. The strategy of $u=0$ represents a local maximum of the adaptive landscape, with a valley separating it from the cooperative strategies that would yield higher fitness. The ESS of cooperation cannot evolve continuously from total non-cooperation, and the only way for cooperation to become established is with the introduction of a small fraction of individuals with a sufficiently high level of existing cooperation. When non-cooperation is the ESS, then the reverse situation happens with regard to total cooperation being a local maximum of the adaptive landscape, but not an ESS. At this local maximum, there are distant strategies that can invade the resident strategy. When the ESS represents a mix of cooperators and non-cooperators, either extreme strategy represents a local maximum of the adaptive landscapes and Darwinian dynamics may not be able to achieve the ESS (figure 4).

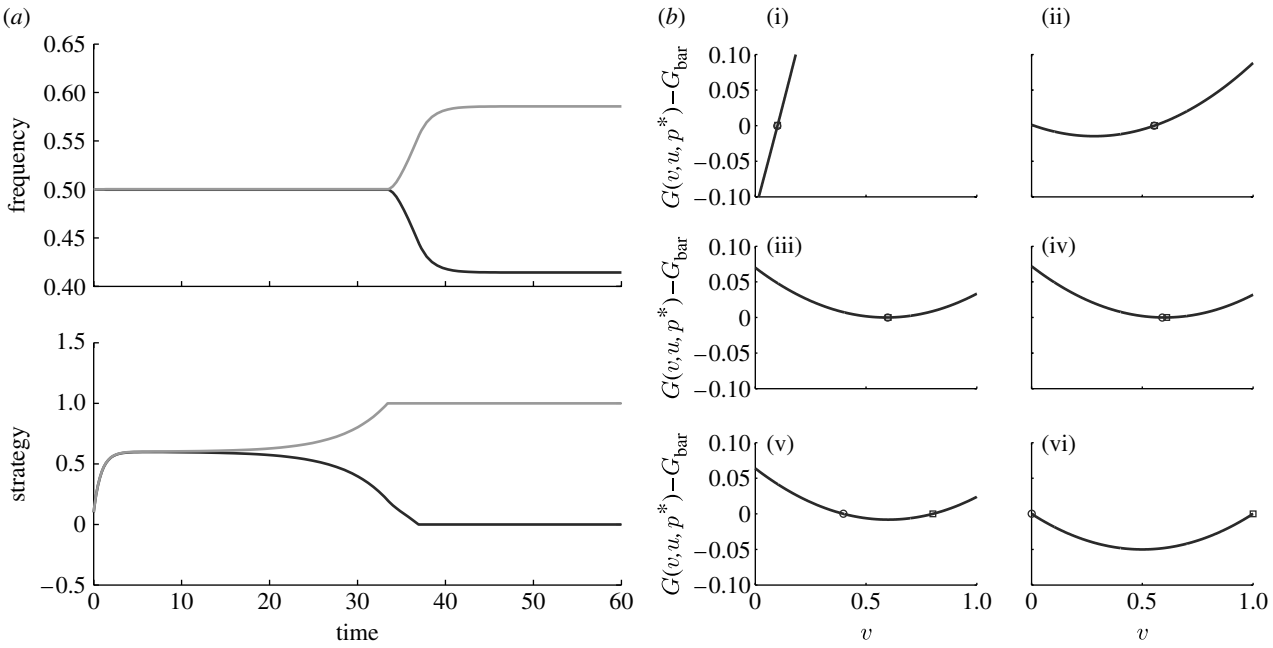

Figure 3. Depicted are the strategy dynamics and resulting ESS for a convergent-stable minimum for the Snowdrift game. (*a*) Starting with two very similar strategies ($u_1=0.1$ and $u_2=0.101$), the strategies evolve in tandem towards the convergent-stable minimum of 0.6, at which point the two strategies diverge and evolve to their ESS values of $u_1=0$ and $u_2=1$. The frequencies of the two strategies stay very close to their starting values of 0.5, with changes manifest only after the two strategies have diverged considerably from each other. (*b*) The adaptive landscape changes shape as shown at different time intervals, while the strategies first evolve towards the minimum and then diverge to the ESS. Near the convergent-stable minimum, a valley appears to the left of the strategies and then moves under and between the two strategies. With respect to collective pay-offs, the strategy $u_1=0$ is at minimum fitness and the strategy $u_2=1$ is at maximum fitness. We used $b_1=6$, $b_2=-1.4$, $c_1=4.56$ and $c_2=-1.6$ as in Doebeli *et al.* (2004). (i) $t=0$, (ii) $t=2$, (iii) $t=5$, (iv) $t=15$, (v) $t=30$ and (vi) $t=60$.

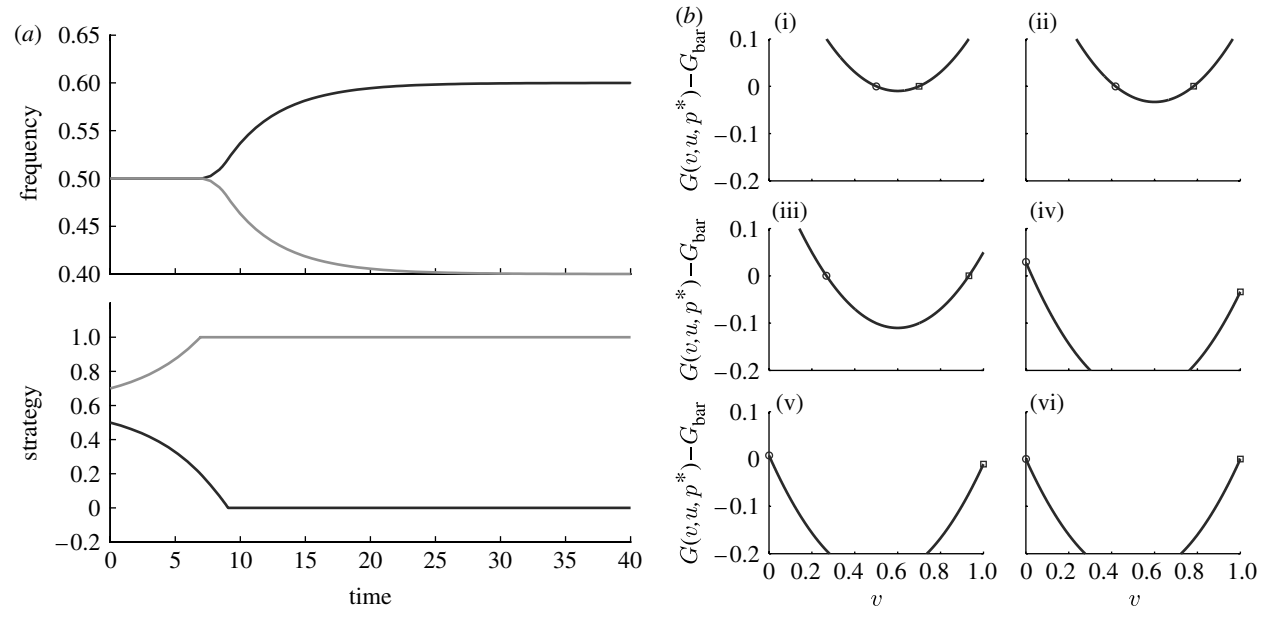

Figure 4. Depicted are the strategy dynamics and resulting ESS for an unstable minimum in the Snowdrift game. (*a*) Starting with two strategies ($u_1=0.5$ and $u_2=0.7$), the strategies immediately diverge from the unstable minimum of 0.6 Unlike the convergent-stable minimum (figure 3), the strategies do not initially evolve to the critical value of 0.6. Similar to the convergent-stable minimum, the strategies diverge and evolve to their ESS values of $u_1=0$ and $u_2=1$. Note, had the initial strategies been at $u_1=0.1$ and $u_2=0.101$, they both would have evolved to $u=0$, a non-ESS local maximum of the adaptive landscape. In fact, any two starting strategies that are to the left (or right) of $u_1=0.6$ will evolve to the non-ESS solution of $u=0$ (or $u=1$). Compare this result with the convergent-stable minimum case, where the two strategies can have very similar values and still evolve to the ESS, whereas for the unstable minimum the initial strategy values must be sufficiently different for the two strategies to evolve to the ESS. (*b*) By starting the strategy values on either side of the unstable minimum, the adaptive landscape begins with a valley between the two strategies, which allows them to evolve towards the ESS. For this unstable minimum case, we used $b_1=3.4$, $b_2=-0.5$, $c_1=4$ and $c_2=-1.5$ as in Doebeli *et al.* (2004). (i) $t=0$, (ii) $t=3$, (iii) $t=6$, (iv) $t=10$, (v) $t=15$ and (vi) $t=40$.

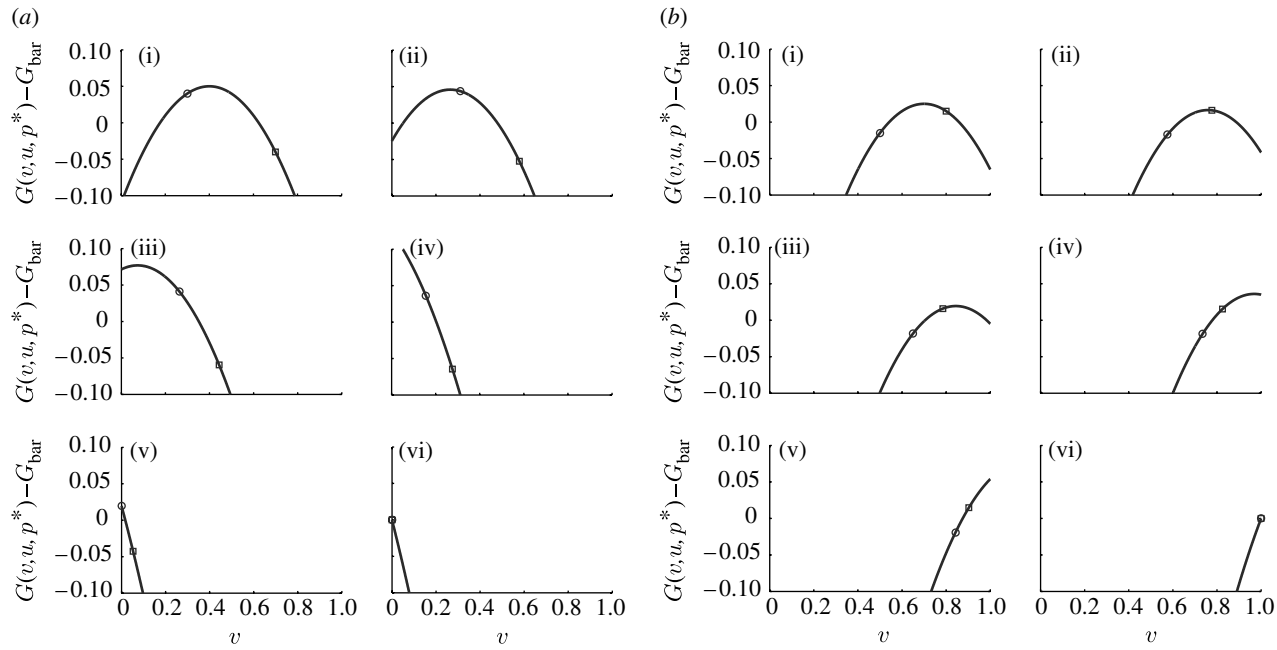

Figure 5. The Snowdrift game with an unstable maximum has two ESS configurations represented by non-cooperation or complete cooperation. (*a*) By starting with two strategies ($u_1=0.3$ and $u_2=0.7$) at values averaging less than the unstable maximum of 0.6, strategy dynamics result in convergent evolution along the adaptive landscape to the ESS of $u_1=u_2=0$. Collective fitness is minimized at this ESS of non-cooperation. (*b*) By starting with two strategies ($u_1=0.5$ and $u_2=0.8$) at values averaging greater than the unstable maximum, strategy dynamics result in convergent evolution towards the alternative ESS of $u_1=u_2=1$. At this ESS, collective fitness is maximized. For the unstable maximum, we used $b_1=2.8$, $b_2=2$, $c_1=4$ and $c_2=3$. One might wonder about the case of starting values that have an average equal to 0.6. In this case, evolution will drive both strategies to $u=0.6$; however, this solution is unstable. The slightest change in one of the strategies at this point will result in the system further evolving to one boundary or the other. (i) $t=0$, (ii) $t=2$, (iii) $t=4$, (iv) $t=6$, (v) $t=8$ and (vi) $t=50$.

An unstable minimum produces a much richer array of outcomes with respect to cooperation and non-cooperation. The outcome will always involve total cooperation or total non-cooperation. Such worlds will always represent maxima on the adaptive landscape, but they may not be an ESS. Sometimes, the ESS represents a mix of total cooperators and/or total non-cooperators that has properties identical to the ESS of the convergent-stable minimum (see fig. 1*c* in Doebeli *et al.* 2004).

Starting the system with two arbitrary strategies does not necessarily result in the evolution to the ESS. If both strategy values are sufficiently large, then the Darwinian dynamics result in $u_1=u_2=1$ and vice versa for starting values sufficiently low. Hence, when there is an intermediate value of cooperation that results in an unstable minimum, both total cooperation and total non-cooperation become local, convergent-stable maxima. These solutions are convergent stable but not resistant to invasion. Getting out of these local maxima is harder than getting out of a convergent-stable minimum (Cohen *et al.* 1999). In terms of strategies, the local maximum is locally, but not globally, resistant to invasion. The actual ESS contains a mix of strategies. If the starting conditions include strategies on either side of the unstable minimum, then strategy dynamics achieve the ESS. Even if the strategies start close to but to one side of the minimum, the ESS coalition of the two emerges from the dynamics. Thus, there is a bounded domain of attraction for the ESS in the neighbourhood of the minimum.

If equation (4.1) results in an unstable maximum, then the candidate solution is not an ESS. The ESS always involves alternate stable states in which the total cooperation is an ESS, and the total non-cooperation is an ESS (figure 5). This requires that each ESS be local with respect to starting conditions. The candidate represents a knife edge. If the system starts at the candidate solution, then it will remain there and no alternative strategies can invade. If perturbed to one side or the other with respect to the strategy value, then the Darwinian dynamics will carry the system to an ESS state of total cooperation or non-cooperation. To an observer, the cooperative ESS would be seen as a triumph of enlightened self-interest, and the non-cooperative ESS would be a total Tragedy of the Commons. To shift the population from one of these solutions to another requires a major perturbation of strategy values and frequencies.

Each ESS has its own domain of attraction. Whether the strategy evolves to a point of total non-cooperation or to the total cooperation solution depends upon the initial conditions.

Under Partial Altruism ($\alpha=0.5$ and $\beta=0.5$), the critical values for determining invasion resistance and convergence stability reduce to

$$S_1 = 0.25b_2 - c_2,$$
$$S_2 = 0.75b_2 - 1.5c_2$$

with an interior candidate solution of

$$u_1 = \frac{c_1 - 0.5b_1}{1.5c_2 - 3c_2}.$$

Depending upon parameter values, these conditions can produce all of the same cases as the Snowdrift game.

(b) *Case 2. Prisoner's Dilemma* ($\alpha=\beta=0$) *and Prisoner's Dilemma with public costs* ($\alpha=0$ *and* $\beta=1$)

Both forms of the Prisoner's Dilemma result in an ESS of total non-cooperation, $u=0$. This is to be expected. Of interest is the nature of any non-ESS solutions that involve some level of cooperation. Such solutions might permit Darwinian dynamics to produce outcomes other than the ESS of $u=0$. When $\alpha=\beta=0$, the stability conditions are given by

$$S_1 = -c_2,$$

$$S_2 = -c_2$$

with the solution for u_1 given by

$$u_1 = -\frac{c_1}{2c_2}.$$

The Prisoner's Dilemma with public costs, $\alpha=0$ and $\beta=1$, has similar stability conditions

$$S_1 = -c_2,$$

$$S_2 = -2c_2$$

with an interior solution of

$$u_1 = -\frac{c_1}{4c_2}.$$

For the Prisoner's Dilemma and the Prisoner's Dilemma with public costs, we obtain a convergent-stable maximum with $c_2>0$. With $c_2<0$ there is an unstable minimum. But, an interior solution of $u>0$ requires that c_1 and c_2 be of opposite sign. Hence, if there is an interior solution it must always be an unstable minimum. With a single strategy starting greater than this minimum, Darwinian dynamics drives the solution to $u=1$; however, this is not an ESS solution. This solution cannot be invaded by neighbouring strategies but can be invaded by other strategies (e.g. $u=0$). With starting strategies less than the minimum, Darwinian dynamics produce the ESS of non-cooperation, $u=0$.

The Prisoner's Dilemma permits just one ESS solution, $u_1=0$. It can, however, produce a non-ESS convergent-stable maximum of total cooperation depending upon the parameters and the initial conditions.

(c) *Case 3. Cooperation under public costs and benefits* ($\alpha=\beta=1$)

When costs and benefits are completely public, $\alpha=1$ and $\beta=1$ there are three possible ESS solutions of total non-cooperation, total cooperation or an intermediate level of cooperation.

The stability parameters are given by

$$S_1 = b_2 - c_2,$$

$$S_2 = 2b_2 - 2c_2$$

with the interior solution of

$$u_1 = \frac{c_1 - b_1}{4(b_2 - c_2)}.$$

The stability conditions are either both positive (unstable minimum) or both negative (convergent-stable maximum). A wide range of parameter values can produce a convergent-stable maximum with an intermediate level of cooperation. This strategy is also the ESS. Also, with a convergent-stable maximum of $u_0<1$ the ESS will be $u_1=0$, or with $u_1>1$ the ESS will be $u_1=1$. A wide range of parameter values can also produce an unstable minimum with an intermediate level of cooperation. Darwinian dynamics will drive the population to either complete cooperation or non-cooperation depending on which side of the minimum the population's strategy begins. Only one of these outcomes will actually be the ESS. The other extreme solution, such as the Prisoner's Dilemma, will be a local maximum only. This strategy can result from Darwinian dynamics, but the strategy can be invaded by strategies with levels of cooperation sufficiently distant from the local maximum.

5. DISCUSSION

Four types of interior solutions become possible for most evolutionary games based on whether a strategy is convergent stable or unstable, and whether the strategy is at a minimum or maximum of the adaptive landscape.

	$\beta=0$	$\beta=1$
$\alpha=0$	unstable minimum	unstable minimum
$\alpha=1$	stable maximum, unstable maximum, stable minimum, unstable minimum	stable maximum, unstable minimum

For the cost–benefit game, the possible types of solutions depend strongly on the extent to which the cooperator shares in the benefits (α) and the opponent shares in the costs (β). For the extreme values of α and β, these are the possible types of outcomes for solutions of $0<u<1$.

Also, when altruism is partial with $\alpha=0.5$ and $\beta=0.5$ all the four types of interior solutions are possible.

The property of directly benefiting from one's cooperative act determines whether an intermediate level of cooperation can be an ESS. If there is no direct benefit, $\alpha=0$ then complete non-cooperation is the global ESS. But, being the global ESS does not mean that total non-cooperation will evolve from Darwinian dynamics. If the strategy of the population begins with a degree of cooperation greater than the value at the unstable minimum, then complete cooperation can result even though the solution represents only a local maximum on the adaptive landscape, and individuals with sufficiently low levels of cooperation can invade.

When both benefits and costs are completely public ($\alpha=1$ and $\beta=1$), the model results in a single, global ESS. Depending upon the parameter values, this ESS can take on any value between total non-cooperation and total cooperation. Furthermore, the ESS will be achieved by Darwinian dynamics independent of the starting conditions.

The fullest range of outcomes and ESS solutions occurs when the benefits are partially or wholly public ($0<\alpha\leq1$), and the costs are not ($0\leq\beta<1$). With a stable minimum, the ESS contains the coexistence of total cooperators with total non-cooperators. Darwinian dynamics with a single starting strategy will evolve to the stable minimum at which point adaptive speciation or the invasion of another strategy permits evolution to the ESS. Both decreasing

returns to benefits and costs from increasing cooperation favour this outcome. The outcome is independent of initial conditions and small changes in parameter values will simply cause small changes in the frequency of cooperators and non-cooperators at the ESS and the position of the convergent-stable minimum.

With a convergent-stable maximum, there may be a global ESS, achievable by Darwinian dynamics at some intermediate level of cooperation. An ESS with an interior level of cooperation occurs when there are diminishing returns to benefits and increasing returns to costs. The outcome is independent of initial conditions and small changes in parameter values will cause correspondingly small changes in the level of cooperation at the ESS.

The unstable minimum and the unstable maximum permit both total non-cooperation and total cooperation to be a local ESS. Increasing returns to both costs and benefits favours an unstable maximum, and decreasing returns to both costs and benefits favours an unstable maximum. As alternative stable states, which ESS results from Darwinian dynamics depends upon the starting conditions. These scenarios under Darwinian dynamics can also result in total non-cooperation or total cooperation as a local maximum of the adaptive landscape. This arrangement allows for shifts from total non-cooperation to total cooperation with small changes in the parameter values and/or initial starting conditions.

(a) *Incentive structures*

The degree to which individuals benefit from their own cooperative acts or can foist some of the cost onto the recipient represents the incentive structure for motivating cooperation among individuals. In economics, this can come about through social contracts or government intervention where cooperative acts may be subsidized or recipients may be taxed or pay user fees for the public good. In ecology, this generally comes about from forms of ecological engineering (Jones *et al*. 1997) where an organism modifies its environment to make it more favourable, such as dam building by beavers, nest construction by woodpeckers that nest in tree cavities or the extensive burrow constructions of aardvarks.

Failure to see higher levels of cooperative behaviour can result from the incentive structure itself or from too high a frequency of non-cooperative individuals. The incentive structure represents the rules of the game itself, and the degree of cooperation among residents of the population represents a kind of culture. If the cost–benefit game has an ESS with an intermediate level of cooperation (stable maximum) or if the ESS has a mixture of totally cooperative and totally non-cooperative individuals (stable minimum), then small changes in the incentive structure will cause small changes in the degree of cooperation or the frequency of extreme cooperators, respectively. Altering the 'culture' by externally changing the initial degree of cooperation within the population will have no effect on the value of the ESS or on achieving the ESS through strategy dynamics. In a sense, these solutions are robust to incentive structure and culture.

On the other hand, when the incentive structure promotes an interior solution that is either an unstable maximum or an unstable minimum, then there are alternative stable states of either total cooperation or total non-cooperation. One of these strategies or both of them may represent the ESS of the system, but when total non-cooperation is not the ESS it is still a stable maximum that cannot be invaded by total cooperation. Seemingly small changes in the incentive structure can swing the system from total cooperation to total non-cooperation or vice versa. When humans or animal societies show drastic changes in the degree of cooperation (or aggression), it may suggest the existence of unstable maximum or unstable minimum within their social game.

(b) *Conclusion*

A matrix game form of the Prisoner's Dilemma ($\alpha=1$ and $\beta=0$ in the cost–benefit matrix game of this paper) provided a starting point for using game theory to examine the evolution of cooperation (Axelrod & Hamilton 1981; Brown *et al*. 1982). It is a game of unmitigated altruism. The altruist bestows a benefit and incurs a cost with no direct benefit to self, and the opponent receives the benefit at no cost to self. Cooperation can evolve in this game provided non-random interactions are introduced or it is played as an iterative game. Non-random interactions permit kin selection where like-minded strategies are more likely to interact with each other than by chance alone. Iterative games permit reciprocal altruism where individuals can learn to recognize, reward and punish others based on the strategies revealed by these individuals during previous plays of the game (Fletcher & Zwick 2006; Nowak 2006; Kummerli *et al*. 2007). In our cost–benefit matrix game, cooperation can also evolve so long as a sufficiently large fraction of the benefit is 'public' and directly enjoyed by the cooperator as well.

We are grateful to the comments and suggestions from two anonymous reviewers and Nils Stenseth and Christopher Whelan. This work was partially supported by a grant from the National Science Foundation to J.S.B. and Henry Howe.

ENDNOTE

[1]More comprehensive conditions for convergent stability that requires stability of both u_1 and p are available (Cohen *et al*. 1999) but are too cumbersome for ease of use.

REFERENCES

Axelrod, R. & Hamilton, W. D. 1981 The evolution of cooperation. *Science* **211**, 1390–1396. (doi:10.1126/science.7466396)

Baalen, M. V. & Jansen, V. A. A. 2006 Kinds of kindness: classifying the causes of altruism and cooperation. *J. Evol. Biol.* **19**, 1377–1379. (doi:10.1111/j.1420-9101.2006.01176.x)

Bilodeau, M. & Gravel, N. 2004 Voluntary provision of a public good and individual morality. *J. Public Econ.* **88**, 645–666. (doi:10.1016/S0047-2727(02)00178-0)

Brown, J. S. 2001 Ngongas and ecology: on having a worldview. *Oikos* **94**, 6–16. (doi:10.1034/j.1600-0706.2001.11309.x)

Brown, J. S., Sanderson, M. J. & Michod, R. E. 1982 The evolution of social behavior by reciprocation. *J. Theor. Biol.* **99**, 319–339. (doi:10.1016/0022-5193(82)90008-X)

Cohen, Y., Vincent, T. L. & Brown, J. S. 1999 A *G*-function approach fitness minima, fitness maxima, evolutionary stable strategies and adaptive landscapes. *Evol. Ecol. Res.* **1**, 923–942.

Cressman, R. & Hofbauer, J. 2005 Measure dynamics on a one-dimensional continuous trait space: theoretical foundations for adaptive dynamics. *Theor. Popul. Biol.* **67**, 47–59. (doi:10.1016/j.tpb.2004.08.001)

Doebeli, M. & Hauert, C. 2005 Models of cooperation based upon the Prisoner's Dilemma and Snowdrift game. *Ecol. Lett.* **8**, 748–766. (doi:10.1111/j.1461-0248.2005.00773.x)

Doebeli, M., Hauert, C. & Killingback, T. 2004 The evolutionary origin of cooperators and defectors. *Science* **306**, 859–862. (doi:10.1126/science.1101456)

Dugatkin, L. 2006 *The altruism equation: seven scientists search for the origins of goodness*. Princeton, NJ: Princeton University Press.

Eshel, I. 1983 Evolutionary and continuous stability. *J. Theor. Biol.* **103**, 99–111. (doi:10.1016/0022-5193(83)90201-1)

Eshel, I. 1996 On the changing concept of population stability as a reflection of a changing problematics in the quantitative theory of evolution. *J. Math. Biol.* **34**, 485–510. (doi:10.1007/BF02409747)

Fletcher, J. A. & Zwick, M. 2006 Unifying the theories of inclusive fitness and reciprocal altruism. *Am. Nat.* **168**, 252–262. (doi:10.1086/506529)

Geritz, S. A. H., Kisdi, E., Meszena, G. & Metz, J. A. J. 1998 Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evol. Ecol.* **12**, 35–57. (doi:10.1023/A:1006554906681)

Giraldeauand, L. & Caraco, T. 2000 *Social foraging theory*. Princeton, NJ: Princeton University Press.

Hamilton, W. D. 1963 The evolution of altruistic behavior. *Am. Nat.* **97**, 354–356. (doi:10.1086/497114)

Hardin, G. 1968 The Tragedy of the Commons. *Science* **162**, 1243–1248. (doi:10.1126/science.162.3859.1243)

Jones, C. G., Lawton, J. H. & Shachak, M. 1997 Positive and negative effects of organisms as physical ecosystem engineers. *Ecology* **78**, 1946–1957.

Killingback, T. & Doebeli, M. 2002 The continous Prisoner's Dilemma and the evolution of cooperation through reciprocal altruism with variable investment. *Am. Nat.* **160**, 421–438. (doi:10.1086/342070)

Kummerli, R., Colliard, C., Fiechter, N., Petitpierre, B., Russier, F. & Keller, L. 2007 Human cooperation in social dilemmas: comparing the Snowdrift game with the Prisoner's Dilemma. *Proc. R. Soc. B* **274**, 2965–2970. (doi:10.1098/rspb.2007.0793)

Maynard-Smith, J. 1982 *Evolution and the theory of games*. Cambridge, UK: Cambridge University Press.

Maynard-Smith, J. & Price, G. R. 1973 The logic of animal conflicts. *Nature* **246**, 15–18. (doi:10.1038/246015a0)

Nowak, M. A. 2006 Five rules for the evolution of cooperation. *Science* **314**, 1560–1563. (doi:10.1126/science.1133755)

Nowak, M. A. & Sigmund, K. 2005 Evolution of indirect reciprocity. *Nature* **437**, 1291–1298. (doi:10.1038/nature04131)

Queller, D. C. 1985 Kinship, reciprocity and synergism in the evolution of social behavior. *Nature* **318**, 366–367. (doi:10.1038/318366a0)

Trivers, R. L. 1971 The evolution of reciprocal altruism. *Q. Rev. Biol.* **46**, 35–57. (doi:10.1086/406755)

Vincent, T. L. & Brown, J. S. 1988 The evolution of ESS theory. *Annu. Rev. Ecol. Syst.* **19**, 423–443. (doi:10.1146/annurev.es.19.110188.002231)

Vincent, T. L. & Brown, J. S. 2005 *Evolutionary game theory, natural selection, and Darwinian dynamics*. Cambridge, UK: Cambridge University Press.

Wahl, L. M. & Nowak, M. A. 1999 The continuous Prisoner's Dilemma: I. Linear reactive strategies. *J. Theor. Biol.* **200**, 307–321. (doi:10.1006/jtbi.1999.0996)

West, S. A., Griffin, A. S. & Gardner, A. 2007 Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. *J. Evol. Biol.* **20**, 415–432. (doi:10.1111/j.1420-9101.2006.01258.x)