

## 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:

$$\Delta W_{\text{IF}} = S_f + S_{\text{Pr}}, \quad (1)$$

(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,  $S_{\text{Pr}}$ , 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,  $S_{\text{Pr}} = 0$  and the inclusive fitness effect ( $\Delta W_{\text{IF}}$ ) 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  $w \equiv w(z_\bullet, \dots, z_j, \dots)$  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

$$\Delta W_{\text{IF}} = \frac{\partial w}{\partial z_\bullet} Q_\bullet + \sum_j \frac{\partial w}{\partial z_j} Q_j. \quad (2)$$

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 ( $z_\bullet = \dots = z_j = \dots = z$ ) where  $z$  is a candidate evolutionary stable strategy (ESS). When  $\Delta W_{\text{IF}} > 0$ , selection favours the behaviour and the candidate ESS is found at  $\Delta W_{\text{IF}} = 0$ . 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

$$\Delta W_{\text{IF}} = \frac{\partial w}{\partial z_{\bullet}} Q_{\bullet} + \frac{\partial w}{\partial z_{\text{d}}} Q_{\text{d}} + \frac{\partial w}{\partial z_0} Q_0. \quad (3)$$

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  $\partial w/\partial z_0 = -(\partial w/\partial z_{\text{d}} + \partial w/\partial z_{\bullet})$  into eq. 3. Hence

$$\begin{aligned} \Delta W_{\text{IF}} &\propto \frac{\partial w}{\partial z_{\bullet}} + \frac{\partial w}{\partial z_{\text{d}}} r \\ &\propto rb - c, \end{aligned} \quad (4)$$

where  $\partial w/\partial z_{\bullet} \equiv -c$  is the effect of the behaviour of the FI on its fitness,  $\partial w/\partial z_{\text{d}} \equiv b$  is the effect of the behaviour of individuals of the related class on the fitness of the FI and  $r \equiv (Q_{\text{d}} - Q_0)/(Q_{\bullet} - Q_0)$  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 ( $\Delta W_{\text{IF}} > 0$ ) when

$$rb - c > 0 \quad (5)$$

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

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

where  $N$  is the number of individuals in the group (Rousset, 2004, p. 121). The first term in this equation ( $\partial f$ ) 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 ( $\partial g$ ) 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

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

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  $\omega$  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

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

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

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

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 ( $\Delta W_{\text{IF}}$ ) of both the initial move ( $\tau$ ) and response slope ( $\beta$ ) can be explicitly expressed as a function of the model's parameters (i.e., memory  $m$  and probability of interacting again  $\omega$ ). This is done by solving the system of equations

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

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  $j$

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

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

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

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

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

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

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

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

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

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

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

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

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

This condition is obtained by applying Hamilton's rule (eq. 5) for both evolving traits ( $\tau$  and  $\beta$ ) and simplifying (for simplicity of notations,  $-c$  and  $b$  will be used

throughout the Target Review and this supplementary material to designate 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  $\zeta B$  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  $\beta$  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  $\alpha$  (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 ( $\alpha_0$ ) of an individual sampled at random from the population. The image score of the FI is given by

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

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

$$g_0(t) = q\alpha_0 g_0(t-1) \quad (22)$$

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

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

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

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

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

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

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

$$\omega q \alpha B - C > 0. \quad (26)$$

This equation is consistent with the results of Nowak and Sigmund (1998). Indeed, when the number of rounds of interactions is infinite ( $\omega \rightarrow 1$ ) and when the response slope of the partners ( $\alpha_0$ ) is equal to one, the condition for helping to evolve is given

by  $qB - C > 0$ , which is equivalent to eq. 59 of Nowak and Sigmund (1998). Inequality 26 has also a similar form as ineq. 6 in the main text. Indeed,  $\alpha$  is akin to  $\beta$  so that cooperation can spread 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

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

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

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

$$\Delta W_{\text{IF}} = \frac{\partial w}{\partial \tau_\bullet} + \sum_{l=0}^{n_d-1} \frac{\partial w}{\partial \tau_j} Q_j^D, \quad (28)$$

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  $\tau_{\bullet} = \dots = \tau_j = \dots = 0$  (Rousset, 2004, p. 124). Then, substituting the fitness function (eq. 27) into the inclusive fitness effect (eq. 28), we obtain after rearrangements

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

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

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

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

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

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

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

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

on Fourier analysis, we find after rearranging that

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

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

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

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

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

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 Baalen 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 ( $p \rightarrow 1$  hence  $d_1 \rightarrow (1 - d_0)$  and  $d_j \rightarrow 0$  for  $j > 1$ ) to Wright's island model of dispersal ( $p \rightarrow 0$  hence  $d_1 = d_2 = \dots = 0$ ). 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 ( $p \rightarrow 0$ ), that is when  $\psi(y) \rightarrow m_0$  and  $\mathcal{L}_0(\mathcal{F}) \rightarrow \mathcal{F}$ . Accordingly,

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

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.<sup>1</sup>

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,

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

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

---

<sup>1</sup>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 ( $\tau_1^D=\tau_2^D=\tau_3^D=\tau_j^D$  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  $\tau_0^D$ ) and individuals from different demes (with average phenotype  $\tau_1^D$ ). 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,  $\partial w/\partial \tau_0^D \equiv b$  is the effect of the behaviour of the other individuals in the same deme of the FI on its fitness and  $r = F_{\text{ST}}$ , where  $F_{\text{ST}}$  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  $\tau_0^R$ , which is the average level of helping of the parental generation, or to size  $N_l$  with complementary probability  $1 - \tau_0^R$ . 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:  $\Delta W_{IF} = S_f + S_{Pr}$ <sup>2</sup>. To calculate this selective pressure, we need

---

<sup>2</sup>In the infinite island model of dispersal with fluctuations of deme size, the explicit expressions

the fitness function

$$w_p(n, n') = \frac{n'(1-d)(1-C\tau_\bullet)}{n(1-d)(1-C\tau_0^R) + N_{\text{eq}}d(1-C\tau_1^D)}, \quad (40)$$

which measures the FI's expected number of offspring reaching adulthood in the focal deme of size  $n'$  ( $N_1$  or  $N_h$ ) in the offspring generation, which was of size  $n$  ( $N_1$  or  $N_h$ ) in the parental generation. In this fitness function  $N_{\text{eq}} = \tau_1^D N_h + (1 - \tau_1^D) N_1$  is the equilibrium deme size in the population. We also need

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

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

$$\mathbf{P} = \begin{pmatrix} \Pr(N_1 | N_1) & \Pr(N_1 | N_h) \\ \Pr(N_h | N_1) & \Pr(N_h | N_h) \end{pmatrix} = \begin{pmatrix} 1 - \tau_0^R & 1 - \tau_0^R \\ \tau_0^R & \tau_0^R \end{pmatrix}, \quad (42)$$

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

$$S_f = \sum_{n'} \sum_n \nu(n') \Pr(n' | n) \left[ \frac{\partial f_p(n, n')}{\partial z_\bullet} + \frac{\partial f_p(n, n')}{\partial z_0^R} Q_0^R(n) \right. \\ \left. \sum_l \Pr(l) \left( \frac{\partial f_d(n, n', l)}{\partial z_\bullet} + \frac{\partial f_d(n, n', l)}{\partial z_0^R} Q_0^R(l) \right) \right] \Pr(n) \quad (38)$$

and

$$S_{\text{Pr}} = \sum_{n'} \sum_n \nu(n') \frac{\partial \Pr(n' | n)}{\partial z_0^R} f_p(n, n') Q_0^R(n) \Pr(n), \quad (39)$$

where  $\nu(n')$  is the relative reproductive value of a deme of size  $n'$ ,  $\Pr(n' | n)$  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  $f_p(n, n') \equiv w_p(n, n')n/n'$ , 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  $f_d(n, n', l) \equiv w_p(n, n', l)l/n'$ , 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  $\Pr(n' | n)$  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  $\tau_{\bullet} = \tau_0^R = \tau_1^D = 0$ , 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

$$S_f = -C\nu(N_1)(1 - Q_0^D(N_1)), \quad (43)$$

where  $\nu(N_1)$  is the relative reproductive value of a deme of size  $N_1$  and  $Q_0^D(N_1)$  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

$$S_{Pr} = (\nu(N_h) - \nu(N_l)) (1 - d)Q_0^R(N_1), \quad (44)$$

where  $Q_0^R(N_1)$  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.,  $\nu(N_h) - \nu(N_l) > 0$ ). Writing  $N_l = kN_h$ , 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

$$\nu(N_h) - \nu(N_l) = d \left( \frac{1}{k} - \frac{k}{1 - d(1 - k)} \right), \quad (45)$$

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 ( $\Delta W_{\text{IF}} > 0$ ). After evaluation of the probabilities of genetic identity and some simplification, this inequality is satisfied when

$$\frac{(1-d)(1-k)(1-d+k)}{k(1-d(1-k))N_1} - C(2-d) > 0. \quad (46)$$

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  $Q_0^{\text{R}}(N)$  by  $1/N$  in the formulae that are used to evaluate the inclusive fitness effect ( $\Delta W_{\text{IF}}$ ). Here, helping is altruistic when

$$\frac{(1-d)(1-k)(1-d+k)}{k(1-d(1-k))N_1} - C\frac{1}{d}\left(1 - \frac{(1-d)^2}{N_1}\right) < 0, \quad (47)$$

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.,  $\Delta W_{\text{IF}} > 0$ ) 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 reciprocity. *Journal of Theoretical Biology* 231:475–486.
- 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* 428:643–646.
- Irwin, A. J. and P. D. Taylor (2001). Evolution of altruism in stepping-stone populations with overlapping generations. *Theoretical Population Biology* 60:315–325.
- Kirkpatrick, M., T. Johnson, and N. Barton (2002). General models of multilocus evolution. *Genetics* 161:1727–1750.
- 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* 267:1979–1985.
- Le Galliard, J., R. Ferrière, and U. Dieckmann (2003). The adaptive dynamics of altruism in spatially heterogeneous populations. *Evolution* 57:1–17.
- 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* 194:561–574.
- 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* 91:4877–4881.

- Ohtsuki, H. and Y. Iwasa (2004). How should we define goodness? reputation dynamics in indirect reciprocity. *Journal of Theoretical Biology* 231:107–120.
- 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* 13:814–825.
- 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* 54:1135–1141.
- Taylor, P. D. and A. J. Irwin (2000). Overlapping generations can promote altruistic behavior. *Evolution* 54:1135–1141.
- van Baalen, 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 theoretic models of finite populations. *Proceedings of the Royal Society of London Series B-Biological Sciences* 271:2345–2349.