

## TARGET REVIEW

# The evolution of cooperation and altruism – a general framework and a classification of models

L. LEHMANN\*† &amp; L. KELLER\*

\*Department of Ecology and Evolution, University of Lausanne, Biophore, Lausanne, Switzerland

†Department of Genetics, University of Cambridge, Downing Street, Cambridge, UK

**Keywords:**

altruism;  
cooperation;  
kin selection;  
group selection;  
punishment;  
strong reciprocity.

**Abstract**

One of the enduring puzzles in biology and the social sciences is the origin and persistence of intraspecific cooperation and altruism in humans and other species. Hundreds of theoretical models have been proposed and there is much confusion about the relationship between these models. To clarify the situation, we developed a synthetic conceptual framework that delineates the conditions necessary for the evolution of altruism and cooperation. We show that at least one of the four following conditions needs to be fulfilled: direct benefits to the focal individual performing a cooperative act; direct or indirect information allowing a better than random guess about whether a given individual will behave cooperatively in repeated reciprocal interactions; preferential interactions between related individuals; and genetic correlation between genes coding for altruism and phenotypic traits that can be identified. When one or more of these conditions are met, altruism or cooperation can evolve if the cost-to-benefit ratio of altruistic and cooperative acts is greater than a threshold value. The cost-to-benefit ratio can be altered by coercion, punishment and policing which therefore act as mechanisms facilitating the evolution of altruism and cooperation. All the models proposed so far are explicitly or implicitly built on these general principles, allowing us to classify them into four general categories.

**Introduction**

When interacting individuals are related, the evolution of intraspecific cooperation and altruism (collectively referred to as helping, see later for a more formal definition) is generally studied within the framework of kin selection (Hamilton, 1964; Grafen, 1984; Taylor, 1992a; Frank, 1998; West *et al.*, 2002). By contrast, numerous theoretical models have been proposed to account for how helping can evolve when individuals are unrelated. In most cases the similarities and differences between these models and their relationship with kin selection models is obscure. In a recent paper

Sachs *et al.* (2004) proposed a useful hierarchical framework to compare models, but they did not clearly distinguish between helping behaviours that result in positive effects on the direct fitness of the actor from those that result in negative effects on the direct fitness of the actor. For instance, it remains unclear in their discussion whether the investment into helping of an individual under direct reciprocation actually increases or decreases its fitness (Sachs *et al.*, 2004, p. 139). Here we argue that such a distinction is useful because it forces one to analyse the selective forces responsible for the evolution of helping in terms of the two fundamental components of selection, i.e. direct and indirect selection (Hamilton, 1964; Grafen, 1984). This is illustrated by developing a simple conceptual framework based on the analysis of a model, which allows us to delineate the prerequisites necessary for the evolution of intraspecific altruism and cooperation.

*Correspondence:* Laurent Lehmann, Department of Ecology and Evolution, University of Lausanne, Biophore, 1015 Lausanne, Switzerland.  
Tel.: +41 21 692 4173; fax: +41 21 692 4165;  
e-mail: ll316@cam.ac.uk

The model is framed within the direct fitness approach (Taylor & Frank, 1996; Frank, 1998; Rousset & Billiard, 2000; Rousset, 2004). In the supplementary material we further develop this model to explicit the connections with classical approaches. Using this framework, we clarify the relationships between available models and categorize them into a few broad categories.

## The model

In our model we first consider a large (infinite) and unstructured (panmictic) population where individuals interact in successive rounds of pair-wise interactions (see supplementary material section for other demographic situations such as geographically structured populations). We assume that the number of rounds of interaction (1, 2, 3, ...) for each individual follows a Geometric distribution with parameter  $\omega$ , which designates the probability that an individual interacts again with a partner after a round of interaction took place (definitions of the symbols are given in Table 1). We also assume that a focal individual (FI) can interact with two classes of individuals. The first class, defined as related, consists of those individuals that have a positive probability of bearing genes identical in state with those of the FI. The second class consists of those individuals that have a lower probability of bearing such genes. The probability of interacting nonrandomly with an individual of the related class is denoted by  $x$ . With complementary probability  $1 - x$  interactions occur randomly with any member of the population. All repeated rounds of interactions take place with the same partner (see supplementary material for other situations such as indirect reciprocity). During each round of interaction the FI invests  $I_{\bullet}$  into helping with  $I_{\bullet}$  varying between 0 and 1. This investment incurs a cost  $CI_{\bullet}$  to the FI and generates a benefit  $BI_{\bullet}$ . A fraction  $\zeta$  of the benefit generated by helping directly returns to the FI and the complementary fraction  $1 - \zeta$  goes to the partner. Both the costs and the benefits are measured in terms of offspring produced. Accordingly, helping may have divergent effects on the fecundity of the FI and its partner. The effect on the FI's fecundity can be either positive or negative depending on the value of  $\zeta BI_{\bullet} - CI_{\bullet}$  while the effect on the partner's fecundity  $(1 - \zeta)BI_{\bullet}$  is always positive unless the FI gets all the benefits of its helping act (i.e.  $\zeta = 1$ ) or does not invest into helping (i.e.  $I_{\bullet} = 0$ ).

As the FI can interact with two classes of individuals who may invest differently into helping, the fecundity of the FI depends on the class of individuals with which he interacts. The relative fecundity of the FI when interacting with a class- $j$  individual is given by

$$F_{\bullet,j} = 1 + \sum_t \omega^{t-1} (B(\zeta I_{\bullet,j}(t) + (1 - \zeta)I_{j,\bullet}(t)) - CI_{\bullet,j}(t)) \quad (1)$$

(see also supplementary material, eqn 8). In this eqn, 1 designates the relative baseline fecundity of an individual,  $I_{\bullet,j}(t)$  the level of investment of the FI into helping at round  $t$  when playing against an individual of class  $j$  (i.e. a member of the class of closely related individuals or a random member of the population) and  $I_{j,\bullet}(t)$  the level of investments into helping of its partner at that round. Taking the average of the fecundities over the different classes of individuals determines the expected fecundity of the FI and the fitness of the FI is then defined as the expected number of offspring reaching adulthood (Hamilton, 1964):

$$w = \frac{x F_{\bullet,d} + (1 - x) F_{\bullet,0}}{F_0} \quad (2)$$

This is the expected fecundity of the FI relative to the expected fecundity ( $F_0$ ) of an individual randomly sampled from the population. In the fitness function,  $F_{\bullet,d}$  designates the fecundity of the FI when interacting with a closely related individual and  $F_{\bullet,0}$  is its fecundity when interacting with a random individual in the population.

To study the dynamics of investment in helping, we assume that the investment level into helping at a given round depends linearly on the partner's investment at the preceding round (Wahl & Nowak, 1999a; Killingback & Doebeli, 2002). Hence, the investment depends on three traits: the investment on the first round  $\tau$ , the response slope  $\beta$  on the partner's investment for the preceding round and the memory  $m$  (varying between zero and one) of the partner's investment at the preceding round. The variable  $m$  can be interpreted as the probability of not making an assignment error by mistakenly considering that a partner has not cooperated in the previous move when in fact he has (Ohtsuki, 2004). The two first traits ( $\tau$  and  $\beta$ ) can evolve and the dynamics of investment of the FI engaged in repeated reciprocal interactions with a partner of class  $j$  then reads

$$I_{\bullet,j}(t+1) = m\beta_{\bullet} I_{j,\bullet}(t) \text{ and } I_{j,\bullet}(t+1) = m\beta_j I_{\bullet,j}(t), \quad (3)$$

where the investments at the first round are given by  $I_{\bullet,j}(1) = \tau_{\bullet}$  and  $I_{j,\bullet}(1) = \tau_j$ .

Solving the equations of the dynamics of investments (see supplementary material) and substituting into the fitness function  $w$  (eqn 2) allows us to determine the inclusive fitness effect (Hamilton, 1964) and to establish the direction of selection on the two evolving traits  $\tau$  and  $\beta$ . In the direct fitness approach, the inclusive fitness effect is calculated by considering the effects of all 'actors' in the population (including the FI himself) on the fitness  $w$  of the FI (Taylor & Frank, 1996; Rousset, 2004). Accordingly, one counts the increment (or decrement) in the FI's fitness stemming from the expression of the behaviour of all its relatives in the population. Then, the inclusive fitness effect of the initial move  $\tau$  reads as (see supplementary material eqns 4, 13 and 14).

**Table 1** List of symbols.

|   |   |
|---|---|
| FI  | Abbreviation for focal individual   |
| $w$                                       | Fitness of a focal individual defined as its expected number of offspring reaching adulthood. It is the fecundity of the focal individual relative to the average fecundity in the population                                     |
| $l$                                       | Level of investment into helping (varying between 0 and 1)  |
| $l_{i,t}(t)$                              | Level of investment into helping at round $t$ of an individual of class $i$ engaged in repeated interactions with an individual of class $j$  |
| $l_{\bullet,j}(t)$                        | Level of investment into helping at round $t$ of the focal individual engaged in repeated interactions with an individual of class $j$  |
| $C$                                       | Fecundity cost per unit investment into helping   |
| $B$                                       | Fecundity benefit per unit investment into helping  |
| $-c$                                      | Effect of the behaviour of the focal individual on its fitness  |
| $b$                                       | Can be interpreted in two ways, either as the effect of the behaviour of the focal individual on the fitness of its related partner or as the effect of the partner when bearing the same gene as the FI on the fitness of the FI |
| $F$                                       | Total relative fecundity of an individual resulting from the repeated reciprocal interactions with its partners   |
| $\tau$                                    | Evolving level of investment into helping (varying between 0 and 1) on the first round  |
| $\beta$                                   | Evolving response slope (varying between 0 and 1) on the partner's investment at the previous round   |
| $\alpha$                                  | Evolving response slope (varying between 0 and 1) on the partner's image score  |
| $z$                                       | Generic designation of an evolving phenotype, here $\tau$ , $\beta$ or $\alpha$   |
| $z_j$                                     | Average phenotype of an individuals of category $j$   |
| $z_{\bullet}$                             | Phenotype of the focal individual   |
| $z_d$                                     | Average phenotype of an individuals of the 'related' class  |
| $z_0$                                     | Average phenotype of an individual randomly sampled from the population in a randomly mixing population or from the focal group in a geographically structured population   |
| $\zeta$                                   | Proportion of the benefits generated by a helping act that directly return to the focal individual  |
| $\omega$                                  | Probability that an individual interacts again with a partner once an interaction took place  |
| $m$                                       | Probability that an individual knows the investment into helping of its partner at the previous move  |
| $q$                                       | Probability that an individual knows the image score of its partner   |
| $x$                                       | Probability that an individual interacts nonrandomly with an individual of the related class  |
| $a$                                       | Probability that a individual interacts nonrandomly with another individual that bears the same genes at the altruistic locus   |
| $Q_j$                                     | Probability of genetic identity between pairs of homologous genes, one sampled from the FI and the other from a category $j$ member   |
| $Q_{\bullet}$                             | Probability of genetic identity between two randomly sampled homologous genes in the FI. In haploid organisms $Q_{\bullet} = 1$   |
| $Q_d$                                     | Probability of genetic identity between one gene sampled in the FI and another one sampled from the related class of individuals  |
| $Q_0$                                     | Probability of genetic identity between one gene sampled in the FI and another one randomly sampled in the population but excluding the FI  |
| $r = \frac{Q_d - Q_0}{Q_{\bullet} - Q_0}$ | Coefficient of relatedness, which is a ratio of difference of probabilities of genetic identity   |
| $rb - c > 0$                              | Hamilton's rule   |
| $N$                                       | Group size  |
| $n_d$                                     | Number of groups (demes) in the population  |
| $d_i$                                     | Dispersal probability at distance $i$   |
| $d_0$                                     | Probability of staying in the natal patch ( $d_0 = 1 - \sum_{i=1}^{n_d-1} d_i$ )  |
| $s$                                       | Survival probability of an adult to the next generation   |
| $k = \frac{N_l}{N_h}$                     | Relative population size, where $N_l$ is the number of individuals in a deme of low density and $N_h$ is the number of individuals in a deme of high density  |

$$\Delta W_{IF} = \frac{\zeta B + (1 - \zeta)\omega m \beta B - C}{\underbrace{(1 + (B - C) - \omega m \beta)(1 + \omega m \beta)}_{-c}} + r \frac{x[(1 - \zeta)B + \omega m \beta(\zeta B - C)]}{\underbrace{(1 + (B - C) - \omega m \beta)(1 + \omega m \beta)}_b} \quad (4)$$

As in Hamilton's (1964) framework, the inclusive fitness effect is broken down into the fitness cost of helping for the FI ( $-c$ ) and the fitness benefits ( $b$ ) provided to partners of the related class multiplied by the coefficient of relatedness ( $r$ ) between the FI and individuals of that class. The trait spreads when the inclusive fitness effect is positive, that is when Hamilton's rule  $rb - c > 0$  is satisfied. Following Hamilton's and

Rousset's terminology, we categorize as cooperation those cases where the act of helping is associated with an increase in the FI's direct fitness (i.e. when  $-c > 0$ ) and as altruism cases where helping is associated with a decrease in the FI's direct fitness (i.e.  $-c < 0$ ). As we shall see later, using cooperation and altruism as originally defined by Hamilton is important because the different conditions are required for the evolution of helping when it results in positive vs. negative effects on the direct fitness of the FI.

Because the inclusive fitness effect of the response slope is proportional to eqn (4) (see supplementary material eqns 13–14 and eqns 15–16) the same conditions must be satisfied for the inclusive fitness effect of  $\tau$  and  $\beta$  to be positive, i.e.

**Table 2** Classification of selective pressures promoting helping.

|  |  |  |  |
|--|--|--|--|
| Helping<br>$rb - c > 0$<br>$\underbrace{\zeta B + (1 - \zeta)\omega m\beta B - C}_{-c} + r x \underbrace{[(1 - \zeta)B + \omega m\beta(\zeta B - C)]}_b > 0 \text{ (eqn 8)}$ |  |  |  |
| Cooperation<br>$-c > 0$  |  | Altruism<br>$-c < 0$   |  |
| Direct benefits<br>$\omega = 0$ and $r = 0$<br>$\underbrace{\zeta B - C}_{-c} > 0 \text{ (eqn 6)}$   | Reciprocation<br>$\zeta = 0$ and $r = 0$<br>Direct reciprocity<br>$\underbrace{\omega m\beta B - C}_{-c} > 0 \text{ (eqn 7)}$<br>Indirect reciprocity<br>$\underbrace{\omega q\alpha B - C}_{-c} > 0 \text{ (eqn 26 of the appendix)}$ | Kin selection<br>$\omega = 0$ and $\zeta = 0$<br>$r x \underbrace{B}_b - \underbrace{C}_c > 0 \text{ (eqn 8)}$ | Greenbeard<br>$\omega = 0$ and $\zeta = 0$<br>$\underbrace{aB}_b - \underbrace{C}_c > 0 \text{ (eqn 9) here } r = 1$ |

The various situations encapsulated in eqn 8 of the main text and shown in Table 2 are not mutually exclusive, but we consider them separately for simplicity.

$$\underbrace{\zeta B + (1 - \zeta)\omega m\beta B - C}_{-c} + r x \underbrace{[(1 - \zeta)B + \omega m\beta(\zeta B - C)]}_b > 0 \quad (5)$$

Inequality 5 allows us to delineate the different conditions where cooperation and altruism are favoured. Table 2 summarizes these conditions that will now be discussed in detail in the next sections.

**The evolution of cooperation**

There are two general situations where helping can evolve and the act is cooperative [i.e. in eqn (5) is satisfied and  $-c > 0$ ]. The first is when the FI gets some direct benefit (i.e.  $\zeta > 0$ ) from its investment in helping. The other is when the FI benefits indirectly from repeated interactions with a partner who also invests in helping (i.e.  $\omega m\beta > 0$ ). Although both situations are not mutually exclusive, we shall consider them separately for simplicity.

**Direct benefits**

When  $\zeta > 0$  helping can evolve even in the absence of discrimination between more and less related individuals ( $x = 0$ ) and in the absence of repeated interactions ( $\omega = 0$ ) when the inequality

$$\zeta B - C > 0 \quad (6)$$

is satisfied. Helping is cooperative because the action results in increased fitness for both the FI (by  $\zeta B - C$ ) and its partner (by  $(1 - \zeta)B$ ). A similar result can be obtained from models, which consider a situation where unrelated individuals in a group equally share the benefits of a cooperative act (Uyenoyama & Feldman, 1980; Nunney, 1985). In that case  $\zeta$  is equal to  $1/N$  where  $N$  is the number of individuals in the group. Clearly, small group size facilitates the evolution of cooperation when the

benefits are equally shared between group members. It is important to note that when  $\zeta B - C > 0$  is satisfied helping evolves simply because the FI increases its direct fitness by performing such an act. This situation has also been previously referred to as weak altruism (Wilson, 1979; Sachs *et al.*, 2004) or by-product mutualism (Brown, 1983). More recently it has been discussed under the heading of ‘snowdrift game’ (Hauert, 2004).

There are several situations under which helping generates direct benefits for the FI (i.e.  $\zeta > 0$ ). A classical case is when individuals invest in communal activities such as nest defence, nest building and group hunting. While the benefits of such cooperation are usually shared equally among all individuals in the group, the value of  $\zeta$  will also depend on the cooperative behaviour of other group members when there are synergistic effects of cooperation (Queller, 1985). The selective pressure on helping is also expected to be high when the fitness of an individual critically depends on its investment in cooperation, for example if helping significantly increases survival (Eshel & Shaked, 2001) or the chance of inheriting a territory.

**Repeated interactions and information**

When individuals interact repeatedly (i.e.  $\omega > 0$ ) helping can evolve even if the FI gets no direct benefits from its investment in helping ( $\zeta = 0$ ) and in the absence of discrimination between more and less related individuals ( $x = 0$ ) when the inequality

$$\omega m\beta B - C > 0 \quad (7)$$

is satisfied. In this case, helping is again cooperative because  $-c > 0$ . Interestingly, when repeated interactions occur with certainty ( $\omega = 1$ ) and individuals have a perfect memory ( $m = 1$ ), in eqn (7) reduces to the threshold theorem of Killingback & Doebeli (2002) when individuals do not take into account their own investment in the previous move while interacting with their

partner. This result emphasizes that cooperation can spread only if interacting individuals have an initial tendency to be cooperative (i.e.  $\beta > 0$ ). At the other extreme when  $\omega = 0$  cooperation can never evolve. In order for cooperation to evolve, there must be a minimal probability to interact again with the same partner and this probability must be greater when the ratio  $C/B$  is small (Friedman, 1971; Trivers, 1971; Axelrod & Hamilton, 1981).

Several mechanisms may lead to  $m$  greater than zero. It is common knowledge that humans have strong capacity to keep track of the nature of their previous interactions with partners as well as detecting cheating (Fehr & Fischbacher, 2003). Experimental studies also revealed that humans are more likely to cooperate with individuals that have been cooperative in previous interactions (Fehr & Fischbacher, 2003). These are the required conditions for cooperation to evolve by direct reciprocity. While direct reciprocity is certainly an important force underlying altruism in humans, its role in other organisms is highly debated and probably of low significance (Hammerstein, 2003; Stevens, 2004). One of the reasons for this difference lies in the higher cognitive abilities of humans that allows for a much higher  $m$ -value than in other organisms. Good memory is for example crucial in 'negotiation games' where players exchange offers back and forth in a negotiation phase until they converge to a final pair of contributions (Taylor & Day, 2004).

In addition to the memory of a partner's previous moves, information on whether a given individual is likely to be cooperative may come from its reputation (Nowak & Sigmund, 1998). Here individuals have some information on the overall level of cooperative tendency of individuals they randomly meet for an interaction. Accordingly, they can adjust their investment in cooperation based on the reputation of their partner and cooperation can evolve by indirect reciprocity (Nowak & Sigmund, 1998). The difference between direct and indirect reciprocity lies in the mechanism underlying the evaluation of the cooperative tendency of the partner. In the supplementary material, we derive a model for the evolution of helping in the presence of indirect reciprocity where reputation of the partner depends on its image score and where assignment errors can occur. The condition for the evolution of helping by image score is then similar to that in eqn (7) with the only difference that  $m$  describes the probability of correctly assessing the partner's reputation [i.e. likelihood to know its social score, which is designated by  $q$  in Nowak & Sigmund (1998), see Table 1 and eqn 26 in the supplementary material]. In other words, the main difference between direct and indirect reciprocity lies in the source of information rather than a difference in the type of selective force involved.

Whether or not repeated interaction leads to stable cooperation is still unclear. Two cases can be distinguished. The first is when no errors occur in the

implementation of helping (i.e.  $m = 1$ ). In that case the initial move and the slope converge respectively towards  $\tau = 1$  and  $\beta = 1$  (Wahl & Nowak, 1999a). The optimal strategy is thus to be generous on the first move because it elicits cooperation in return. While simulations suggest that such a strategy is stable and immune to the invasion of cheaters (Wahl & Nowak, 1999a; Killingback & Doebeli, 2002), analytical work seems to indicate that this may not be the case when players interact long enough (Lorberbaum, 1994). The second situation is when errors occur. While it has been suggested that direct reciprocity can then be stable (Lorberbaum *et al.*, 2002), this seems not to be generally the case. For instance, in the direct reciprocity setting of Wahl & Nowak (1999b), discriminator cooperative strategies can invade defectors but when discriminator cooperative strategies have reached a high frequency, nondiscriminative cooperative strategies may emerge. This, in turn, enables defectors to invade, resulting in a population that cycles between cooperation and defection. The same conclusion holds for indirect reciprocity when reputation through image scores is based on individuals past actions (Nowak & Sigmund, 1998). By contrast, sustained cooperation over time seems possible under indirect reciprocity when specific assumptions are made on the distribution of the number of rounds of interactions (Brandt & Sigmund, 2004) or when reputation is modelled as standing, where an individual's standing is not negatively affected by refusing to provide help to partners in bad standing (Panchanathan & Boyd, 2003). It is not clear why these different conditions lead to such contrasting results, and more generally, whether cooperation can be stable with imperfect memory and a limited number of interactions as is the case for most natural systems.

## The evolution of altruism

When  $-c < 0$  (i.e. helping is altruistic as it is associated with a decrease in the FI's direct fitness but an increase in the direct fitness of individuals receiving help), in eqn (4) can only be satisfied when there are different kin classes in the population and helping is preferentially directed toward individuals of the related class (i.e.  $x > 0$ ). In the following sections we will differentiate two situations that differ depending on whether the kin classes are defined on the basis of the average genetic similarity over the whole genome (genetic relatedness) or similarity at particular loci (greenbeard effect).

## Preferential interactions and helping between kin

When  $xr > 0$ , helping can evolve even if the FI gets no direct benefits from its investment in helping ( $\zeta = 0$ ) and when there is no repeated interactions between individuals (i.e.  $\omega = 0$ ) when the inequality

$$xrB - C > 0 \quad (8)$$

is satisfied. When  $x = 1$  (i.e. perfect discrimination between more and less related partners), the inequality simplifies to  $rB - C > 0$ , which is the condition for the spread of helping when altruistic acts are directed only toward relatives. Because competition occurs at random in the population, this situation represents the family-structured model as originally envisioned by Hamilton (1964). In eqn (8),  $xr$  measures the extent to which individuals are more related in altruistic than in competitive interactions, in line with the view that individuals must be more related in altruistic than in competitive interactions for helping to evolve when it results in a net fecundity cost (Queller, 1994). Inversely, when helping is provided irrespective of relatedness, helping cannot be selected for unless it results in a direct fecundity benefit, a result which usually holds whatever the genetic structure of the population (Taylor, 1992b; Rousset, 2004) (see supplementary material).

Several mechanisms may generate an  $x$  greater than 0. The most common in nature is probably the use of spatial cues with individuals expressing conditional altruism in the natal nest or colony. Indeed most of the extreme cases of altruism are found within families such as in social insects (Keller & Chapuisat, 2001). A more active and refined mechanism is phenotype-matching, with individuals being able to actively estimate their genetic similarity by comparing their own phenotypic characteristics with those of other individuals (Reeve, 1989). As common genealogy generates phenotypic similarity for genetically determined traits, each trait can be used as an independent value to estimate average genetic identity. This is a process of statistical inference with arbitrary phenotypic traits being used as quantitative or qualitative variables. Importantly, both spatial recognition and phenotype matching lead to uniform genetic similarity over the whole genome. Hamilton's rule is then broadly satisfied and there is no intragenomic conflict. In other words, altruism is stable and immune to cheating (Seeger, 1993). However, deception may occur when individuals can circumvent the recognition mechanism. This may occur when individuals succeed in infiltrating a foreign family. An excellent example of this is social parasitism in ants, where queens enter foreign established colonies and secure help from the resident workers to raise their brood of reproductive individuals. Importantly, however, these cases of parasitism are expected to be relatively rare because frequency-dependent selection on the recognition system of the hosts should maintain the rate of parasitism under check (Reeve, 1989; Axelrod *et al.*, 2004).

#### *Greenbeard effect*

The other possible mechanism leading to altruism is when preferential interactions between the FI and related individuals at the helping loci are mediated by a linkage disequilibrium between the gene encoding a

phenotypic trait used for recognition and the gene(s) responsible for helping. Imagine the simple case of two genes, one causing a specific phenotypic effect and the other determining the level of helping and allowing its bearers to determine whether or not other individuals exhibit a specific phenotype expressed by the first gene. Whether or not helping may evolve will depend on the linkage disequilibrium between these two genes. In case of perfect linkage, the situation is that of a greenbeard gene, a concept invented by Hamilton (1964) and named by Dawkins (1976). A greenbeard gene is defined as a gene that causes a phenotypic effect (e.g. the presence of a greenbeard or any other conspicuous feature) that allows the bearer of this feature to recognize it in other individuals, and results in the bearer to behave differently toward other individuals depending on whether or not they possess the feature. If a haploid greenbearded individual has a probability  $a$  to correctly identify and preferentially interact with another greenbearded individual investment into helping is selected when

$$aB - C > 0 \quad (9)$$

is satisfied and one recovers the conditions described by Hamilton (1975). Importantly, this inequality is similar to in eqn (8), the parameter  $a$  being equivalent to  $x$ . The coefficient of relatedness is equal to one here because the probability of genetic identity at the altruistic locus is one. This situation of preferential interactions between individuals sharing the same altruistic gene is also sometimes referred to as 'assortative meeting' models (Eshel & Cavalli-Sforza, 1982). If recombination can break down the linkage between 'recognition' and 'altruistic' genes, the situation becomes quite different because altruism becomes intrinsically unstable. This is because individuals with the gene conferring the greenbeard phenotype but without the gene coding altruism will have greatest fitness and there will be a rapid decrease in frequency of the altruism gene. In contrast, if the recognition and altruistic effects are the product of a single gene or two completely linked genes, a breakdown of the system can occur only after the evolution of a new gene, which confers the greenbeard but not response effect. In other words, greenbeard systems should essentially be unstable over evolutionary time, with rapid collapse if there are two genes and recombination and significantly slower collapse when the greenbeard and response effect are the product of a single gene or two genes without recombination.

#### **Cost and benefit of helping**

In the previous sections we highlighted four situations conducive to the evolution of helping. For each of these situations, the condition required for helping to be favoured is directly dependent on the cost to benefit ratio ( $C/B$ ) of this behaviour. The importance of this ratio has been repeatedly recognized. For example, both the

role of ecological factors and species-specific idiosyncratic characteristics, which benefit altruism are all important in promoting the evolution of reproductive altruism in social insects. Thus, it has been suggested that the presence of a sting and the raising of brood in a complex nest are preadaptations responsible for the disproportionate number of eusocial evolution in Hymenoptera (Seger, 1993). Similarly, living in a relatively invariable and warm climate coupled with low annual mortality possibly predisposes certain taxonomic lineages of birds to cooperative breeding (Arnold & Owens, 1999).

Another central issue that has received increased attention over the last decade is that the costs and benefits of helping are not fixed variables since other group members can actively alter them. This can occur by coercion, punishment and policing (collectively called punishment hereafter) which, in essence, imply that a fine is imposed on defectors. As a result, the relative cost of defecting becomes greater compared with the alternate option of helping.

Numerous models of coercion, punishment and policing have been developed and they can be broadly separated in two classes. The first class mainly conceives punishment as a mechanism channelling the behaviour of defectors toward higher levels of cooperation (Boyd & Richerson, 1992; Clutton-Brock & Parker, 1995; Bowles & Gintis, 2004). The general idea of these models is that when individuals interact repeatedly, punishment is selected because the ensuing cost is more than compensated by the shift to cooperation of the partner. There are several important assumptions in these models (Boyd & Richerson, 1992; Bowles & Gintis, 2004). First, the punishing and cooperative traits are frequently assumed linked, constituting a so-called 'strong reciprocator' gene. However, there is no *a priori* reason to assume that these traits are linked. In fact, it is more likely that these traits will be unlinked (Gardner & West, 2004) and simulations suggest that cooperation is not stable when cooperation and punishing can co-evolve (L. Lehmann & L. Keller, unpublished data). The second and related assumption is that these models assume that 'strong reciprocators' can recognize and punish defectors conditionally. In other words, these models are akin to greenbeard models with helping behaviour being used as the 'recognition cue'. Accordingly, strong reciprocators are always harmful towards defectors. It remains to be studied what the consequences would be of allowing conditional expression of both cooperation and punishment as well as the possibility of these two traits to evolve independently with explicit gene dynamics.

The second class of models envisions punishment as a mechanism suppressing selfish behaviour, which may threaten group integrity and/or productivity (Clutton-Brock & Parker, 1995; Frank, 1995; Reeve & Keller, 1997). This would, for example, be the case of a behaviour that would increase the relative share of an individual at a cost to overall group productivity. These

models show that punishment should co-evolve with cooperation if the cost of being punished is sufficiently high to make it a better option not to behave selfishly and if the cost of punishment is smaller than the benefit gained by the punisher in terms of increased group productivity and survival. An important simplifying assumption of these models is that individuals cannot develop countermeasures or retaliate to punishment. It would be of interest to determine the evolutionary consequences of countermeasures and/or arms races between conflicting parties on the stability of the punishment and cooperation.

This brief overview of models reveals that punishment and other behaviours of that type have the potential to influence the cost/benefit ratio of cooperative acts. These behaviours can thus alter the social and demographic conditions where cooperation may evolve. However, these models are still in their infancy and it remains to be studied whether their predictions would be altered if some of their crucial assumptions were not fulfilled.

### A classification of models of helping

Our general model revealed that there are four general situations where helping is favoured. The first is when the act of helping provides direct benefits to the FI that outweighs the cost of helping (i.e. there are direct benefits). In that case helping simply evolves because it is associated with an increase of the direct fitness of the FI. The second situation is when the FI can alter the behaviour response of its partners by helping and thereby receives in return benefits that outweigh the cost of helping. In both situations, the helping act is cooperative as it results in an increase of the fitness of both the FI and its partners. A difference, however, is that in the first situation the increase of the FI's fitness is because of its own behaviour while in the second situation it results from the behavioural change induced in its partner(s). The third situation conducive to altruism is when the FI interacts and provides help to related individuals (i.e. kin selection). In that case, Hamilton's rule provides the conditions when helping can evolve even when it is associated with a decrease in the FI's direct fitness (i.e. when the act is altruistic). The fourth situation is a special case of the third with, in this case, recognition and helping being coded by two individual loci (i.e. greenbeard effect). In that case, helping can also evolve and remain stable when the two loci are linked and the conditions of Hamilton's rule are fulfilled.

In this section we shall briefly review several models proposed for the evolution of cooperation and altruism and investigate whether they can be classified in the four general categories outlined above or whether there are other general selected forces that may select for cooperation and altruism. We list in Table 3 a subset of models selected on the criteria of representing to us 'influential or original models'. This table reveals that all

| References                       | Helping         |               |               |            |
|----------------------------------|-----------------|---------------|---------------|------------|
|                                  | Cooperation     |               | Altruism      |            |
|                                  | -c>0            |               | -c<0          |            |
|                                  | Direct benefits | Reciprocation | Kin selection | Greenbeard |
| Friedman (1971)                  |                 | +             |               |            |
| Hamilton (1975)                  |                 |               | +             | +          |
| Cohen & Eshel (1976)             | +               |               |               |            |
| Wilson (1977)                    | +               |               |               |            |
| Uyenoyama & Feldman (1980)       | +               |               | +             |            |
| Axelrod & Hamilton (1981)        |                 | +             |               |            |
| Eshel & Cavalli-Sforza (1982)    |                 |               | +             | +          |
| Vehrencamp (1983)                | +               |               |               |            |
| Nunney (1985)                    | +               |               | +             |            |
| Queller (1985)                   | +               | +             | +             |            |
| Nowak & May (1992)               |                 |               | +             |            |
| Nowak & Sigmund (1992)           |                 | +             | +             |            |
| Dugatkin <i>et al.</i> (1994)    |                 | +             | +             |            |
| Frank (1995)                     | +               |               | +             |            |
| Nakamaru <i>et al.</i> (1997)    |                 | +             | +             |            |
| Wilson & Dugatkin (1997)         |                 |               | +             | +          |
| Nowak & Sigmund (1998)           |                 | +             |               |            |
| Van Baalen & Rand (1998)         |                 |               | +             |            |
| Michod (1998)                    | +               |               | +             |            |
| Roberts & Sherratt (1998)        |                 | +             |               |            |
| Reeve <i>et al.</i> (1998)       | +               |               | +             |            |
| Killingback <i>et al.</i> (1999) |                 |               | +             |            |
| Taylor & Irwin (2000)            | +               | +             | +             |            |
| Kokko <i>et al.</i> (2001)       | +               |               | +             |            |
| Eshel & Shaked (2001)            | +               |               |               |            |
| Aviles (2002)                    |                 |               |               | +          |
| Killingback & Doebeli (2002)     |                 | +             |               |            |
| Pepper & Smuts (2002)            |                 |               |               | +          |
| Le Galliard <i>et al.</i> (2003) |                 |               | +             |            |
| Bowles <i>et al.</i> (2003)      | +               |               | +             |            |
| Boyd <i>et al.</i> (2003)        | +               |               | +             |            |
| Traulsen & Schuster (2003)       |                 |               |               | +          |
| Axelrod <i>et al.</i> (2004)     |                 |               | +             |            |
| Panchanathan & Boyd (2004)       |                 | +             |               |            |
| Ohtsuki & Iwasa (2004)           |                 | +             |               |            |
| Taylor & Day (2004)              |                 | +             |               |            |
| Hauert (2004)                    | +               |               | +             |            |
| Pfeiffer <i>et al.</i> (2005)    |                 | +             |               |            |

**Table 3** Subset of models selected on the criteria of representing to us ‘influential or original models’.

these models fall within one of the four general categories at least once. Some models actually fall in several categories, and it is not always easy to disentangle the relative roles of the forces promoting altruism or cooperation.

Four types of models deserve special attention in that they have been proposed as providing new principles for the evolution of cooperation and altruism. The first class consists of ‘spatial structuring’ models (Nowak & May, 1992; Killingback *et al.*, 1999). A close inspection of these models shows that the actual selective force operating in the system is generally kin selection. Thus, in the simulations of Nowak & May (1992), altruists are more likely to be surrounded by altruists than defectors

at the beginning of the simulation, with the effect that altruists generally do better than defectors. In this situation, altruism can be maintained as long as individuals are more related in altruistic than in competitive interactions, which are the conditions required for kin selection to operate (Queller, 1994). Hence, the heuristic eqn (p. 1725) of Killingback *et al.* (1999) exactly gives Queller’s requirements for kin selection to be effective.

Several demographic factors can sustain the spread of an altruistic gene under ‘spatial structuring’ when initially rare. Thus, overlapping generations have a greater effect on the kin selected benefits of altruism



than on kin competition (Taylor & Irwin, 2000) (see supplementary material). Similarly, the capacity of the population to expand as a consequence of helping can facilitate the spread of altruism. This might occur when the population remains unsaturated through environmental and/or demographic stochasticity (Van Baalen & Rand, 1998; Mitteldorf & Wilson, 2000; Le Galliard *et al.*, 2003) or when helping increases group survival or carrying capacity (see supplementary material).

Realising that the actual force promoting altruism in spatially structured models in kin selection is important because it helps to identify the actual demographic and biological processes promoting the trait. Frequently, it is claimed that a new mechanism favouring cooperation or altruism has been identified. However, what has usually been found is a new situation (e.g. demographic or environmental stochasticity, overlapping generations, particular recognition mechanism) underlying higher relatedness during cooperative rather than competitive interactions. In the supplementary material we show that it is possible to disentangle between components of direct and kin selection in spatial structuring models and thus to identify the selective forces promoting investment into helping.

The second class of models are reproductive skew models which include ecological, genetic and social factors in a single explanatory framework and aim at determining how these factors jointly influence the apportionment of reproduction (reproductive skew) among group members (Vehrencamp, 1983; Reeve & Ratnieks, 1993; Reeve & Keller, 1995). In essence, reproductive skew models delineate the possible reproductive strategies available to a FI and define the conditions under which the best strategy is to cooperate and sacrifice part or all of its direct offspring production. Importantly, all these models are based on the explicit comparison of inclusive fitness of individuals adopting alternate reproductive strategies. An analysis of these models reveals that individuals will stay in the group and forego direct reproduction only when such an act provides either direct benefits (i.e.  $\zeta B - C > 0$  and the act is cooperative, for example because such a strategy increases group survival or the probability of inheriting a territory (Kokko & Johnston, 1999; Ragsdale, 1999) or because individuals can increase the reproductive output of related individuals (i.e.  $xrB - C > 0$  and the act is altruistic; e.g., (Reeve & Keller, 1995; Reeve *et al.*, 1998).

The third class of models are so-called 'tag-recognition' (Riolo *et al.*, 2001) and 'grouping' (Aviles, 2002) models. The tag-recognition system is when an altruistic gene is partially linked to a tag that can be recognized by other members in the population. In other words, these models fall in the greenbeard category with incomplete linkage between the altruistic and recognition traits. Realising this is useful for at least two reasons. First, it would have prevented confusion about the actual selective force at work. Second, it would have helped to realise that the

system is not stable over time because the association between the tag and altruistic genes is bound to decay just as any greenbeard mechanism (Roberts & Sherratt, 2002). Similarly, the 'grouping' model leads to altruism because altruistic individuals are more likely to group and interact. This is once again a special case of a greenbeard mechanism, which cannot be stable over time. Indeed, selection should favour nonaltruistic individuals to preferentially associate with altruists. As a result, the association between the altruist gene and the recognition trait (in that case grouping behaviour) will decay and altruism will disappear.

The final class of models to be discussed are the 'group selection' models. The general idea of these models is to use a multi-level selection approach to partition selection into components of within group and between group selection. Contrary to what is sometimes claimed, group selection models are not fundamentally different from classical models and it is possible in every instance to translate from one approach to the other without disturbing the mathematics describing the net result of selection, (see eqn A6 of the supplementary material) (Hamilton, 1975; Grafen, 1984; Dugatkin & Reeve, 1994; Frank, 1998; Rousset, 2004).

The transition from unicellularity to multicellularity is a classical example used to exemplify the role of group selection (Michod, 1998). Importantly, however, the high level of cooperation between cells in a multicellular organism can just as well be explained by kin selection (Queller, 2000). Indeed, a key factor necessary for the evolution of the highly cooperative nature of interactions between cells is probably a high relatedness, which is generally attained by multicellular organisms going through a unicellular phase such as the egg stage (Wolpert & Szathmary, 2002).

The other important selective force that operates in many group selection models is cooperative action providing direct benefits to the FI. A classical example is Wilson's (1977) model of random group formation where cooperation evolves only so far that the direct benefits to the FI exceeds the costs. Unfortunately, in-group selection models it is not always easy to determine the relative importance of relatedness and direct benefits. This is particularly true in settings where groups compete against each other and reproduce as, for instance, in the stochastic corrector model (Szathmary & Demeter, 1987). The difficulty with this class of models is that the costs and benefits are functions of group composition and growth rate, which are highly dependent on interactions within and between groups. The complexity of the situation makes it difficult to delineate analytically the relative importance of kin selection and direct benefits. But realising that the actual force promoting cooperation under group selection is a combination of kin selection and direct benefits allows us to delineate more clearly the role of the factors promoting or repressing cooperation and altruism.

## Conclusion

The conceptual framework developed here emphasizes that there are four general situations conducive to helping and that all models proposed so far can be classified accordingly. Hence, cooperation and altruism can evolve only when there are direct benefits to the FI performing a cooperative act, repeated interactions with direct or indirect information on the behaviour of the partner in previous moves, preferential interactions between related individuals and/or a linkage disequilibrium between genes coding for altruism and phenotypic traits that can be identified. In the three later cases helping evolves because there is a positive association between individuals at the genotypic and/or phenotypic levels. The other parameter of paramount importance is the cost-to-benefit ratio of helping acts that can be altered by coercion, punishment and policing. However, because these later behaviours are costly they can evolve and remain stable only when at least one of the four general conditions necessary for the evolution of cooperation and altruism is fulfilled.

The synthetic model we developed to study the evolution of helping made several assumptions such as dyadic interactions between individuals, reputation dynamics dependent only on the previous move of the partner, linear payoff stream, the cost and benefits of interactions varying linearly with the intensity of helping and independently of the number of interactions; and evolution proceeding in an unstructured population held at a constant size. Some of these assumptions are relaxed in the supplementary material, where it is shown that they do not affect our general conclusions. More generally, Rousset & Ronce (2004) recently studied the inclusive effects of behavioural traits in complex demographics and an inspection of their eqn (23) reveals that the conditions required for the evolution of helping can always be broken down into direct and indirect effects on the FI's fitness resulting from its own behaviour and that of various classes of relatives (see supplementary material). In other words, we are not aware of situations conducive to helping when at least one of our four conditions is not fulfilled.

In the future it would be very useful if new models of cooperation and altruism explicitly referred to these four general principles. Using a general framework will help to clarify the relationship between new and old models and to classify different situations belonging to the same mechanism. This will enable us to clearly determine whether the mechanism in question allows stable cooperation or whether it is likely to be unstable as in the case where linkage between altruistic and recognition genes decays over time. Finally, the use of a general framework will also greatly help readers to determine the originality of new models and whether or not they really provide new insights on the forces promoting cooperation and altruism in nature.

## Acknowledgments

We thank Stuart West for encouraging us to write our paper as a target article as well as him, Robert Boyd, François Balloux, Michel Chapuisat, Ernst Fehr, Pierre Fontanillas, Sebastian Bonhoeffer, Timothy Killingback, Karen Parker, Nicolas Perrin, Virginie Ravigné, Max Reuter, Karl Sigmund, Claus Wedekind, Tom Wenseleers, David Sloan Wilson and two anonymous reviewers for very useful comments on the manuscript and François Rousset for providing a draft of his book. This work was funded by grants of Swiss NSF to LK and Nicolas Perrin.

## Conflicts of interest

The authors declare that they have no competing financial interests.

## References

- Arnold, K.E. & Owens, I.P.F. 1999. Cooperative breeding in birds: the role of ecology. *Behav. Ecol.* **10**: 465–471.
- Aviles, L. 2002. Solving the freeloaders paradox: genetic associations and frequency-dependent selection in the evolution of cooperation among nonrelatives. *Proc. Natl. Acad. Sci. USA* **99**: 14268–14273.
- Axelrod, R. & Hamilton, W.D. 1981. The evolution of cooperation. *Science* **211**: 1390–1396.
- Axelrod, R., Hammond, R.A. & Grafen, A. 2004. Altruism via kin-selection strategies that rely on arbitrary tags with which they coevolve. *Evolution* **58**: 1833–1838.
- Bowles, S., Choi, J.K. & Hopfensitz, A. 2003. The co-evolution of individual behaviors and social institutions. *J. Theor. Biol.* **223**: 135–147.
- Bowles, S. & Gintis, H. 2004. The evolution of strong reciprocity: cooperation in heterogeneous populations. *Theor. Popul. Biol.* **65**: 17–28.
- Boyd, R. & Richerson, P.J. 1992. Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethol. Sociobiol.* **13**: 171–195.
- Boyd, R.G.H., Bowles, S. & Richerson, P.J. 2003. The evolution of altruistic punishment. *Proc. Natl. Acad. Sci. USA* **100**: 3531–3535.
- Brandt, H. & Sigmund, K. 2004. The logic of reprobation: assessment and action rules for indirect reciprocation. *J. Theor. Biol.* **231**: 475–486.
- Brown, J.L. 1983. Cooperation – a biologists dilemma. *Adv. Stud. Behav.* **13**: 1–37.
- Clutton-Brock, T.H. & Parker, G.A. 1995. Punishment in animal societies. *Nature* **373**: 209–216.
- Cohen, D. & Eshel, I. 1976. On the founder effect and the evolution of altruistic traits. *Theor. Popul. Biol.* **10**: 276–302.
- Dawkins, R. 1976. *The Selfish Gene*. Oxford university press, Oxford.
- Dugatkin, L.A. & Reeve, H.K. 1994. Behavioral ecology and levels of selection – dissolving the group selection controversy. *Adv. Stud. Behav.* **23**: 101–133.
- Dugatkin, L.A., Wilson, D.S., Farrand, L. III & Wilkens, R.T. 1994. Altruism, tit for tat and outlaw genes. *Evol. Ecol.* **8**: 431–437.

- Eshel, I. & Cavalli-Sforza, L.L. 1982. Assortment of encounters and evolution of cooperativeness. *Proc. Natl. Acad. Sci. USA* **79**: 1331–1335.
- Eshel, I. & Shaked, A. 2001. Partnership. *J. Theor. Biol.* **208**: 457–474.
- Fehr, E. & Fischbacher, U. 2003. The nature of human altruism. *Nature* **425**: 785–791.
- Frank, S.A. 1995. Mutual policing and repression of competition in the evolution of cooperative groups. *Nature* **377**: 520–522.
- Frank, S.A. 1998. *Foundations of Social Evolution*. Princeton University Press, Princeton, NJ.
- Friedman, M. 1971. A noncooperative equilibrium for super-games. *Rev. Econ. Stud.* **38**: 1–12.
- Gardner, A. & West, S.A. 2004. Cooperation and punishment, especially in humans. *Am. Nat.* **164**: 753–764.
- Grafen, A. 1984. Natural selection, kin selection and group selection. In: *Behavioural Ecology. An Evolutionary Approach* (J. R. Krebs & N. B. Davies, eds), pp. 62–84. Blackwell Scientific Publications, Oxford.
- Hamilton, W.D. 1964. The genetical evolution of social behaviour. *I. J. Theor. Biol.* **7**: 1–16.
- Hamilton, W.D. 1975. Innate social aptitudes of man, an approach from evolutionary genetics. In: *Biosocial Anthropology* (R. Fox, ed.), pp. 133–157. Malaby Press, London.
- Hammerstein, P. 2003. Why is reciprocity so rare in social animals? A protestant appeal. In: *Genetic and Cultural Evolution of Cooperation* (P. Hammerstein, ed.), pp. 481–496. MIT Press, Cambridge, MA.
- Hauert, C. & Doebeli, M. 2004. Spatial structure often inhibits the evolution of cooperation in the snowdrift game. *Nature* **428**: 643–646.
- Keller, L. & Chapuisat, M. 2001. Eusociality and cooperation. In: *Encyclopedia of Life Sciences*. Nature Publishing Group, London. www.els.net.
- Killingback, T. & Doebeli, M. 2002. The continuous prisoner's dilemma and the evolution of cooperation through reciprocal altruism with variable investment. *Am. Nat.* **160**: 421–438.
- Killingback, T., Doebeli, M. & Knowlton, N. 1999. Variable investment, the continuous prisoner's dilemma, and the origin of cooperation. *P. Roy. Soc. Lond. B Biol. Sci.* **266**: 1723–1728.
- Kokko, H. & Johnston, R.A. 1999. Social queuing in animal societies: a dynamic model of reproductive skew. *Philos. T. Roy. Soc. B* **226**: 1–8.
- Kokko, H., Johnston, R.A. & Clutton-Brock, T.H. 2001. The evolution of cooperative breeding through group augmentation. *Philos. T. Roy. Soc. B* **268**: 187–196.
- Le Galliard, J.F., Ferrière, R. & Dieckmann, U. 2003. The adaptive dynamics of altruism in spatially heterogeneous populations. *Evolution* **57**: 1–17.
- Lorberbaum, J. 1994. No strategy is evolutionarily stable in the repeated prisoners dilemma. *J. Theor. Biol.* **168**: 117–130.
- Lorberbaum, J.P., Bohning, D.E., Shastri, A. & Sine, L.E. 2002. Are there really no evolutionarily stable strategies in the iterated prisoner's dilemma? *J. Theor. Biol.* **214**: 155–169.
- Michod, R.E. 1998. Evolution of individuality. *J. Evol. Biol.* **11**: 225–227.
- Mitteldorf, J. & Wilson, D.S. 2000. Population viscosity and the evolution of altruism. *J. Theor. Biol.* **204**: 481–496.
- Nakamaru, M., Matsuda, H. & Iwasa, Y. 1997. The evolution of cooperation in a lattice-structured population. *J. Theor. Biol.* **184**: 65–81.
- Nowak, M.A. & May, R.M. 1992. Evolutionary games and spatial chaos. *Nature* **359**: 826–829.
- Nowak, M.A. & Sigmund, K. 1992. Tit-for-tat in heterogeneous populations. *Nature* **355**: 250–253.
- Nowak, M.A. & Sigmund, K. 1998. Evolution of indirect reciprocity by image scoring. *Nature* **393**: 573–577.
- Nunney, L. 1985. Group selection, altruism, and structured-deme models. *Am. Nat.* **126**: 212–230.
- Ohtsuki, H. 2004. Reactive strategies in indirect reciprocity. *J. Theor. Biol.* **227**: 299–314.
- Ohtsuki, H. & Iwasa, Y. 2004. How should we define goodness? Reputation dynamics in indirect reciprocity. *J. Theor. Biol.* **231**: 107–120.
- Panchanathan, K. & Boyd, R. 2003. A tale of two defectors: the importance of standing for evolution of indirect reciprocity. *J. Theor. Biol.* **224**: 115–126.
- Panchanathan, K. & Boyd, R. 2004. Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature* **432**: 499–502.
- Pepper, J.W. & Smuts, B.B. 2002. A mechanism for the evolution of altruism among nonkin: positive assortment through environmental feedback. *Am. Nat.* **160**: 205–213.
- Pfeiffer, T., Rutte, C., Killingback, T., Taborsky, M. & Bonhoeffer, S. 2005. Evolution of cooperation by generalized reciprocity. *P. Roy. Soc. B-Biol. Sci.* **272**: 1115–1120.
- Queller, D.C. 1985. Kinship, reciprocity and synergism in the evolution of social behavior. *Nature* **318**: 366–367.
- Queller, D.C. 1994. Genetic relatedness in viscous populations. *Evol. Ecol.* **8**: 70–73.
- Queller, D.C. 2000. Relatedness and the fraternal major transitions. *Philos. T. Roy. Soc. B* **355**: 1647–1655.
- Ragsdale, J.E. 1999. Reproductive skew theory extended: the effect of resource inheritance on social organization. *Evol. Ecol. Res.* **1**: 859–874.
- Reeve, H.K. 1989. The evolution of conspecific acceptance thresholds. *Am. Nat.* **133**: 407–435.
- Reeve, H.K., Emlen, S.T. & Keller, L. 1998. Reproductive sharing in animal societies: reproductive incentives or incomplete control by dominant breeders? *Behav. Ecol.* **9**: 267–278.
- Reeve, H.K. & Keller, L. 1995. Partitioning of reproduction in mother-daughter versus sibling associations: a test of optimal skew theory. *Am. Nat.* **145**: 119–132.
- Reeve, H.K. & Keller, L. 1997. Reproductive bribing and policing evolutionary mechanisms for the suppression of within-group selfishness. *Am. Nat.* **150**: S42–S58.
- Reeve, H.K. & Ratnieks, F.L.W. 1993. Queen–queen conflict in polygynous societies: mutual tolerance and reproductive skew. In: *Queen Number and Sociality in Insects* (L. Keller, ed.), pp. 45–85. Oxford: Oxford University Press.
- Riolo, R.L., Cohen, M.D. & Axelrod, R. 2001. Evolution of cooperation without reciprocity. *Nature* **414**: 441–443.
- Roberts, G. & Sherratt, T.N. 1998. Development of cooperative relationship through increasing investment. *Nature* **9**: 175–179.
- Roberts, G. & Sherratt, T.N. 2002. Behavioural evolution – does similarity breed cooperation? *Nature* **418**: 499–500.
- Rousset, F. 2004. *Genetic Structure and Selection in Subdivided Populations*, Princeton edn. Princeton University Press, Princeton, NJ.
- Rousset, F. & Billiard, S. 2000. A theoretical basis for measures of kin selection in subdivided populations: finite populations and localized dispersal. *J. Evol. Biol.* **13**: 814–825.

- Rousset, F. & Ronce, O. 2004. Inclusive fitness for traits affecting metapopulation demography. *Theor. Popul. Biol.* **65**: 127–141.
- Sachs, J.L., Mueller, U.G., Wilcox, T.P. & Bull, J.J. 2004. The evolution of cooperation. *Q. Rev. Biol.* **79**: 135–160.
- Seger, J. 1993. Cooperation and conflict in social insects. In: *Behavioural Ecology. An Evolutionary Approach* (J. R. Krebs & N. B. Davies, eds), pp. 338–373. Blackwell Scientific Press, Oxford.
- Stevens, J.R. 2004. The selfish nature of generosity: harassment and food sharing in primates. *P. Roy. Soc. Lond. B Bio.* **271**: 451–456.
- Szathmary, E. & Demeter, L. 1987. Group selection of early replicators and the origin of life. *J. Theor. Biol.* **128**: 463–486.
- Taylor, P.D. 1992a. Altruism in viscous populations – an inclusive fitness model. *Evol. Ecol.* **6**: 352–356.
- Taylor, P.D. 1992b. Inclusive fitness in a homogeneous environment. *P. Roy. Soc. Lond. B Bio.* **249**: 299–302.
- Taylor, P.D. & Day, D. 2004. Stability in negotiation games and the emergence of cooperation. *P. Roy. Soc. Lond. B Bio.* **271**: 669–674.
- Taylor, P.D. & Frank, S.A. 1996. How to make a kin selection model. *J. Theor. Biol.* **54**: 1135–1141.
- Taylor, P.D. & Irwin, A.J. 2000. Overlapping generations can promote altruistic behavior. *Evolution* **54**: 1135–1141.
- Traulsen, A. & Schuster, H.G. 2003. Minimal model for tag-based cooperation. *Phys. Rev. E* **68**: 1–8.
- Trivers, R.L. 1971. The evolution of reciprocal altruism. *Q. Rev. Biol.* **46**: 35–57.
- Uyenoyama, M. & Feldman, M.W. 1980. Theories of kin and group selection – a population-genetics perspective. *Theor. Popul. Biol.* **17**: 380–414.
- Van Baalen, M. & Rand, D.A. 1998. The unit of selection in viscous populations and the evolution of altruism. *J. Theor. Biol.* **193**: 631–648.
- Vehrencamp, S.L. 1983. A model for the evolution of despotic versus egalitarian societies. *Anim. Behav.* **31**: 667–682.
- Wahl, L.M. & Nowak, M.A. 1999a. The continuous prisoner's dilemma: I. Linear reactive strategies. *J. Theor. Biol.* **200**: 307–321.
- Wahl, L.M. & Nowak, M.A. 1999b. The continuous prisoner's dilemma: II. Linear reactive strategies with noise. *J. Theor. Biol.* **200**: 323–338.
- West, S.A., Pen, I. & Griffin, A.S. 2002. Conflict and cooperation – cooperation and competition between relatives. *Science* **296**: 72–75.
- Wilson, D.S. 1977. Structured demes and evolution of group-advantageous traits. *Am. Nat.* **111**: 157–185.
- Wilson, D.S. 1979. Structured demes and trait-group variation. *Am. Nat.* **113**: 606–610.
- Wilson, D.S. & Dugatkin, L.A. 1997. Group selection and assortative interactions. *Am. Nat.* **149**: 336–351.
- Wolpert, L. & Szathmary, E. 2002. Multicellularity: evolution and the egg. *Nature* **420**: 745–745.

### Supplementary material

The following supplementary material is available for this article online:

**Appendix S1.** Measuring selection on helping

**Appendix S2.** Selection on helping in a randomly mixing population

**Appendix S3.** Selection on helping in a geographically structured population

These materials are available as part of the online article from <http://www.blackwell-synergy.com>

Received 9 November 2005; revised 18 January 2006; accepted 18 January 2006