The evolution of cooperation and altruism.

A general framework and a classification of models

Laurent Lehmann[1,2] and Laurent Keller[1]

[1]Department of Ecology and Evolution, University of Lausanne, Biophore, 1015 Lausanne, Switzerland..

[2]Department of Genetics, University of Cambridge, Downing Street, CB2 3EH Cambridge, UK

Lehmann: Phone: + 44 -1223-332584 Email: ll316@cam.ac.uk

Keller: Phone: +41-21-6924173 Email: Laurent.Keller@unil.ch

To be published as a "Target Review" in Journal of Evolutionary Biology

Summary

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, that is, 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. The model is framed within the direct fitness approach (Taylor and Frank 1996, Frank 1998, Rousset and 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 ω, 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 non-randomly 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 ζ of the benefit generated by helping directly return 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., ζ = 1) or does not invest into helping (i.e., $I_\bullet = 0$).

Since 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

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

(see also supplementary material, eq.8). In this equation, 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):

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

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 and Nowak 1999a, Killingback and Doebeli 2002). Hence, the investment depends on three traits: the investment on the first round τ, the response slope β 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 (τ and β) can evolve and the dynamics of investment of the focal individual 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 (eq.2) allows us to determine the inclusive fitness effect (Hamilton 1964) and to establish the direction of selection on the two evolving traits τ and β. 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 and Frank 1996, Rousset 2004). Accordingly, one counts the increment (or decrement) in the focal individual's fitness stemming from the expression of the behavior of all its relatives in the population. Then, the inclusive fitness effect of the initial move τ reads as (see supplementary material eqs. 4, 13 and 14)

$$\Delta W_{IF} = \underbrace{\frac{\zeta B + (1-\zeta)\omega m\beta B - C}{(1+(B-C)-\omega m\beta)(1+\omega m\beta)}}_{-c'} + r\underbrace{\frac{x\left[(1-\zeta)B + \omega m\beta(\zeta B - C)\right]}{(1+(B-C)-\omega m\beta)(1+\omega m\beta)}}_{b'}. \qquad (4)$$

As in Hamilton's framework (1964), 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 versus negative effects on the direct fitness of the FI.

Because the inclusive fitness effect of the response slope is proportional to equation (4) (see supplementary material eqs.13-14 and eqs.15-16) the same conditions must be satisfied for the inclusive fitness effect of τ and β to be positive, that is

$$\underbrace{\zeta B + (1-\zeta)\omega m\beta B - C}_{-c'} + r\,\underbrace{x\left[(1-\zeta)B + \omega m\beta(\zeta B - C)\right]}_{b'} > 0\,. \qquad (5)$$

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., ineq. (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

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

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 and Feldman 1980, Nunney 1985). In that case ζ 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 focal individual (i.e., ζ >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 ζ 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 surival (Eshel and 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

$$\omega m \beta B - C > 0 \tag{7}$$

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 ($m = 1$), inequ.(7) reduces to the threshold theorem of Killingback and Doebeli (2002) when individuals do not take into account their own investment in the previous move while interacting with their partner. This result emphasize 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 and 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 and Fischbacher 2003). Experimental studies also revealed that humans are more likely to cooperate with individuals that have been cooperative in previous interactions (Fehr and 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 2000, 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 and 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 and 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 and 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

ineq.(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 and Sigmund 1998, see Table 1 and eq.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 and 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 and Nowak 1999a, Killingback and 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 and Nowak (1999b), discriminator cooperative strategies can invade defectors but when discriminator co-operative strategies have reached a high frequency, non-discriminative 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 and 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 and 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 and 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 since it is associated with a decrease in the FI's direct fitness but an in increase in the direct fitness of individuals receiving help), inequ. (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 (greanbeard 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

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

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 ineq.(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 help is provided irrespective of relatedness ($x = 0$) such altruism cannot evolve when there is no repeated interactions (i.e., $\omega = 0$), a result which usually holds whatever the genetic structure of the population (Taylor 1992b, Rousset 2004) (see supplementary material).

Several mechanisms may generate a 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 and 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). Since 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 (Seger 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 greanbeard gene, a concept invented by Hamilton (1964) and named by

Dawkins (1976). A greanbeard gene is defined as a gene that causes a phenotypic effect (e.g., the presence of a greanbeard or any other conspicuous feature) that allows the bearer of this feature to recognize it in other individuals, and results in the bearer to behaving 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

$$aB - C > 0 \tag{9}$$

is satisfied and one recovers the conditions described by Hamilton (1975). Importantly, this inequality is similar to ineq.(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 and Cavalli-Sforza 1982). If recombination can break down the linkage between "recognition" and "altruistic" genes, the situation become 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 pre-adaptations 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 and 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 to 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 channeling the behaviour of defectors toward higher levels of cooperation (Boyd and Richerson 1992, Clutton-Brock and Parker 1995, Bowles and Gintis 2004). The general idea of these models is that when individuals interact repeatedly, punishment is selected for 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 and Richerson 1992, Bowles and 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 trait will be unliked (Gardner and West 2004) and simulations suggest that cooperation is not stable when cooperation and punishing can co-evolve (L. Lehmann and 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". Thus, individuals are always cooperative with strong reciprocators and harmful towards other individuals. 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 envision punishment as a mechanism suppressing selfish behaviour which may threaten group integrity and/or productivity (Clutton-Brock and Parker 1995, Frank 1995, Reeve and 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 are 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 since 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 due to 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 together 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 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 actively 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 and 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 and 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 equation (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 and 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 and Rand 1998, Mitteldorf and 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 is 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 and Ratnieks 1993, Reeve and Keller 1995). In essence reproductive skew models delineate the possible reproductive strategies available to a focal individual and define the conditions under which the best strategy is to cooperate and sacrifice part or all of it 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 and 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 and 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 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 and 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 non-altruistic 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 eq.A6 of the supplementary material) (Hamilton 1975, Grafen 1984, Dugatkin and 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 and Szathmary 2002).

The other important selective force that operates in many group selection models is cooperative action providing direct benefits to the focal individual. A classical example is Wilson's model (1977) of random group formation where cooperation evolves only so far that the direct benefits to the focal individual 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, the stochastic corrector model (Szathmary and 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 focal individual 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 and Ronce (2004) recently studied the inclusive effects of behavioural traits in complex demographies and an inspection of their eq.(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 refers 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.

Reference and Notes

Arnold, K. E., and I. P. F. Owens. 1999. Cooperative breeding in birds: the role of ecology. Behavioral Ecology **10**:465-471.

Aviles, L. 2002. Solving the freeloaders paradox: genetic associations and frequency-dependent selection in the evolution of cooperation among nonrelatives. Proceedings of the National Academy of Sciences of the United States of America **99**:14268-14273.

Axelrod, R., and W. D. Hamilton. 1981. The evolution of cooperation. Science **211**:1390-1396.

Axelrod, R., R. A. Hammond, and A. Grafen. 2004. Altruism via kin-selection strategies that rely on arbitrary tags
with which they coevolve. Evolution **58**:1833-1838.

Bowles, S., J. K. Choi, and A. Hopfensitz. 2003. The co-evolution of individual behaviors and social institutions. Journal of Theoretical Biology **223**:135-147.

Bowles, S., and H. Gintis. 2004. The evolution of strong reciprocity: cooperation in heterogeneous populations. Theoretical Population Biology **65**:17-28.

Boyd, R., and P. J. Richerson. 1992. Punishment allows the evolution of cooperation (or anything else) in sizable groups. Ethology and Sociobiology **13**:171-195.

Boyd, R. G., H.; Bowles, S. & Richerson, P. J. 2003. The evolution of altruistic punishment. Proceedings of the National Academy of Sciences of the United States of America **100**:3531--3535.

Brandt, H., and K. Sigmund. 2004. The logic of reprobation: assessment and action rules for indirect reciprocation. Journal of Theoretical Biology **231**:475-486.

Brown, J. L. 1983. Cooperation - a biologists dilemma. Advances in the Study of Behavior **13**:1-37.

Clutton-Brock, T. H., and G. A. Parker. 1995. Punishment in animal societies. Nature **373**:209-216.

Cohen, D., and I. Eshel. 1976. On the founder effect and the evolution of altruistic traits. Theoretical Population Biology **10**:276-302.

Dawkins, R. 1976. The Selfish Gene. Oxford university press, Oxford.

Dugatkin, L. A., and H. K. Reeve. 1994. Behavioral ecology and levels of selection - Dissolving the group selection controversy. Pages 101-133 *in* Advances in the Study of Behavior, Vol 23.

Dugatkin, L. A., D. S. Wilson, L. Farrand III, and R. T. Wilkens. 1994. Altruism, tit for tat and outlaw genes. Evolutionary Ecology **8**:431-437.

Eshel, I., and L. L. Cavalli-Sforza. 1982. Assortment of encounters and evolution of cooperativeness. Proceedings of the National Academy of Sciences of the United States of America **79**:1331-1335.

Eshel, I., and A. Shaked. 2001. Partnership. Journal of Theoretical Biology **208**:457-474.

Fehr, E., and U. Fischbacher. 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.

Friedman, M. 1971. A non-cooperative equilibrium for supergames. The Review of Economic Studies **38**:1-12.

Gardner, A., and S. A. West. 2004. Cooperation and punishment, especially in humans. American Naturalist **164**:753-764.

Grafen, A. 1984. Natural selection, kin selection and group selection. Pages 62-84 *in* J. R. Krebs and N. B. Davies, editors. Behavioural ecology. An evolutionary approach. Blackwell Scientific Publications.

Hamilton, W. D. 1964. The genetical evolution of social behaviour. I. Journal of Theoretical Biology **7**:1-16.

Hamilton, W. D. 1975. Innate social aptitudes of man, an approach from evolutionary genetics. Pages 133-157 *in* R. Fox, editor. Biosocial anthropology. Malaby Press, London.

Hammerstein, P. 2003. Why is reciprocity so rare in social animals? A protestant appeal. Pages 481-496 *in* P. Hammerstein, editor. Genetic and Cultural Evolution of Cooperation. MIT Press.

Hauert, C. D., M. 2004. Spatial structure often inhibits the evolution of cooperation in the snowdrift game. Nature **428**:643--646.

Keller, L., and M. Chapuisat. 2001. Eusociality and cooperation. *in* Encyclopedia of Life Sciences. London: Nature
Publishing Group.

Killingback, T., and M. Doebeli. 2002. The continuous prisoner's dilemma and the evolution of cooperation through reciprocal altruism with variable investment. American Naturalist **160**:421-438.

Killingback, T., M. Doebeli, and N. Knowlton. 1999. Variable investment, the continuous prisoner's dilemma, and the origin of cooperation. Proceedings of the Royal Society of London Series B-Biological Sciences **266**:1723-1728.

Kokko, H., and R. A. Johnston. 1999. Social queuing in animal societies: A dynamic model of reproductive skew. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences **226**:1-8.

Kokko, H., R. A. Johnston, and T. H. Clutton-Brock. 2001. The evolution of cooperative breeding through group augmentation. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences **268**:187-196.

Le Galliard, J. F., R. Ferrière, and U. Dieckmann. 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. Journal of Theoretical Biology **168**:117-130.

Lorberbaum, J. P., D. E. Bohning, A. Shastri, and L. E. Sine. 2002. Are there really no evolutionarily stable strategies in the iterated prisoner's dilemma? Journal of Theoretical Biology **214**:155-169.

Michod, R. E. 1998. Evolution of individuality. Journal of Evolutionary Biology **11**:225-227.

Mitteldorf, J., and D. S. Wilson. 2000. Population viscosity and the evolution of altruism. Journal of Theoretical Biology **204**:481-496.

Nakamaru, M., H. Matsuda, and Y. Iwasa. 1997. The evolution of cooperation in a lattice-structured population. Journal of Theoretical Biology **184**:65-81.

Nowak, M. A., and R. M. May. 1992. Evolutionary games and spatial chaos. Nature **359**:826-829.

Nowak, M. A., and K. Sigmund. 1992. Tit-for-Tat in heterogeneous populations. Nature **355**:250-253.

Nowak, M. A., and K. Sigmund. 1998. Evolution of indirect reciprocity by image scoring. Nature **393**:573-577.

Nunney, L. 1985. Group selection, altruism, and structured-deme models. American Naturalist **126**:212-230.

Ohtsuki, H. 2004. Reactive strategies in indirect reciprocity. Journal of Theoretical Biology **227**:299-314.

Ohtsuki, H., and Y. Iwasa. 2004. How should we define goodness? Reputation dynamics in indirect reciprocity. Journal of Theoretical Biology **231**:107-120.

Panchanathan, K., and R. Boyd. 2003. A tale of two defectors: the importance of standing for evolution of indirect reciprocity. Journal of Theoretical Biology **224**:115-126.

Panchanathan, K., and R. Boyd. 2004. Indirect reciprocity can stabilize cooperation without the second-order free rider problem. Nature **432**:499-502.

Pepper, J. W., and B. B. Smuts. 2002. A mechanism for the evolution of altruism among nonkin: Positive assortment through environmental feedback. American Naturalist **160**:205-213.

Pfeiffer, T., C. Rutte, T. Killingback, M. Taborsky, and S. Bonhoeffer. 2005. Evolution of cooperation by generalized reciprocity. Proceedings of the Royal Society B-Biological Sciences **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. Evolutionary Ecology **8**:70-73.

Queller, D. C. 2000. Relatedness and the fraternal major transitions. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences **355**:1647-1655.

Ragsdale, J. E. 1999. Reproductive skew theory extended: the effect of resource inheritance on social organization. Evolutionary Ecology Research **1**:859-874.

Reeve, H. K. 1989. The evolution of conspecific acceptance thresholds. American Naturalist **133**:407-435.

Reeve, H. K., S. T. Emlen, and L. Keller. 1998. Reproductive sharing in animal societies: reproductive incentives or incomplete control by dominant breeders? Behavioral Ecology **9**:267-278.

Reeve, H. K., and L. Keller. 1995. Partitioning of reproduction in mother-daughter versus sibling associations: a test of optimal skew theory. American Naturalist **145**:119-132.

Reeve, H. K., and L. Keller. 1997. Reproductive bribing and policing evolutionary mechanisms for the suppression of within-group selfishness. American Naturalist **150**:S42-S58.

Reeve, H. K., and F. L. W. Ratnieks. 1993. Queen-queen conflict in polygynous societies: mutual tolerance and reproductive skew. *in* L. Keller, editor. Queen number and sociality in insects. Oxford, Oxford.

Riolo, R. L., M. D. Cohen, and R. Axelrod. 2001. Evolution of cooperation without reciprocity. Nature **414**:441-443.

Roberts, G., and T. N. Sherratt. 1998. Development of cooperative relationship through increasing investment. Nature **9**:175-179.

Roberts, G., and T. N. Sherratt. 2002. Behavioural evolution - Does similarity breed cooperation? Nature **418**:499-500.

Rousset, F. 2004. Genetic structure and selection in subdivided populations, Princeton edition. Princeton University Press, NJ.

Rousset, F., and S. Billiard. 2000. A theoretical basis for measures of kin selection in subdivided populations: finite populations and localized dispersal. Journal of Evolutionary Biology **13**:814-825.

Rousset, F., and O. Ronce. 2004. Inclusive fitness for traits affecting metapopulation demography. Theoretical Population Biology **65**:127-141.

Sachs, J. L., U. G. Mueller, T. P. Wilcox, and J. J. Bull. 2004. The evolution of cooperation. The Quarterly Review of Biology **79**:135-160.

Seger, J. 1993. Cooperation and conflict in social insects. Pages 338-373 *in* J. R. Krebs and N. B. Davies, editors. Behavioural Ecology. An evolutionary Approach. Oxford, Blackwell Scientific Press.

Stevens, J. R. 2004. The selfish nature of generosity: harassment and food sharing in primates. Proceedings of the Royal Society of London Series B-Biological Sciences **271**:451-456.

Szathmary, E., and L. Demeter. 1987. Group selection of early replicators and the origin of life. Journal of Theoretical Biology **128**:463-486.

Taylor, P. D. 1992a. Altruism in viscous populations - an inclusive fitness model. Evolutionary Ecology **6**:352-356.

Taylor, P. D. 1992b. Inclusive fitness in a homogeneous environment. Proceedings of the Royal Society of London Series B-Biological Sciences **249**:299-302.

Taylor, P. D., and D. Day. 2004. Stability in negotiation games and the emergence of cooperation. Proceedings of the Royal Society of London Series B-Biological Sciences **271**.

Taylor, P. D., and S. A. Frank. 1996. How to make a kin selection model. Journal of Theoretical Biology **54**:1135-1141.

Taylor, P. D., and A. J. Irwin. 2000. Overlapping generations can promote altruistic behavior. Evolution **54**:1135-1141.

Traulsen, A., and H. G. Schuster. 2003. Minimal model for tag-based cooperation. Physical Review E **68**.

Trivers, R. L. 1971. The evolution of reciprocal altruism. Quarterly Review of Biology **46**:35-57.

Uyenoyama, M., and M. W. Feldman. 1980. Theories of kin and group selection - a population-genetics perspective. Theoretical Population Biology **17**:380-414.

Van Baalen, M., and D. A. Rand. 1998. The unit of selection in viscous populations and the evolution of altruism. Journal of Theoretical Biology **193**:631-648.

Vehrencamp, S. L. 1983. A model for the evolution of despotic versus egalitarian societies. Animal Behaviour **31**:667-682.

Wahl, L. M., and M. A. Nowak. 1999a. The continuous prisoner's dilemma: I. Linear reactive strategies. Journal of Theoretical Biology **200**:307-321.

Wahl, L. M., and M. A. Nowak. 1999b. The continuous prisoner's dilemma: II. Linear reactive strategies with noise. Journal of Theoretical Biology **200**:323-338.

West, S. A., I. Pen, and A. S. Griffin. 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. American Naturalist **111**:157-185.
Wilson, D. S. 1979. Structured demes and trait-group variation. American Naturalist **113**:606-610.
Wilson, D. S., and L. A. Dugatkin. 1997. Group selection and assortative interactions. American Naturalist **149**:336-351.
Wolpert, L., and E. Szathmary. 2002. Multicellularity: evolution and the egg. Nature **420**:745-745.

Supporting Online Material

Supplementary material

Acknowledgements

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.

Competing interests statement The authors declare that they have no competing financial interests.

Table 1. List of symbols

FI	Abreviation 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
I	Level of investment into helping (varying between 0 and 1)
$I_{i,j}(t)$	Level of investment into helping at round t of an individual of class i engaged in repeated interactions with an individual of class j
$I_{\bullet,j}(t)$	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 behavior of the focal individual on its fitness
b'	Can be interpreted in two ways Either as the effect of the behavior 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
τ	Evolving level of investment into helping (varying between 0 and 1) on the first round
β	Evolving response slope (varying between 0 and 1) on the partner's investment at the previous round
α	Evolving response slope (varying between 0 and 1) on the partner's image score
z	Generic designation of an evolving phenotype, here τ, β or α
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
ς	Proportion of the benefits generated by a helping act that directly return to the focal individual
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 non-randomly with an individual of the related class
a	Probability that a individual interacts non-randomly 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 groupes (demes) in the population
d_i	Disperal probability at distance i
d_0	Probability of staying in the natal patch $\left(d_0 = 1 - \sum_{i=1}^{n_d - 1} m_i \right)$
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

Table 2 Classification of selective pressures promoting helping

<table>
<tr><td colspan="4">Helping

$rb'-c'>0$

$\underbrace{\zeta B+(1-\zeta)\omega m\beta B-C}_{-c'}+r\,\underbrace{x\left[(1-\zeta)B+\omega m\beta(\zeta B-C)\right]}_{b'}>0$ (eq.8)</td></tr>
<tr><td colspan="2">Cooperation

$-c'>0$</td><td colspan="2">Altruism

$-c'<0$</td></tr>
<tr><td>Direct
Benefits

$\omega=0$ and $r=0$</td><td>Reciprocation

$\zeta=0$ and $r=0$</td><td>Kin
Selection

$\omega=0$ and $\zeta=0$</td><td>Greenbeard

$\omega=0$ and $\zeta=0$</td></tr>
<tr><td rowspan="2">$\underbrace{\zeta B-C}_{-c'}>0$ (eq6)</td><td>Direct reciprocity

$\underbrace{\omega m\beta B-C}_{-c'}>0$ (eq7)</td><td rowspan="2">$r\,\underbrace{xB}_{b'}-\underbrace{C}_{c'}>0$ (eq8)</td><td rowspan="2">$\underbrace{aB}_{b'}-\underbrace{C}_{c'}>0$ (eq9)
here $r=1$</td></tr>
<tr><td>Indirect reciprocity

$\underbrace{\omega q\alpha B-C}_{-c'}>0$
(eq26 of the appendix)</td></tr>
</table>

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

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

References	Helping			
	Cooperation $-c > 0$		**Altruism** $-c < 0$	
	Direct benefits	Reciprocation	Kin selection	Greenbeard
(Friedman 1971)		+		
(Hamilton 1975)			+	+
(Cohen and Eshel 1976)	+			
(Wilson 1977)	+			
(Uyenoyama and Feldman 1980)	+		+	
(Axelrod and Hamilton 1981)		+		
(Eshel and Cavalli-Sforza 1982)			+	+
(Vehrencamp 1983)	+			
(Nunney 1985)	+		+	
(Queller 1985)	+	+	+	
(Nowak and May 1992)			+	

(Nowak and Sigmund 1992)		+	+?	
(Dugatkin et al. 1994)		+	+	
(Frank 1995)	+		+	
(Nakamaru et al. 1997)		+	+	
(Wilson and Dugatkin 1997)			+	+
(Nowak and Sigmund 1998)		+		
(Van Baalen and Rand 1998)			+	
(Michod 1998)	+		+	
(Roberts and Sherratt 1998)		+		
(Reeve et al. 1998)	+		+	
(Killingback et al. 1999)			+	
(Taylor and Irwin 2000)	+	+	+	
(Kokko et al. 2001)	+		+	
(Eshel and Shaked 2001)	+			
(Aviles 2002)				+
(Killingback and Doebeli 2002)		+		
(Pepper and Smuts 2002)				+
(Le Galliard et al. 2003)			+	
(Bowles et al. 2003)	+		+**?**	

(Boyd 2003)	+		+**?**	
(Traulsen and Schuster 2003)				+
(Axelrod et al. 2004)			+	
(Panchanathan and Boyd 2004)		+		
(Ohtsuki and Iwasa 2004)		+		
(Taylor and Day 2004)		+		
(Hauert 2004)	+		+	
(Pfeiffer et al. 2005)		+		

Supplementary material for "The evolution of cooperation and altruism. A general framework and a classification of models"

This supplementary material is divided into three sections. In the first section, we provide a description of the selective pressure on a helping trait under broad demographic and ecological conditions. In particular, we highlight how Hamilton's rule emerges as a particular situation of the general selective pressure and delineate its relationship to multilevel selection theory. In the second section, we analyze the explicit selective pressure on helping under both *direct* and *indirect* reciprocity when evolution occurs in an a panmictic population. Finally, in the third section we analyze the selective pressure on helping under one-shot interactions when evolution occurs in a geographically structured population. This sections illustrates the fundamental role played by kin selection under "spatial structuring" models.

1.Measuring selection on helping

The inclusive fitness effect

Here we provide a description of the selective pressure acting on a social trait (z) affecting the fecundity and/or survival of actors and recipients. The strength of both effects is a function of the phenotype of the actors which in turn is determined by their genotype. Selection on such a trait can be analysed by considering a mutant allele coding for a phenotypic value deviating by small magnitude from that expressed by individuals bearing a resident allele fixed in the population. The selective pressure on such a mutant allele, which determines whether it will increase or decrease in frequency in the population is given by Hamilton's inclusive fitness effect (Hamilton, 1964, pp. 6, 15). The inclusive fitness effect represents the marginal contribution of the allele to fitness and there are two standard ways by which it can be

evaluated. First, by summing up the effects of individual bearing the mutant allele (i.e., the actors) on the fitness of all individuals in the population (i.e., the receptors) weighted by their genetic similarity to the actors. Alternatively, the inclusive fitness effect can be evaluated in the direct fitness manner as a relatedness weighted sum of the effects of all individuals in the population on the fitness of individuals bearing the mutant allele (Taylor and Frank, 1996; Frank, 1998; Rousset and Billiard, 2000). Both ways of evaluating selection are equivalent (Rousset, 2004, p. 108). We use here the direct fitness method as developed in Rousset and Ronce (2004) and Rousset (2003, 2004), which fits within a one locus population genetic model and provides an exact descriptions of the first-order effects of selection (weak selection) on allele frequency change. Accordingly, all components of selection are taken into account, whether these are described as individual, kin, or group selection. This approach can be interpreted as a particular case of the general multilocus selection theory developed by Kirkpatrick *et al.* (2002). Relaxation of the one locus assumption does not change our general conclusions albeit introducing additional component of selection resulting from the association of genes within and between individuals.

In the presence of additive gene action and weak selection, the inclusive fitness effect can be decomposed into two terms:

$$\Delta W_{\mathrm{IF}} = S_f + S_{\mathrm{Pr}}, \tag{1}$$

(Rousset and Ronce, 2004, eq. 25). The first term in this equation (S_f) is a weighted effect of all individuals in the population on the expected number of actor's offspring reaching adulthood. The second term, S_{Pr}, is a weighted effect of all individuals, through changes in the demographic states of the population, on the reproductive value of these offspring. In a population of constant size or when a trait cannot affect the demography, $S_{\mathrm{Pr}} = 0$ and the inclusive fitness effect (ΔW_{IF}) reduces to the classical selective pressure for structured population in the presence of kin selection (Taylor, 1990, 1996; Taylor and Frank, 1996; Frank, 1998), where the structure can for instance be by sex, age or geography.

For simplicity, we now consider a population of constant size (see below for a relaxation of this latter assumption) that can be geographically structured. In such a population, the fitness of a focal individual $w \equiv w(z_\bullet, ..., z_j, ...)$ can be expressed as a function of its own phenotype ($z_\bullet$) and the average phenotypes (z_j) of different classes of actors (labelled j) affecting its fitness (Taylor and Frank, 1996; Frank, 1998; Rousset, 2004). The inclusive fitness effect then reads

$$\Delta W_{\mathrm{IF}} = \frac{\partial w}{\partial z_\bullet} Q_\bullet + \sum_j \frac{\partial w}{\partial z_j} Q_j. \tag{2}$$

This equation sums up the effect of the action of all actors in the population on the fitness of the focal individual, that is, on its expected number of offspring reaching adulthood. The effect of each category of actors comes under the form of a weighted partial derivative of the fitness of the focal individual (w) with respect to the average phenotype of individual's in that category. The weight is the probability that an individual of that category bears a copy of a randomly drawn homologous gene in the focal individual. The probability of genetic identity between two randomly sampled genes in the FI is denoted $Q_\bullet$ while Q_j denotes the probability of identity of two homologous genes, one sampled from the FI and the other from a class j actor. The derivatives of the fitness function with respect to phenotypes are evaluated at $(z_\bullet = ... = z_j = ... = z)$ where z is a candidate evolutionary stable strategy (ESS). When $\Delta W_{\mathrm{IF}} > 0$, selection favours the behaviour and the candidate ESS is found at $\Delta W_{\mathrm{IF}} = 0$. Here, our goal is not to find the evolutionary stable level of helping but to establish the conditions under which selection favours helping.

Hamilton's rule

For the model presented in the main text, we assumed that only three classes of actors affect the FI's fitness. These are the FI itself, individuals of the class defined as related (with average phenotype z_{d}) and individuals that are encountered at random in the population (with average phenotype z_0). The probabilities of genetic identity between the FI and these two classes of actors are designated by Q_{d} and Q_0

for, respectively, an individual of the related class and the other class. From eq. 2, the inclusive fitness effect is then given by

$$\Delta W_{\mathrm{IF}} = \frac{\partial w}{\partial z_\bullet} Q_\bullet + \frac{\partial w}{\partial z_{\mathrm{d}}} Q_{\mathrm{d}} + \frac{\partial w}{\partial z_0} Q_0. \tag{3}$$

Using the property that the partial derivatives of the fitness function add up to zero (Rousset, 2004, pp. 96), we rearrange the inclusive fitness effect by substituting $\partial w/\partial z_0 = -(\partial w/\partial z_{\mathrm{d}} + \partial w/\partial z_\bullet)$ into eq. 3. Hence

$$\begin{aligned} \Delta W_{\mathrm{IF}} &\propto \frac{\partial w}{\partial z_\bullet} + \frac{\partial w}{\partial z_{\mathrm{d}}} r \\ &\propto rb' - c', \end{aligned} \tag{4}$$

where $\partial w/\partial z_\bullet \equiv -c'$ is the effect of the behaviour of the FI on its fitness, $\partial w/\partial z_{\mathrm{d}} \equiv b'$ is the effect of the behaviour of individuals of the related class on the fitness of the FI and $r \equiv (Q_{\mathrm{d}} - Q_0)/(Q_\bullet - Q_0)$ is the coefficient of relatedness that comes as a ratio of differences of probabilities of genetic identity ($\propto$ means proportional to). This coefficient of relatedness measures the extent to which an individual of the related class is more likely to bear genes identical in state with the FI than is an individual taken at random (but excluding the FI) from the population. The selective pressure on the trait is positive ($\Delta W_{\mathrm{IF}} > 0$) when

$$rb' - c' > 0 \tag{5}$$

is satisfied, which is Hamilton's rule. Accordingly, Hamilton's rule provides a condition for selection on a trait in a particular simple demographic situation, that is, when three classes of actors are affecting the trait under selection.

Multilevel selection

One could equally well describe the fitness of the FI as $w = gf$, where g is the expected number of offspring of the focal group that reach adulthood and f is the focal individual's share of these offspring. Then, the inclusive fitness effect given by

eq. 4 is equivalently given by

$$\Delta W_{\mathrm{IF}} \propto \underbrace{\left(\frac{\partial f}{\partial z_\bullet} + r\frac{\partial f}{\partial z_{\mathrm{d}}}\right)\frac{1}{N}}_{\partial f} + \underbrace{\left(\frac{\partial g}{\partial z_\bullet} + r\frac{\partial g}{\partial z_{\mathrm{d}}}\right)N}_{\partial g}, \tag{6}$$

where N is the number of individuals in the group (Rousset, 2004, p. 121). The first term in this equation (∂f) measures the effects of all actors in the focal group on the FI's share of the offspring of that group. This effect is in fact equivalent to the first order effect of selection on the covariance between individual trait value and relative fitness of the Price equation, which is interpreted as a component of selection within groups (Frank, 1998). The second term (∂g) measures the effects of all actors on the total number of offspring of the focal group that reach adulthood. This term is in fact the first order effect of selection on the covariance between group trait value and fitness of the Price equation, which is interpreted as the component of selection between groups. Importantly, both components (within and between-group selection) involve relatedness and thus a kin selection components.

2. Selection on helping in a randomly mixing population

For simplicity, we assumed a randomly mixing (i.e., panmictic) population of very large size in the model presented in the main text. The timing of the life-cycle was the following: (1) Repeated interactions occur between pairs of adult individuals. With probability x an individual interact repeatedly with an individual of the related class. With complementary probability $(1-x)$, the stream of repeated interactions occur with an individual encountered at random in the population. Under indirect reciprocity, such streams of interactions occur with different individuals from the respective classes. (2) Each adult produces a large number of juveniles depending on the costs and benefits of social interactions and then dies. (3) Competition occurs with the effect of regulating the population to a constant size.

Under such a life-cycle, the fitness function of a FI (eq. 2 of the main text) can

be written

$$w = \frac{xF_{\bullet,d} + (1-x)F_{\bullet,0}}{F_0}, \tag{7}$$

where $F_{\bullet,j}$ is the relative fecundity of the FI when interacting with an individual of class j and F_0 is the expected relative fecundity of an individual randomly sampled from the population. From the assumption of a very large population size (say infinite), F_0 is independent of the FI's fecundity and depends only on the interaction of two individuals bearing the same average phenotype z_0 in our pairwise interaction setting. Since the population is infinite and randomly mixing we also have $Q_0 = 0$ in the relatedness coefficient of eq. 5 and the model corresponds to Hamilton's original model of interaction among family members (family structured population).

As explained in the main text, the fecundity of each individual depends on successive rounds of pairwise interactions between individuals. The number (T) of rounds of interaction (1,2,3,...) for each individual is assumed to follow a Geometric distribution with parameter ω and the fecundity of an individual being the sum of the payoff of all rounds of interaction. Then, from the assumptions that costs and benefits are linear functions of investment into helping, the average relative fecundity of the FI interacting repeatedly with an individual of class j is

$$\begin{aligned} F_{\bullet,j} &= 1 + \sum_{T=1}^{\infty}(1-\omega)\omega^{T-1}\sum_{t=1}^{T}\left[\zeta B I_{\bullet,j}(t) + B(1-\zeta)I_{j,\bullet}(t) - C I_{\bullet,j}(t)\right] \\ &= 1 + \sum_{t=1}^{\infty}\omega^{t-1}\left[\zeta B I_{\bullet,j}(t) + B(1-\zeta)I_{j,\bullet}(t) - C I_{\bullet,j}(t)\right], \end{aligned} \tag{8}$$

where $I_{\bullet,j}(t)$ designates the level of investment of the FI into helping at round t when interacting with an individual of class j and $I_{j,\bullet}$ is the level of investments into helping of its partner at that round. The second equality is eq. 1 of the main text and follows by noting that $\sum_{T=1}^{\infty}\sum_{t=1}^{T} g(t,T) = \sum_{t=1}^{\infty}\sum_{T=t}^{\infty} g(t,T)$. The relative fecundity of an individual randomly sampled from the population is given by

$$F_0 = 1 + \sum_{t=1}^{\infty}\omega^{t-1}\left[\zeta B I_{0,0}(t) + B(1-\zeta)I_{0,0}(t) - C I_{0,0}(t)\right], \tag{9}$$

where $I_{0,0}(t)$ is the level of investment into helping at round t of an individual randomly sampled from the population when interacting with another individual randomly sampled from the population.

Using the same approach we also investigated the consequences of changing the functional relationships between investment into helping and fecundity (e.g., multiplicative streams of payoffs or non-linear cost and benefits) and the impact of different types of distribution of the number of rounds. These analyses reveal that the general conclusions are robust and not directly influenced by the assumed setting (L.Lehmann, unpublished results).

Helping and direct reciprocity

The inclusive fitness effect (ΔW_{IF}) of both the initial move (τ) and response slope (β) can be explicitly expressed as a function of the model's parameters (i.e., memory m and probability of interacting again ω). This is done by solving the system of equations

$$I_{i,j}(t) = m\beta_i I_{j,i}(t-1) \tag{10}$$

describing the investment into helping $I_{i,j}(t)$ of an individual of class i playing with an individual of class j at round t for three pairs of interacting actors: the FI and an individual of the related class; the FI and a randomly sampled individual from the population; and two randomly sampled individuals from the population (see eq. 3 of the main text). The initial conditions of these equations are given by the initial moves $I_{i,j}(1) = \tau_i$, which are the investment into helping of actors at the first round of interaction. Solving these equations gives the relative fecundity of the FI (eq. 8) when interacting repeatedly with an individual of class j

$$F_{\bullet,j} = 1 + B\left[\zeta\frac{\tau_\bullet + \omega m\beta_\bullet\tau_j}{1-\omega^2m^2\beta_j\beta_\bullet} + (1-\zeta)\frac{\tau_j + \omega m\beta_j\tau_\bullet}{1-\omega^2m^2\beta_j\beta_\bullet}\right] - C\left[\frac{\tau_\bullet + \omega m\beta_\bullet\tau_j}{1-\omega^2m^2\beta_j\beta_\bullet}\right] \tag{11}$$

and the relative fecundity of a randomly sampled individual in the population (eq. 9):

$$F_0 = 1 + (B - C)\left[\frac{\tau_0 + \omega m \beta_0 \, \tau_0}{1 - \omega^2 m^2 \beta_0^2}\right]. \tag{12}$$

Substituting the relative fecundities (eq. 11 and eq. 12) into the fitness function (eq. 7 here or eq. 2 in the main text) allows us to evaluate the effect of the behaviour of all actors on the FI's fitness. The effect of the FI on its own fitness when helping a partner at the first round is

$$\left.\frac{\partial w}{\partial \tau_\bullet}\right|_{\tau_\bullet = \tau_0 = \tau_d = \tau} = \frac{\zeta B + (1-\zeta)\,\omega m \beta B - C}{(1 + (B-C)\tau - m\beta\omega)\,(1 + \omega m \beta)}. \tag{13}$$

The effect of an individual of the related class on the FI's fitness by helping the FI at the first round reads

$$\left.\frac{\partial w}{\partial \tau_d}\right|_{\tau_\bullet = \tau_0 = \tau_d = \tau} = \frac{x\,[(1-\zeta)\,B + \omega m \beta(\zeta B - C)]}{(1 + (B-C)\tau - m\beta\omega)\,(1 + \omega m \beta)}. \tag{14}$$

Similarly, the effect of the FI on its fitness when responding to the investment into helping of its partner is

$$\left.\frac{\partial w}{\partial \beta_\bullet}\right|_{\beta_\bullet = \beta_0 = \beta_d = \beta} = \frac{\omega m \tau\,[\zeta B + (1-\zeta)\,\omega m \beta B - C]}{(1 + (B-C)\tau - m\beta\omega)\left(1 - (\omega m \beta)^2\right)}. \tag{15}$$

Finally, the effect of an individual of the related class on the fitness of the FI by reciprocating the help of the FI is

$$\left.\frac{\partial w}{\partial \beta_d}\right|_{\beta_\bullet = \beta_0 = \beta_d = \beta} = \frac{\omega m \tau x\,[(1-\zeta)\,B + \omega m \beta(\zeta B - C)]}{(1 + (B-C)\tau - m\,\beta\omega)\left(1 - (\omega m \beta)^2\right)}. \tag{16}$$

Since the effects on the FI's fitness of the initial move and the response slope are proportional to each other, the condition for the spread of both helping behaviours is given equivalently by

$$\underbrace{\zeta B - C + (1-\zeta)\,\omega m \beta B}_{-c'} + r\,\underbrace{x\,[(1-\zeta)\,B + \omega m \beta(\zeta B - C)]}_{b'} > 0. \tag{17}$$

This condition is obtained by applying Hamilton's rule (eq. 5) for both evolving traits (τ and β) and simplifying (for simplicity of notations, -c' and b' designates effects

on fitness up to a constant of proportionality). The effect of the behaviour of the FI on its fitness ($-c$) depends on two components. The first is the net effect of the act of helping on its fecundity, which involves the cost of helping $-C$ and the benefit ζB that directly return to him. The other stems from the helping reciprocated by the partner as a result of the FI investing into helping. This benefit depends on the cooperative tendency β of the partner, here the average response slope in the population. The effect of the behaviour of the FI on the fitness of its related partner (b) also depends on two components, the benefit $(1-\zeta)B$ received by the partner as a result of the helping of the FI, and the net effect $(\zeta B - C)$ on the partner fecundity resulting from the partner investing into helping to reciprocate the helping expressed by the FI.

In a population where there are no direct benefits ($\zeta = 0$) and where no individual initially express any helping, the inclusive fitness effects of the initial move and the response have to be evaluated at zero ($\tau_\bullet = \tau_0 = \tau_d = \tau = 0$ and $\beta_\bullet = \beta_0 = \beta_d = \beta = 0$). In such a resident population, the initial move and the response evolve when the condition

$$-C + rxB > 0 \tag{18}$$

is satisfied. Accordingly, the evolution of helping through reciprocal interactions takes off only in the presence of preferential interaction among close kin selection and is therefore altruistic ($-c' = -C < 0$ and $b' = xB > 0$).

Helping and indirect reciprocity

We consider here a simple situation of the evolution of helping under indirect reciprocity. Following previous analyses (Nowak and Sigmund, 1998) we assume that the level of investment into helping of an individual is a linear function of its partner's image score which can take only two values: *good* or *bad*. In this situation, investment into helping at any round is assumed to depend on the evolving response

slope α (varying between 0 and 1) on the partner's probability g of having a good image score. The probability that an individual has a good image score at a given round depends on two events. First, we assume that an observer of the behaviour of that individual at the previous round assigns an image score to that individual based on its investment into helping at that round (Ohtsuki and Iwasa, 2004). We also assume that assignment errors occur (Ohtsuki and Iwasa, 2004), and we denote by q the probability that an observer correctly attributes the image score to the player he observes. Hence, $1-q$ can be interpreted as the probability that the observer mistakenly assigns a bad image score when he should assign a good one and we do not consider (for simplicity) the situation where the observer assigns a good image score when he should assign a bad one. The strategy described here is of the "Scoring" type in the classification of Brandt and Sigmund (2004).

We posit that the FI encounters randomly its partner ($x = 0$) and that it does not interact twice with the same partner. Then, the investment of the partner of the FI at round t is

$$I_{0,\bullet}(t) = \alpha_0 g_\bullet(t), \tag{19}$$

which depends on the response slope (α_0) of an individual sampled at random from the population. The image score of the FI is given by

$$\begin{aligned} g_\bullet(t) &= qI_{\bullet,0}(t-1) \\ &= q\alpha_\bullet g_0(t-1). \end{aligned} \tag{20}$$

because the investment of the FI into helping a randomly encountered partner at round t is

$$I_{\bullet,0}(t) = \alpha_\bullet g_0(t). \tag{21}$$

In this equation, $\alpha_\bullet$ is the response slope of the FI and $g_0(t)$ is the probability that its partner has a good image score at that round t. This image score obeys the

recursion

$$g_0(t) = q\alpha_0 g_0(t-1) \tag{22}$$

because in a population of very large size, the average reputation dynamics is independent of the reputation of the FI. All these equations can be solved once the image scores at the initial moves are known. For simplicity, we assume that each individual has a good image score at the initial move (i.e., $g_0(1) = 1$ and $g_\bullet(1) = 1$).

Following the same stream of calculations as in the previous section, we can evaluate the relative fecundity of the FI (eq. 8) as

$$F_{\bullet,0} = 1 + B\left[\zeta\left(\frac{\alpha_\bullet}{1-\omega q\tau_0}\right) + (1-\zeta)\left(\alpha_0 + \frac{\omega q\alpha_0\alpha_\bullet}{1-\omega q\alpha_0}\right)\right] - C\left[\frac{\alpha_\bullet}{1-\omega q\alpha_0}\right] \tag{23}$$

and the relative fecundity of a randomly sampled individual from the population (eq. 9) as

$$F_0 = 1 + (B-C)\left[\frac{\alpha_0}{1-\omega q\alpha_0}\right]. \tag{24}$$

Because we assumed random interactions ($x = 0$), the direct fitness of the FI (eq. 7) is simply $w = F_{\bullet,0}/F_0$ and the inclusive fitness effect of helping is given directly by the effect of the FI on its fitness

$$\Delta W_{\mathrm{IF}} = \left.\frac{\partial w}{\partial \alpha_\bullet}\right|_{\alpha_\bullet=\alpha_0=\alpha_d=\alpha} = \frac{\zeta B + (1-\zeta)\,\omega q\alpha B - C}{(1 + (B-C)\alpha - q\alpha\omega)}. \tag{25}$$

In the absence of direct benefits ($\zeta = 0$), helping spreads through indirect reciprocity when

$$\omega q\alpha B - C > 0. \tag{26}$$

This equation is consistent with the results of Nowak and Sigmund (1998). Indeed, when the number of rounds of interactions is infinite ($\omega \to 1$) and when the response slope of the partners (α_0) is equal to one, the condition for helping to evolve is given

by $qB - C > 0$, which is equivalent to **?**, eq. 59f Nowak and Sigmund (1998). Inequality 26 has also a similar form as ineq. 6 in the main text. Indeed, α is akin to β so that cooperation can spreads only if interacting individuals have an initial tendency to be cooperative. Similarly, q is akin to m therefore requiring that individuals can evaluate the cooperative tendency of their partners for cooperation to evolve. Finally, cooperation can spread only if individuals are engaged in several rounds of reciprocal interactions with partners (i.e., $\omega > 0$), but where the partners are different at each round under indirect reciprocity. Accordingly, the main difference between direct and indirect reciprocity lies in the source of information individuals have to evaluate the cooperative tendency of their partner. Our model can be interpreted as a particular case encapsulated in the setting of Ohtsuki and Iwasa (2004) who consider very generally the co-evolution of reputation and cooperation.

3. Selection on helping in a geographically structured population

In order to illustrate the action of kin selection in a structured population, we consider in this section two situations of the evolution of helping where individuals interact randomly within demes connected by dispersal. First, we present a generalization of the overlapping generation model of Taylor and Irwin (2000), which takes isolation by distance into account in a manner similar to Rousset (2004, p. 124) in the absence of such overlapping generations. Second, we present a model where demes can fluctuate between two different sizes: low and high number of individuals and where helping can increase the probability of occurrence of the deme with the high density of individuals.

Helping and overlapping generations

We assume that individuals are haploid ($Q_\bullet = 1$) and that the population lies on a one-dimensional habitat where n_d demes of finite size N are regularly arrayed on a circle (circular lattice). Starting from a focal deme as origin, the different demes

can be numbered positively by moving clockwise or negatively by moving counterclockwise to represent distance between demes. The life-cycle in the population is the following: (1) Adult individuals express a helping act at a direct cost C to themselves. This act generates a benefit B that is shared equally among all other individuals in the deme. (2) Each adult then produces a large number (infinite) of juveniles and has a probability s to survive to the next breeding season. (3) Each juvenile disperses independently from each other juvenile with probability d_i to a deme at a distance i from the natal deme (the probability of staying in the natal patch is denoted d_0). The dispersal distribution is identical for all juveniles, moving clockwise or counterclockwise, and for all demes (isotropic dispersal). The resulting dispersal distribution encompasses a large class of population structures. (4) Regulation occurs. The proportion of juveniles competing in a deme reaching adulthood is $(1-s)N$, which corresponds to the average number of empty breeding spots in a deme resulting from the death of adult individuals.

The fitness function of a focal individual under this life-cycle reads

$$w = s + (1-s)\left[\sum_{i=0}^{n_{\mathrm{d}}-1} d_i \frac{1 + B\tau_0^{\mathrm{D}} - C\tau_\bullet}{\sum_{j=0}^{n_{\mathrm{d}}-1} d_{i-j}(1 + (B-C)\tau_j^{\mathrm{R}})}\right], \tag{27}$$

where $\tau_\bullet$ is the phenotype of the FI. The superscript in τ_0^{D} emphasizes that the average phenotype in the focal deme is computed after dispersal by excluding the FI. By contrast, the superscript in τ_j^{R} emphasizes that all individuals (including the FI) are taken into account when computing the average phenotype in a deme at distance j. We have $\tau_j^{\mathrm{R}} = \tau_j^{\mathrm{D}}$ except that $\tau_0^{\mathrm{R}} = 1/N \times \tau_\bullet + (N-1)/N \times \tau_0^{\mathrm{D}}$.

This inclusive fitness effect for this life-cycle is from eq.(2)

$$\Delta W_{\mathrm{IF}} = \frac{\partial w}{\partial \tau_\bullet} + \sum_{l=0}^{n_{\mathrm{d}}-1} \frac{\partial w}{\partial \tau_j} Q_j^{\mathrm{D}}, \tag{28}$$

where Q_j^{D} is the probability of genetic identity between two different individuals sampled *without* replacement at distance j. Since we are only interested in the selective pressure on helping when the population is initially filled with individuals

that do not express the act, we evaluate the partial derivatives at $\tau_\bullet = ... = \tau_j = ... = 0$ (Rousset, 2004, p. 124). Then, substituting the fitness function (eq. 27) into the inclusive fitness effect (eq. 28), we obtain after rearrangements

$$\Delta W_{\mathrm{IF}} = (1-s)\left[-C + BQ_0^{\mathrm{D}} - (B-C)\sum_{i=0}^{n_{\mathrm{d}}-1}\sum_{j=0}^{n_{\mathrm{d}}-1} d_i d_{i-j}\, Q_j^{\mathrm{R}}\right], \tag{29}$$

where the probability of genetic identity between two adults randomly sampled *with* replacement at distance j is $Q_j^{\mathrm{R}} = Q_j^{\mathrm{D}}$ except that $Q_0^{\mathrm{R}} = 1/N + (N-1)/NQ_0^{\mathrm{D}}$. The effect of the behaviour of the FI on its own fitness can be obtained by replacing Q_0^{R} by $1/N$ and Q_0^{D} by 0 in the formula for the inclusive fitness effect, hence

$$\left.\frac{\partial w}{\partial \tau_\bullet}\right|_{\tau_\bullet=...=\tau_j^{\mathrm{D}}=...=0} = -(1-s)\left[C + \frac{(B-C)}{N}\sum_{i=0}^{n_{\mathrm{d}}-1} d_i^2\right], \tag{30}$$

which is a net fitness cost and where the second term in brackets is the increase in competition faced by the offspring of the FI, which results from the help provided to neighbours. This increase in competition depends on the probability of the offspring of the FI dispersing into the same deme as an offspring produced in the focal deme.

We can solve the inclusive fitness effect in closed form if we know the stationary probabilities of genetic identity. Combining previous analyses of kin selection theory with overlapping generations (Taylor and Irwin, 2000, eq. A2) and classical analyses of isolation by distance models (Rousset, 2004, eq. 3.46), we find that the equilibrium probability of genetic identity between two individuals randomly sampled without replacement after dispersal in the same deme is

$$Q_0^{\mathrm{D}} = s^2 Q_0^{\mathrm{D}} + 2s(1-s)\sqrt{\gamma}\sum_{l=0}^{n_{\mathrm{d}}-1} d_l Q_l^{\mathrm{R}} + (1-s)^2\gamma\sum_{i=0}^{n_{\mathrm{d}}-1}\sum_{j=0}^{n_{\mathrm{d}}-1} d_i d_{i-j} Q_j^{\mathrm{R}}, \tag{31}$$

where $\gamma \equiv (1-\mu)^2$ and μ is the mutation rate. From this equation, we substitute

$$\sum_{i=0}^{n_{\mathrm{d}}-1}\sum_{j=0}^{n_{\mathrm{d}}-1} d_i d_{i-j} Q_j^{\mathrm{R}} = \frac{Q_0^{\mathrm{D}}(1-s^2)}{\gamma(1-s)^2} - \frac{2s}{(1-s)\sqrt{\gamma}}\sum_{l=0}^{n_{\mathrm{d}}-1} d_l Q_l^{\mathrm{R}} \tag{32}$$

into the inclusive fitness effect (eq. 29). Assuming an infinite number of demes ($n_{\mathrm{d}} \to \infty$), no mutations ($\gamma \to 1$) and using the results of Rousset (2004, pp. 46-52)

on Fourier analysis, we find after rearranging that

$$\Delta W_{\mathrm{IF}} \propto \frac{2s\mathcal{L}_0(\mathcal{F})}{(1-s)N}(B-C) - C. \tag{33}$$

The function $\mathcal{L}_0(\mathcal{F}) \equiv \frac{1}{\pi}\int_0^{\pi} \mathcal{F}(y)dy$ is the inverse Fourier transform at zero distance of

$$\mathcal{F}(y) = \frac{(1+s)\psi(y) - 2sd_0}{(1+s) + (1-s)\psi(y)}, \tag{34}$$

where $\psi(y) \equiv \sum_j d_j e^{ijy}$ is the characteristic function of the dispersal distribution (A supplementary Mathematica package that details the derivation is available on request). The function $\mathcal{L}_0(\mathcal{F})$ is positive and depends on the shape of the dispersal kernel with the effect that altruism spreads when

$$\frac{2s\mathcal{L}_0(\mathcal{F})}{(1-s)N}(B-C) - C > 0. \tag{35}$$

Hence, irrespectively of the structure of the population, altruism is favoured only if individuals have a positive probability of surviving from one breeding season to the next ($s > 0$). In this situation, altruistic interactions are favoured because they can occur between parents and their offspring. In the absence of overlapping generation ($s = 0$) the selective pressure on helping reduce to $-C$, which cannot be positive whatever the structure of the population. That overlapping generations is a mechanism promoting the evolution of altruism and cooperation under limited dispersal through the action of kin selection was first formally demonstrated by Taylor and Irwin (2000); Irwin and Taylor (2001) and has been repeatedly observed in the literature (Nowak *et al.*, 1994; Van Balen and Rand, 1998; Koella, 2000; Le Galliard *et al.*, 2003; Hauert and Doebeli, 2004).

The geometric distribution of dispersal ($d_i \equiv (1-d_0)(1-p)^{i-1}p$ for $i > 0$) allows us to investigate the effect of a continuum of spatial structures on the evolution of altruism ranging from the stepping-stone model of dispersal ($p \to 1$ hence $d_1 \to (1-d_0)$ and $d_j \to 0$ for $j > 1$) to Wright's island model of dispersal ($p \to 0$ hence d_1=d_2=...=0). Under the geometric distribution of dispersal, the spatial structure

that minimizes the deleterious effect of kin competition is that which maximizes $\mathcal{L}_0(\mathcal{F})$ with respect to p. The spatial structure the most favourable for altruism to evolve is Wright's island model of dispersal ($p \to 0$), that is when $\psi(y) \to m_0$ and $\mathcal{L}_0(\mathcal{F}) \to \mathcal{F}$. Accordingly,

$$\frac{2sd_0}{N(1+s+(1-s)d_0)}(B-C)-C>0, \tag{36}$$

which is the condition for the spread of altruism given by Taylor and Irwin (2000, eq. A10). Alternatively, this result can also be found by directly applying Hamilton's rule.[1]

Exact results can also be established for a population structured into a finite number of demes (n_d) but they are more cumbersome. When such a population behaves as a single panmictic unit, the direction of selection on altruism becomes independent of the survival rate (s) and depends only on the effect of the FI on its fitness. Hence,

$$\Delta W_\mathrm{IF} = \left.\frac{\partial w}{\partial \tau_\bullet}\right|_{\tau_\bullet=\tau_0=\tau_d=0} = (1-s)\left(-C-\frac{B-C}{Nn_\mathrm{d}}\right), \tag{37}$$

which highlights that the smaller the total population size (Nn_d), the stronger the selection against helping. The reason for this is that the by helping neighbours, the FI not only decreases its own fecundity but also increases the competition faced by its own offspring. This result was established by Rousset (2004, eq. 7.21) in the

[1]Because migration is random in Wright's island model of dispersal ($d_1{=}d_2{=}d_3{=}d_j$ for $j \neq 0$), individuals from all demes (excluding those of the deme of the FI) bear on average the same phenotype ($\tau_1^\mathrm{D}{=}\tau_2^\mathrm{D}{=}\tau_3^\mathrm{D} = \tau_j^\mathrm{D}$ for $j \neq 0$). One can then consider that there are only three classes of individuals in the population and affecting the fitness of the FI (eq. 27): the FI itself (phenotype $\tau_\bullet$), its neighbours in its deme (with average phenotype τ_0^D) and individuals from different demes (with average phenotype τ_1^D). The condition for the spread of altruism (eq. 36) is then equivalently found by applying Hamilton's rule $rb'+c'>0$, where $\partial w/\partial\tau_\bullet \equiv -c'$ is the effect of the behaviour of the FI on its fitness, $\partial w/\partial\tau_0^\mathrm{D} \equiv b'$ is the effect of the behaviour of the other individuals in the same deme of the FI on its fitness and $r = F_\mathrm{ST}$, where F_ST is Wright's measure of population structure.

absence of overlapping generations and by Nowak *et al.* (2004, eq. 2) and Wild and Taylor (2004, eq. 4.4) in the presence of overlapping generations.

Helping and fluctuating demography

Here, we consider a haploid population following an infinite island model of dispersal where each deme can fluctuate between two different states, one characterised by a high number N_{h} of individuals surviving to adulthood and the other characterized by a smaller number N_{l} of individuals reaching adulthood. We assume that helping between individuals within a deme affects the transition probabilities between the deme states through habitat and/or resource engineering, in a way that increases the probability of occurrence of the state with the highest number of individuals. Helping thus exerts an effect on the subsistence of all offspring in the descendant generation. We assume that a focal deme is either set to size N_{h} with probability τ_0^{R}, which is the average level of helping of the parental generation, or to size N_{h} with complementary probability $1-\tau_0^{\mathrm{R}}$. Changes in deme size are independent of their size in the parental generation but are dependent on the helping behaviour in that generation. The life-cycle is the following: (1) Reproduction occurs. Each individual produces an infinite number of juveniles. Helping reduces the relative fecundity of actors by a factor C, which varies linearly with the investment into helping. All adults die. (2) Juveniles disperse independently from each other with probability d to another deme (d is equivalent to $1-d_0$ in the model of the preceding section). (3) Regulation occurs with the number of juveniles reaching adulthood in a deme being determined by the investment into helping of the parental generation residing in that deme.

In order to evaluate whether helping spreads under such a life-cycle with fluctuating demography we use the inclusive fitness effect (eq. 1), which is then given by two components: $\Delta W_{\mathrm{IF}} = S_f + S_{\mathrm{Pr}}$[2]. To calculate this selective pressure, we need

[2]In the infinite island model of dispersal with fluctuations of deme size, the explicit expressions

the fitness function

$$w_{\mathrm{p}}(n,n') = \frac{n'\,(1-d)\,(1-C\tau_{\bullet})}{n\,(1-d)\,(1-C\tau_0^{\mathrm{R}}) + N_{\mathrm{eq}}d(1-C\tau_1^{\mathrm{D}}),}, \tag{40}$$

which measures the FI's expected number of offspring reaching adulthood in the focal deme of size n' (N_{l} or N_{h}) in the offspring generation, which was of size n (N_{l} or N_{h}) in the parental generation. In this fitness function $N_{\mathrm{eq}} = \tau_1^{\mathrm{D}} N_{\mathrm{h}} + (1-\tau_1^{\mathrm{D}})N_{\mathrm{l}}$ is the equilibrium deme size in the population. We also need

$$w_{\mathrm{d}}(n,n',l) = \frac{n'd(1-C\tau_{\bullet})}{(n\,(1-d) + N_{\mathrm{eq}}d)(1-C\tau_1^{\mathrm{D}})}, \tag{41}$$

which is the expected number of offspring reaching adulthood, in demes of size n' in the offspring generation that were of size n in the parental generation, of a FI breeding in a deme of size l (Rousset and Ronce, 2004, e.g., eq. 31-32). We also need the transition probability matrix of the demography of the focal deme

$$\mathbf{P} = \begin{pmatrix} \Pr(N_{\mathrm{l}} \mid N_{\mathrm{l}}) & \Pr(N_{\mathrm{l}} \mid N_{\mathrm{h}}) \\ \Pr(N_{\mathrm{h}} \mid N_{\mathrm{l}}) & \Pr(N_{\mathrm{h}} \mid N_{\mathrm{h}}) \end{pmatrix} = \begin{pmatrix} 1-\tau_0^{\mathrm{R}} & 1-\tau_0^{\mathrm{R}} \\ \tau_0^{\mathrm{R}} & \tau_0^{\mathrm{R}} \end{pmatrix}, \tag{42}$$

of the inclusive fitness effect $\Delta W_{\mathrm{IF}} = S_f + S_{\mathrm{Pr}}$ are obtained from eq. 26 and eq. 27 in Rousset and Ronce (2004); which read

$$S_f = \sum_{n'}\sum_{n} \nu(n')\Pr(n' \mid n)\left[\frac{\partial f_{\mathrm{p}}(n,n')}{\partial z_{\bullet}} + \frac{\partial f_{\mathrm{p}}(n,n')}{\partial z_0^{\mathrm{R}}} Q_0^{\mathrm{R}}(n) \right.$$
$$\left. \sum_{l} \Pr(l)\left(\frac{\partial f_{\mathrm{d}}(n,n',l)}{\partial z_{\bullet}} + \frac{\partial f_{\mathrm{d}}(n,n',l)}{\partial z_0^{\mathrm{R}}} Q_0^{\mathrm{R}}(l)\right)\right] \Pr(n) \tag{38}$$

and

$$S_{\mathrm{Pr}} = \sum_{n'}\sum_{n} \nu(n') \frac{\partial \Pr(n' \mid n)}{\partial z_0^{\mathrm{R}}} f_{\mathrm{p}}(n,n') Q_0^{\mathrm{R}}(n) \Pr(n)\,, \tag{39}$$

where $\nu(n')$ is the relative reproductive value of a deme of size n, $\Pr(n' \mid n)$ is the forward transition probability of a deme of size n to a deme of size n' and $\Pr(n)$ is the stationary probabilities that a deme will be of size n (Rousset and Ronce (2004) use $\alpha(n) = \nu(n)\Pr(n)$ in their formalization, which is the reproductive value of *all* demes of size n). The selective pressures also depend on $f_{\mathrm{p}}(n,n') \equiv w_{\mathrm{p}}(n,n')n/n'$, which is the probability that a gene sampled in a focal deme of size n' in the offspring generation descend from the focal deme that was of size n in the parental generation and on $f_{\mathrm{d}}(n,n',l) \equiv w_{\mathrm{p}}(n,n',l)l/n'$, which is the probability that a gene, conditional on its parental deme being of size l, is sampled presently in a deme of size n' that was of size n in the parental generation.

where $\Pr(n' \mid n)$ is the forward transition probability of a deme of size n to a deme of size n'.

Evaluating the expressions of the selective pressure at $\tau_\bullet = \tau_0^{\mathrm{R}} = \tau_1^{\mathrm{D}} = 0$, that is, in a population where individuals do not initially express helping and which implies that demes are only of the smaller size, we find after simplification that

$$S_f = -C\nu(N_{\mathrm{l}})(1 - Q_0^{\mathrm{D}}(N_{\mathrm{l}})), \tag{43}$$

where $\nu(N_{\mathrm{l}})$ is the relative reproductive value of a deme of size N_{l} and $Q_0^{\mathrm{D}}(N_{\mathrm{l}})$ is the probability of genetic identity of two individuals sampled without replacement in such a deme (A supplementary Mathematica package that details the derivation is available on request). Thus, conditional on the realization of the demographic states, helping results in a loss of the number of adult offspring produced by the FI. The effect of helping on the demography of the focal deme is

$$S_{\mathrm{Pr}} = (\nu(N_{\mathrm{h}}) - \nu(N_{\mathrm{l}}))\,(1-d)Q_0^{\mathrm{R}}(N_{\mathrm{l}}), \tag{44}$$

where $Q_0^{\mathrm{R}}(N_{\mathrm{l}})$ is the probability of genetic identity of two individuals sampled with replacement in a deme of low density, which varies inversely with deme size. The effect of the behaviour on deme demography is a net benefit if the reproductive value of all offspring in a deme of high density exceeds the reproductive value of all offspring in a deme of low density (i.e., $\nu(N_{\mathrm{h}}) - \nu(N_{\mathrm{l}}) > 0$). Writing $N_{\mathrm{l}} = kN_{\mathrm{h}}$, so that a deme of low density is reduced relative to a deme of high density by a factor k (varying between 0 and 1), the difference between the reproductive values simplifies to

$$\nu(N_{\mathrm{h}}) - \nu(N_{\mathrm{l}}) = d\left(\frac{1}{k} - \frac{k}{1 - d\,(1-k)}\right), \tag{45}$$

which is null when both states result in the same deme size ($k = 1$) or when migration vanishes ($d = 0$). This difference varies directly with the difference in deme size because the contribution of a deme to the future of the population is increasing with its size in the present model.

Helping can spread under the present life-cycle when the inclusive fitness effect is positive ($\Delta W_{\mathrm{IF}} > 0$). After evaluation of the probabilities of genetic identity and some simplification, this inequality is satisfied when

$$\frac{(1-d)(1-k)(1-d+k)}{k(1-d(1-k))N_{\mathrm{l}}} - C(2-d) > 0. \tag{46}$$

The selective pressure on helping increases with an increased difference in the sizes of the two types of demes and decreases with an increasing number of individuals in the state with the lowest number of individuals. When helping out propagates its alternative, it results in an expansion of deme size. Is such helping altruistic? The answer to this question depends on the direct effects of the FI on its fitness. This effect can be evaluated by replacing $Q_0^{\mathrm{R}}(N)$ by $1/N$ in the formulae that are used to evaluate the inclusive fitness effect (ΔW_{IF}). Here, helping is altruistic when

$$\frac{(1-d)(1-k)(1-d+k)}{k(1-d(1-k))N_{\mathrm{l}}} - C\frac{1}{d}\left(1 - \frac{(1-d)^2}{N_{\mathrm{l}}}\right) < 0, \tag{47}$$

which thus depends on the value of the parameters, a situation that sometimes occurs under a constant demography as well (Rousset, 2004, p. 114). In the present model, helping spreads (i.e.,$\Delta W_{\mathrm{IF}} > 0$) and is altruistic when dispersal is low ($d \ll 1$) or when the difference in the sizes of both types of demes is small ($k \ll 1$).

References

Brandt, H. and K. Sigmund (2004). The logic of reprobation: assessment and action rules for indirect reciprocation. *Journal of Theoretical Biology* 231(4):475–486.

Frank, S. A. (1998). *Foundations of social evolution.* Princeton University Press, Princeton, NJ.

Hamilton, W. D. (1964). The genetical evolution of social behaviour. I. *Journal of Theoretical Biology* 7:1–16.

Hauert, C. and M. Doebeli (2004). Spatial structure often inhibits the evolution of cooperation in the snowdrift game. *Nature* 428(6983):643–646.

Irwin, A. J. and P. D. Taylor (2001). Evolution of altruism in stepping-stone populations with overlapping generations. *Theoretical Population Biology* 60(4):315–325.

Kirkpatrick, M., T. Johnson, and N. Barton (2002). General models of multilocus evolution. *Genetics* 161(1456):1727–1750.

Koella, J. C. (2000). The spatial spread of altruism versus the evolutionary response of egoists. *Proceedings of the Royal Society of London Series B-Biological Sciences* 267(1456):1979–1985.

Le Galliard, J., R. Ferrière, and U. Dieckmann (2003). The adaptive dynamics of altruism in spatially heterogeneous populations. *Evolution* 57(1):1–17.

Nowak, M., A. Sasaki, C. Taylor, and D. Fudenberg (2004). Emergence of cooperation and evolutionary stability in finite populations. *Nature* 428:646–650.

Nowak, M. A. and K. Sigmund (1998). The dynamics of indirect reciprocity. *Journal of Theoretical Biology* 194(4):561–574.

Nowak, M. A., S. Bonhoeffer, and R. M. May (1994). Spatial games and the maintenance of cooperation. *Proceedings of the National Academy of Sciences of the United States of America* 91(11):4877–4881.

Ohtsuki, H. and Y. Iwasa (2004). How should we define goodness? reputation dynamics in indirect reciprocity. *Journal of Theoretical Biology* 231(1):107–120.

Rousset, F. (2003). A minimal derivation of convergence stability measures. *Journal of Theoretical Biology* 221:665–668.

Rousset, F. (2004). *Genetic structure and selection in subdivided populations*. Princeton University Press, Princeton, NJ.

Rousset, F. and S. Billiard (2000). A theoretical basis for measures of kin selection in subdivided populations: finite populations and localized dispersal. *Journal of Evolutionary Biology* 13(5):814–825.

Rousset, F. and O. Ronce (2004). Inclusive fitness for traits affecting metapopulation demography. *Theoretical Population Biology* 142:1357–1362.

Taylor, P. (1990). Allele-frequency change in a class-structured population. *American Naturalist* 135:95–106.

Taylor, P. D. (1996). Inclusive fitness arguments in genetic models of behaviour. *Journal of Mathematical Biology* 34:654–674.

Taylor, P. D. and S. A. Frank (1996). How to make a kin selection model. *Journal of Theoretical Biology* 54(4):1135–1141.

Taylor, P. D. and A. J. Irwin (2000). Overlapping generations can promote altruistic behavior. *Evolution* 54(180):1135–1141.

Van Balen, M. and A. Rand (1998). The unit of selection in viscous populations and the evolution of altruism. *Journal of Theoretical Biology* 193:631–648.

Wild, G. and P. Taylor (2004). Fitness and evolutionary stability in game theoric models of finite populations. *Proceedings of the Royal Society of London Series B-Biological Sciences* 271:2345–2349.