

# The evolution of helping and harming on graphs: the return of the inclusive fitness effect

L. LEHMANN,\* L. KELLER† & D. J. T. SUMPTER‡

\*Morrison Institute for Population and Resource Studies, University of Stanford, Stanford, CA, USA

†Department of Ecology and Evolution, Biophore, University of Lausanne, Lausanne, Switzerland

‡Mathematics Department, Uppsala University, Uppsala, Sweden

## Keywords:

graph theory;  
harming;  
helping;  
kin selection;  
population structure.

## Abstract

Evolutionary graph theory has been proposed as providing new fundamental rules for the evolution of co-operation and altruism. But how do these results relate to those of inclusive fitness theory? Here, we carry out a retrospective analysis of the models for the evolution of helping on graphs of Ohtsuki *et al.* [Nature (2006) 441, 502] and Ohtsuki & Nowak [Proc. R. Soc. Lond. Ser. B Biol. Sci (2006) 273, 2249]. We show that it is possible to translate evolutionary graph theory models into classical kin selection models without disturbing at all the mathematics describing the net effect of selection on helping. Model analysis further demonstrates that costly helping evolves on graphs through limited dispersal and overlapping generations. These two factors are well known to promote relatedness between interacting individuals in spatially structured populations. By allowing more than one individual to live at each node of the graph and by allowing interactions to vary with the distance between nodes, our inclusive fitness model allows us to consider a wider range of biological scenarios leading to the evolution of both helping and harming behaviours on graphs.

## Introduction

Unlike traditional kin selection or inclusive fitness theory (Hamilton, 1964), which emphasizes how relatedness between individuals promotes helping behaviour, evolutionary graph theory emphasizes the importance of the spatial subdivision of populations (Santos & Pacheco, 2005; Nowak, 2006; Ohtsuki *et al.*, 2006; Ohtsuki & Nowak, 2006). Evolutionary graph theory models allow the effect of space on helping to be investigated in a broad context, while simultaneously making the models mathematically tractable (Lieberman *et al.*, 2005; Ohtsuki *et al.*, 2006; Ohtsuki & Nowak, 2006). Recently, a rule has been derived (Nowak, 2006; Ohtsuki *et al.*, 2006; Ohtsuki & Nowak, 2006) for the evolution of helping on graphs: ‘natural selection favours cooperation if the benefit  $B$  of the act, divided by the cost  $C$ , exceeds the average number  $k$  of neighbours’:

$$\frac{B}{C} > k. \quad (1)$$

This rule has been derived using pair approximations for regular isothermal graphs and tested numerically as a lower bound for the evolution of helping on more complicated graphs.

Does inequality 1 provides us with a new pathway to the evolution of helping behaviours on graphs? We think that this is not the case. Indeed, kin selection operates whenever interactions occur among relatives, that is, among individuals that are more likely to inherit a strategy from a common ancestor than are individuals sampled at random from the population (Hamilton, 1964, 1970, 1971). This may happen when interactions take place among members of a family or when the population is structured through limited dispersal. In both cases, relatives remain close to each other. Because dispersal occurs only to the nearest neighbours in populations structured according to evolutionary graph theory, interactions occur necessarily among relatives.

In this paper, we carry out a retrospective analysis of the models for the evolution of helping on graphs of Ohtsuki *et al.* (2006) and Ohtsuki & Nowak (2006), and

Correspondence: Laurent Lehmann, Morrison Institute for Population and Resource Studies, University of Stanford, Gilbert Building, Rm 16, 371 Serra Mall, Stanford, CA 94305-5020, USA.  
Tel.: 650 724 3570; fax: 650 725 8244;  
e-mail: lehmann@stanford.edu

generalize them by applying inclusive fitness theory for finite populations (Rousset & Billiard, 2000; Rousset, 2004, 2006). This allows us to recover the mathematical results of evolutionary graph theory for 'death–birth' and 'imitation' life cycles as examples of inclusive fitness theory for spatially structured populations. In a companion paper, Grafen (2007) also provides an inclusive fitness analysis of the 'death–birth' and 'birth–death' life cycles of evolutionary graph theory. The reason for our retrospective analysis is to illustrate how results obtained heuristically by pair approximations can be obtained exactly by using inclusive fitness theory and to provide a link between the results of evolutionary graph theory and those of inclusive fitness theory. Further, by allowing more than one individual to live at each node of the graph and by allowing interactions to vary with the distance between nodes, our models allow us to represent different biological scenarios leading to the evolution of both helping and harming behaviours on graphs.

## Model

### Life cycle

Consider a spatially structured population (or graph) consisting of  $n_d$  demes (or nodes), each of which is occupied by a constant number  $N$  of adult individuals. The population is therefore of total size  $N_T = n_d N$  and we assume that demes are linked by edges through which neighbouring individuals interact. Further, we assume that this population is spatially homogeneous, so that all nodes have the same number of neighbours (i.e. the same degree) and all interactions are symmetrical. With these assumptions, the population structure is said to be regular and isothermal (e.g. Lieberman *et al.*, 2005; Ohtsuki *et al.*, 2006; Ohtsuki & Nowak, 2006) and encompasses a circular lattice in one dimension and a torus in two dimensions (e.g. Malécot, 1975; Nagylaki, 1982; Taylor, 1992b; Epperson, 1999; Gandon & Rousset, 1999; Rousset & Billiard, 2000; Irwin & Taylor, 2001; Rousset, 2004; Rousset, 2006). 'Distances' on this graph are denoted by a vector  $\mathbf{j}$ , which is the coordinate (single or pair) of a deme relative to the position of a focal deme, so that  $\mathbf{j} = \mathbf{0} \equiv (0, 0)$  stands for the deme of a focal individual in two dimensions (i.e. focal deme).

We assume that each individual bears either a mutant 'helping' allele or a resident 'defector' allele. A focal individual bearing the mutant helping allele confers, at a direct cost to self, a benefit to each of its neighbours, in the same or in a different deme of the population. Specifically, the focal individual increments the fecundity (or pay-off) of the whole set of neighbours in a deme at vectorial distance  $\mathbf{j}$  by  $B_j$ , an action decrementing its own fecundity by  $C_j$ . An individual bearing the defector allele takes fecundity benefits but pays no cost.

The life cycle is punctuated by the following events. (1) Each individual in the population produces a large

number of juveniles proportional to the total of its benefits minus its costs. (2) Each juvenile in a deme disperses independently of the others with probability  $m_j$  to a deme at distance  $\mathbf{j}$  ( $\sum_j m_j = 1$ ). The dispersal distribution is assumed to be symmetric and identical for all demes (isotropic dispersal). (3) After dispersal, one adult individual is chosen at random from the whole population and dies (i.e. according to the Moran scheme of reproduction; Maruyama, 1974; Ewens, 2004; Ohtsuki *et al.*, 2006). One of the juveniles present at the deme that has the vacant site is chosen at random to occupy this site.

This life cycle is very close to the seminal kin selection model for 'environmental homogeneity' considered by Taylor (1992b) and analysed under finite deme number by Rousset (2004). The only difference is that these authors assumed that all individuals die during stage (3) of the life cycle (i.e. 'Wright–Fisher' scheme of reproduction; Ewens, 2004), whereas here exactly one individual dies per unit time (i.e. 'Moran' scheme of reproduction; Ewens, 2004). This difference will prove crucial for the maintenance of co-operation at a fecundity (pay-off) cost to the actor. If we further assume that all juveniles disperse (i.e. no philopatry  $m_0 = 0$ ) and that dispersal occurs only to the nearest neighbours, then the life cycle corresponds to the 'death–birth' protocol of evolutionary graph theory, where one individual is chosen at random to die and neighbours compete for the empty site proportional to their fitness (Lieberman *et al.*, 2005; Ohtsuki *et al.*, 2006; Ohtsuki & Nowak, 2006). Finally, to provide a linking model between the 'Moran' and the 'Wright–Fisher' process, we also consider in the Appendix a variant of the life-cycle where each adult individual survives independently with probability  $s$  to the next generation (i.e. 'Cannings' scheme of reproduction; Ewens, 2004).

### Inclusive fitness effect

Whether natural selection will favour the fixation of a mutant allele introduced as a single copy in a population fixed for the defector allele can be ascertained by evaluating the inclusive fitness effect (Rousset & Billiard, 2000; Rousset, 2004, 2006), which is the sum of the effects of the behaviour of a focal individual bearing the mutation on the fitness of all recipients in the population, each weighted by the relatedness of the focal individual to the recipient (Hamilton, 1964, 1970). For a mutant allele whose phenotypic effect on fitness deviates by a small magnitude from that of a resident allele (weak selection), the inclusive fitness effect is positive when

$$\frac{\partial w}{\partial z_{\bullet}} + \sum_{\mathbf{j}} \frac{\partial w}{\partial z_{\mathbf{j}}} R_{\mathbf{j}} > 0, \quad (2)$$

where  $w$  is the expected number of recruited offspring of the focal individual (i.e. its fitness) and the relatedness

coefficient is given by  $R_j = (Q_j - \bar{Q}) / (1 - \bar{Q})$ , where  $Q_j$  is the probability that a recipient (excluding the focal individual) sampled in a deme at distance  $\mathbf{j}$  from the focal deme bears a copy of a homologous gene (or meme) drawn in the focal individual, and  $\bar{Q} = \sum_j Q_j / n_d$  is the average probability of identity between pairs of homologous genes sampled in different individuals. The relatedness coefficient  $R_j$  measures the extent to which two individuals sampled at distance  $\mathbf{j}$  from each other are more (or less) likely to have inherited the same genes (or memes) from a recent common ancestor than two individuals sampled at random (but without replacement) from the population (see eqn 12 in the Appendix). The derivatives of  $w$  are the effects of actors on the fitness of the focal individual, the actors being the focal individual itself with phenotype denoted  $z_\bullet$ , and the other individuals in the population, where  $z_j$  is the average phenotype of individuals living in a deme at distance  $\mathbf{j}$  (excluding the focal individual for  $\mathbf{j} = \mathbf{0}$ ). Hence,  $-c \equiv \partial w / \partial z_\bullet$  represents the effect of the behaviour of a focal individual on its fitness; and  $b_j \equiv \partial w / \partial z_j$  can be interpreted in the 'inclusive fitness' way as the effect of the focal individual on the fitness of the whole set of recipients (but excluding himself) living in the deme at distance  $\mathbf{j}$  from the focal deme (Frank, 1998; Rousset, 2004; Grafen, 2006).

In interpreting the inclusive fitness effect, it is useful to note that the definition of the average probability of identity  $\bar{Q}$  implies that the average relatedness of the focal individual to other individuals from the population is equal to zero. Hence

$$\frac{1}{n_d} \sum_j R_j = 0, \quad (3)$$

which informs us that the total number of individuals that are positively related to the focal individual, each weighted by their relatedness to the focal individual, is exactly equal to the total relatedness weighted sum of individuals that are negatively related to him (see eqn 13 in the Appendix). Hence, provided an individual is positively related to some individuals in the population, it must be negatively related to some other individual(s) (e.g. Grafen, 2007). Thus, as a consequence of the inclusive fitness effect (eqn 2), if there exists a value of  $-c$  under which costly helping ( $b_j > 0$ ) can evolve towards positively related individuals, there must also exist another value of  $-c$  under which costly harming ( $b_j < 0$ ) can evolve towards negatively related individuals.

When populations are spatially structured, individuals also compete with positively related individual for local resources (Taylor, 1992a, b). The intensity of competition for resources is affected by helping behaviours because, by incrementing the fecundity of neighbours, an actor increases the number of offspring competing for the same breeding spots as its own offspring. Thus, in spatially structured populations, genetic closeness between indi-

viduals is necessarily associated with competitive closeness. This association means that the fitness effects ( $-c$  and the  $b_j$ s) will be a complicated function comprising the effects on fecundity  $C_j$  and  $B_j$ , and of the dispersal distribution (the  $m_j$ s). To evaluate these effects, we need an expression of the direct fitness function  $w$ .

### Fitness function

From the life-cycle assumptions, the relative fecundity of a focal individual is given by  $1 - \sum_k C_k z_\bullet + \sum_k B_k z_k$ , which depends on the baseline reproductive unit, on the cost of expressing helping and the benefits received by all other individuals in the population expressing the helping allele. A fraction  $m_i$  of the offspring of the focal individual then migrates to a deme at distance  $\mathbf{i}$  from the focal deme. These offspring enter with probability  $1/N_T$  in competition for a vacant breeding spot (i.e. an adult dies in the deme at distance  $\mathbf{i}$ ) with a fraction  $\sum_j m_{i-j} (1 + \sum_k B'_{j-k} z_k^R)$  of the total number of offspring produced in the population, where  $B'_j \equiv B_j$  except that  $B'_0 \equiv B_0 - \sum_i C_i$  because those individuals expressing helping also pay the cost of helping. The phenotype  $z_k^R$  denotes the average phenotype in deme  $\mathbf{k}$  (including the focal individual in the average for the focal deme at  $\mathbf{k} = \mathbf{0}$ ), whereby  $z_k^R = z_k$  except that  $z_0^R = z_0 / N + (N - 1)z_0 / N$ . Collecting all terms allows us to write the direct fitness of the focal individuals as

$$w = \frac{N_T - 1}{N_T} + \frac{1}{N_T} \sum_i m_i \frac{1 - \sum_k (C_k z_\bullet - B_k z_k)}{\sum_j m_{i-j} (1 + \sum_k B'_{j-k} z_k^R)}. \quad (4)$$

With this function, whose second term is equivalent to eqn 7.18 of Rousset (2004), we can now evaluate explicitly the effects of actors on the fitness of the focal individuals that are necessary to compute the gradient of selection (eqn 2).

## Results

### Effects of actor on the fitness of recipients

Substituting the direct fitness function (eqn 4) into the inclusive fitness effect (eqn 2), evaluating the partial derivatives at the phenotypic value of the defector allele, which does not express the helping behaviour ( $z_\bullet = z_0 = \dots = z_j = \dots = 0$ ), we find that the effect of the focal individual on its fitness is given by

$$-c = \frac{1}{N_T} \sum_j \left( -C_j - \frac{1}{N} B'_j P_j \right), \quad (5)$$

where the term inside the parentheses consists of two components. First, on the direct fecundity cost  $C_j$  for the focal individual when helping individuals in a deme at distance  $\mathbf{j}$  from the focal deme. Second, on the indirect cost stemming from the increase in competition faced by the focal individual's own offspring. This increased competition depends on the increment  $B'_j$  in fecundity

of individuals in deme  $\mathbf{j}$  times the probability  $P_j = \sum_i m_i m_{i-j}$  that an offspring of the focal individual competes against an offspring produced in deme  $\mathbf{j}$ .

The effect of the focal individual on the fitness of other individuals in its deme is given by

$$b_0 = \frac{1}{N_T} \left( B_0 - \frac{N-1}{N} \sum_j B'_j P_j \right), \quad (6)$$

which also consists of two components, the direct fecundity benefit  $B_0$  conferred by the focal individual to its deme mates and the indirect costs resulting from the focal individual helping individuals on all nodes of the graph and thereby augmenting the intensity of competition faced by the offspring of his deme mates. This indirect cost depends on the increment  $B'_j$  in the fecundity of individuals in deme  $\mathbf{j}$  times the probability  $P_j$  that an offspring of a focal individual's deme mate competes against an offspring produced in deme  $\mathbf{j}$ . Finally, the effect of the focal individual on the fitness of recipients living in a deme at distance  $\mathbf{k}$  from the focal deme is given by

$$b_k = \frac{1}{N_T} \left( B_k - \sum_j B'_{j-k} P_j \right). \quad (7)$$

This fitness effect depends on the direct fecundity benefit  $B_k$  the focal individual confers to individuals living in deme  $\mathbf{k}$  and on the indirect costs resulting from the focal individual helping individuals on all nodes of the graph and thereby augmenting the intensity of competition faced by the offspring of individuals residing in deme  $\mathbf{k}$ . This competition cost depends on the increment  $B'_{j-k}$  of the fecundity of individuals residing in a deme at distance  $\mathbf{j}$  from deme  $\mathbf{k}$  (hence at  $\mathbf{j} - \mathbf{k}$  from the focal deme) times the probability  $P_j$  that offspring of individuals residing at  $\mathbf{j}$  steps apart enter in competition with each other.

We have now established a relationship between the parameters of evolutionary graph theory and the models of inclusive fitness in spatially subdivided populations. Equations 5–7 illustrate that the fitness costs and benefits ( $-c$  and the  $b_{js}$ , i.e. the parameters of inclusive fitness theory) of the behaviours depend on the phenotypic effects (the  $C_j$ s and the  $B_j$ s, i.e. the parameters of evolutionary graph theory) as well as on the dispersal distribution (the  $m_i$ s), which itself determines the extent to which relatives compete with each other. This illustrates the notion that one cannot simply construct a model and assume that Hamilton's rule applies to parameters one has arbitrarily labelled ' $R$ ', ' $b$ ' and ' $c$ ' (Rousset, 2004; Grafen, 2006). Rather one must carefully consider all life-history parameters to evaluate how behavioural effects translate into fitness costs and benefits.

### Inclusive fitness effect

Using the fitness costs and benefits computed in the previous section, and the stationary probabilities of

identity we evaluate in the Appendix the inclusive fitness effect. We find that the inclusive fitness effect for the Moran process (eqn 33 of Appendix A3) is positive when

$$\sum_j \left[ \frac{B_j}{N} \left( m_j - \frac{2}{n_d} \right) - C_j \left( 1 + \frac{m_0}{N} - \frac{2}{n_d N} \right) \right] > 0, \quad (8)$$

which, when satisfied, means that the probability of fixation of the mutant helping allele is greater than the probability of fixation of a neutral allele (Rousset & Billiard, 2000; Rousset, 2004; Lessard, 2005; Rousset, 2006). Hence, when this inequality holds true, natural selection favours the helping allele.

We now consider eqn 8 under two different types of spatial structures. The first is when the migration rate is the same to all demes and is equal to the probability that an individual remains philopatric ( $m_j = 1/n_d$  for all  $\mathbf{j}$ ) (i.e. the population behaves as a single panmictic unit). In this case all coefficients of relatedness appearing in the inclusive fitness effect are equal to zero ( $R_j = 0$ ) because two individuals sampled at two arbitrary locations in the population have exactly the same probability of bearing the genes (or memes) inherited from the same recent common ancestor. In this case, the condition of invasion of the helping allele is given directly by the effect of the focal individual on its fitness, which, from inequality 8, becomes

$$\sum_j \left[ -C_j - \frac{(B_j - C_j)}{n_d N} \right] > 0. \quad (9)$$

As by definition  $C_j > 0$  and  $B_j > 0$ , the helping mutant allele is counter-selected. The second situation is when dispersal is localized (i.e. there is a spatial structure and isolation by distance,  $m_j \neq 1/n_d$  for some  $\mathbf{j}$  at least). In this case, neighbours on the lattice are more likely to have inherited the same genes from a recent common ancestor than two individuals sampled at random from the population and are positively related. The inclusive fitness effect (inequality 8) informs us that an actor should help the whole set of individuals that are at distance  $\mathbf{j}$  from its deme, whenever the migration rate to that deme exceeds twice the inverse of the number of demes in the population ( $m_j - 2/n_d > 0$ ). Otherwise, the actors should harm ( $B_j < 0$ ) the recipients at a fecundity cost to self.

### Recovering results from evolutionary graph theory

We now turn to the results derived in evolutionary graph theory. From the inclusive fitness effect (inequality 8), we can now obtain the results of Ohtsuki *et al.* (2006) and Ohtsuki & Nowak (2006) by following their assumptions that there is only one individual per node (i.e.  $N = 1$  and  $B_0 = 0$ ) and interactions occur only between the nearest neighbours, of which there are  $k$ . Further, for each nearest neighbour, a focal individual pays a direct

cost  $C$  and provides a benefit  $B$  (i.e.  $B_j = B$  and  $C_j = C$  for the  $k$  nearest neighbour nodes and  $B_j = 0$  and  $C_j = 0$  for all other nodes). Finally, the migration rates are determined by ‘death–birth’ to the nearest neighbours (i.e.  $m_0 = 0$ ,  $m_j = 1/k$  for the  $k$  nearest neighbour nodes and  $m_j = 0$  for all other nodes). Substituting these parameters into eqn 8 shows that the trait spreads when

$$B\left(1 - \frac{2k}{n_d}\right) > kC\left(1 - \frac{2}{n_d}\right). \quad (10)$$

For large numbers of demes or nodes ( $n_d \rightarrow \infty$ ), we recover from this inequality the rule  $B/C > k$  of Ohtsuki *et al.* (2006). For a finite number of nodes and when  $k = 2$ , we recover the rule  $B/C > 2 - 4/(n_d - 4)$  of Ohtsuki & Nowak (2006) for cycles on graphs of size  $n_d$ . Grafen (2007, Table 1) presents the fitness costs and benefits ( $-c$  and the  $b_j$ s) corresponding to this model.

Substituting  $m_0 = 1/(k + 1)$  and  $m_j = 1/(k + 1)$  for the  $k$  nearest neighbour nodes gives ‘imitation’ instead of ‘death–birth’ updating and we recover the appropriate rule  $B/C > 4 - 18/(n_d - 6)$  of Ohtsuki *et al.* (2006) for  $k = 2$  and the rule  $B/C > k + 2$  of Ohtsuki & Nowak (2006) for large  $n_d$ . These derivations of the results of evolutionary graph theory illustrate that the ‘beautiful similarity’ that Ohtsuki *et al.* (2006) note between the rule  $B/C > k$  and Hamilton’s rule arises from a very specific parameterization of eqn 8, which is a specific application of inclusive fitness theory for structured population.

## Discussion

In this paper, we carried out a retrospective analysis of the models for the evolution of helping on graphs of Ohtsuki *et al.* (2006) and Ohtsuki & Nowak (2006). Using inclusive fitness theory (Hamilton, 1964), we recovered exactly the results of Ohtsuki *et al.* (2006) and Ohtsuki & Nowak (2006) for both ‘death–birth’ and ‘imitation’ life cycles as a specific application of kin selection theory for structured populations of finite size (e.g. Rousset & Billiard, 2000; Rousset, 2004, 2006). Grafen’s (2007) analysis demonstrates that inclusive fitness theory allows us to recover the results of Ohtsuki *et al.* (2006) and Ohtsuki & Nowak (2006) for both ‘death–birth’ and ‘birth–death’ life cycles. In other words, both Grafen’s and our analysis show that evolutionary graph theory falls squarely into inclusive fitness theory.

Our inclusive fitness analysis allows us to disentangle the factors necessary to promote the evolution of helping at a fecundity cost to the actor in populations structured according to evolutionary graph theory. The first necessary factor with ‘death–birth’ and ‘imitation’ life cycles is limited migration between actor and recipient, which results in spatial population subdivision. Individuals are then likely to interact with relatives so that kin-selected benefits enter into the equation. However, genetic closeness also results in competition between relatives (Queller, 1992; Taylor, 1992a, b; Gandon & Rousset, 1999).

Indeed, had we assumed for the last stage of the life cycle a Wright–Fisher scheme of reproduction (i.e. all adults die per unit time) instead of the Moran scheme of reproduction (i.e. one individual dies per unit time), kin competition would exactly cancel out the kin-selected fecundity benefits of helping (see Taylor, 1992a, b and eqn 25 in Appendix A3). It is, however, crucial to mention that this does not mean that ‘altruism’ cannot be selected for under the Wright–Fisher scheme of reproduction. Indeed, if altruism is defined from the effects on fitness (as Hamilton did) rather than from the effects on fecundity (as so many subsequent authors assumed), then altruism can be selected for under the Wright–Fisher scheme of reproduction (i.e.  $S > 0$  in eqn 36), while resulting in a fitness cost ( $c > 0$ ) for the focal individual (Rousset, 2004).

The second factor promoting the evolution of helping in populations structured according to evolutionary graph theory is overlapping generations. The assumption of evolutionary graph theory that there is only one adult individual dying and a single juvenile migrating per unit time (Moran process) markedly increases the relatedness between actors and recipients because offspring are then likely to interact directly with their parents (Taylor & Irwin, 2000; Irwin & Taylor, 2001). In Appendix A3 we derive the inclusive fitness effect under the assumption that each adult survives from one generation to the next with probability  $s$  (i.e. ‘Cannings’ scheme of reproduction Ewens, 2004), which represents a linking model between the Moran and the Wright–Fisher process. The analysis of this model reveals that the intensity of selection on helping increases with increased survival probability  $s$  (see eqns 44 and 54). For instance, from eqn 44 and using eqn 50 we find that if interactions occur only with the two nearest neighbours in a one-dimensional habitat of infinite size ( $n_d \rightarrow \infty$ ), the condition of invasion of helping is given by the  $B/C > 1 + 1/\sqrt{s}$  rule. When the life history converges to the Moran process ( $s \rightarrow 1$ ), we recover the  $B/C > 2$  rule of evolutionary graph theory. By contrast, when the life history converges to the Wright–Fisher process ( $s \rightarrow 0$ ), the inequality can never be satisfied and helping at a fecundity cost to the actor cannot evolve (Taylor, 1992a, b). The result that overlapping generations increases the selective pressure on helping has repeatedly been observed in computer simulations or models of the evolution of helping on lattices (e.g. Nowak *et al.*, 1994; Nakamaru *et al.*, 1997; Koella, 2000; Irwin & Taylor, 2001; Hauert, 2002; Hauert & Doebeli, 2004), with helping being more likely to be favoured under ‘asynchronous’ than under ‘synchronous’ updating. Our analysis shows that the results of these simulations also fall under the scope inclusive fitness theory.

The rule presented in this paper (inequality 8) extends previous results in two ways for isothermal regular graphs. The first is to allow more than one individual to live and interact at each node of the graph. By increasing  $N$  we can describe ‘small world’ effects (Watts & Strogatz, 1998), where individuals have strong local

interactions within a node and weaker links to other nodes. Equation 8 shows that increasing within node interactions results in a more stringent condition for the evolution of helping because relatedness is a decreasing function of the number of individuals within demes (nodes). The second extension is to allow interactions to vary with the distance between nodes. This allows us to encompass a much wider range of biological situations conducive to the evolution of helping and harming behaviours. Here, we see that a failure of offspring to disperse from their natal deme (i.e. increasing  $m_0$ ) can lead to a decrease in co-operation because kin competition is increased. This effect is exemplified by a reduced tendency to co-operate under the 'imitation' life cycle when compared with the 'death-birth' life cycle (Ohtsuki *et al.*, 2006; Ohtsuki & Nowak, 2006).

It is useful to recognize that understanding the conditions promoting the emergence and stability of co-operation cannot be achieved without understanding the life cycle of the social system under investigation (Ratnieks, 2006). The results described in this paper are based on the assumptions of a one-locus weak selection model, haploid inheritance, stable population size (i.e. constant demography and environment), overlapping generations but no aging, homogeneous dispersal and homogeneous interactions occurring only between members of the same generation. With these life-cycle assumptions, we were able to use inclusive fitness theory to derive results analytically and to generalize the rule for co-operation proposed by evolutionary graph theorists (Ohtsuki *et al.*, 2006; Ohtsuki & Nowak, 2006). However, despite its generality in the context of the 'death-birth' and 'imitation' life cycles arising in evolutionary graph theory under weak selection, our eqn 8 or even our extension of it (eqn 54) are no more general rules for the evolution of co-operation in structured populations than any of the many rules resulting from considering different life-cycle assumptions (e.g. Eshel, 1972; Aoki, 1982; Taylor, 1992b; van Baalen & Rand, 1998; Frank, 1998; Irwin & Taylor, 2001; Reuter & Keller, 2001; Le Galliard *et al.*, 2003; Roze & Rousset, 2004; Lehmann, 2006; Lehmann *et al.*, 2006). What these rich variety of theoretical studies reveal is that there is no simple mathematical formula for the evolution of co-operative behaviour but that the inclusive fitness framework provides a powerful toolset to disentangle the factors promoting co-operation in spatially structured populations, a problem that will undoubtedly continue to stretch the minds of biologists, sociologists and mathematicians for years to come.

## Acknowledgments

We thank Martine Ehinger, François Rousset, Max Reuter and two anonymous reviewers for useful comments on the manuscript. We have benefited from reading the MS by Grafen (2007) and also thank him for very useful remarks on the manuscript. LK and LL were

both supported by several grants from the Swiss NSF and DJTS is funded by the Royal Society.

## References

- Aoki, K. 1982. A condition for group selection to prevail over counteracting individual selection. *Evolution* **36**: 832–842.
- van Baalen, M. & Rand, A.R. 1998. The unit of selection in viscous populations and the evolution of altruism. *J. Theor. Biol.* **193**: 631–648.
- Epperson, B.K. 1999. Gene genealogies in geographically structured populations. *Genetics* **152**: 797–806.
- Epperson, B.K. 2003. *Geographical Genetics*. Princeton University Press, Princeton, NJ.
- Eshel, I. 1972. On the neighbor effect and the evolution of altruistic traits. *Theor. Popul. Biol.* **11**: 258–277.
- Ewens, W.J. 2004. *Mathematical Population Genetics*. Springer-Verlag, New York.
- Frank, S.A. 1998. *Foundations of Social Evolution*. Princeton University Press, Princeton, NJ.
- Gandon, S. & Rousset, F. 1999. Evolution of stepping-stone dispersal rates. *Proc. R. Soc. Lond. Ser. B. Biol. Sci.* **221**: 2507–2513.
- Grafen, A. 2006. Optimization of inclusive fitness. *J. Theor. Biol.* **238**: 541–563.
- Grafen, A. 2007. An inclusive fitness analysis of altruism on a cyclical network. *J. Evol. Biol.* doi 10.1111/j.1420-9101.2007.01413.x.
- Hamilton, W.D. 1964. The genetical evolution of social behaviour. I. *J. Theor. Biol.* **7**: 1–16.
- Hamilton, W.D. 1970. Selfish and spiteful behavior in an evolutionary model. *Nature* **228**: 1218–1220.
- Hamilton, W.D. 1971. Selection of selfish and altruistic behaviour in some extreme models. In *Man and Beast: Comparative Social Behavior* (J. Eisenberg & W. Dillon, eds), pp. 56–91. Smithsonian Institution Press, Washington, DC.
- Hauert, C. 2002. Effects of space in 22 games. *Int. J. Bifurcat. Chaos* **12**: 1531–1548.
- Hauert, C. & Doebeli, M. 2004. Spatial structure often inhibits the evolution of cooperation in the snowdrift game. *Nature* **428**: 643–646.
- Irwin, A.J. & Taylor, P.D. 2001. Evolution of altruism in stepping-stone populations with overlapping generations. *Theor. Popul. Biol.* **60**: 315–325.
- Koella, J.C. 2000. The spatial spread of altruism versus the evolutionary response of egoists. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **267**: 1979–1985.
- Le Galliard, J., Ferrière, R. & Dieckmann, U. 2003. The adaptive dynamics of altruism in spatially heterogeneous populations. *Evolution* **57**: 1–17.
- Lehmann, L. 2006. The evolution of trans-generational altruism: kin selection meets niche construction. *J. Evol. Biol.* **20**: 181–189.
- Lehmann, L. & Keller, L. 2006. The evolution of cooperation and altruism – a general framework and a classification of models. *J. Evol. Biol.* **19**: 1365–1376.
- Lehmann, L., Perrin, N. & Rousset, F. 2006. Population demography and the evolution of helping behaviors. *Evolution* **60**: 1137–1151.
- Lessard, S. 2005. Long-term stability from fixation probabilities in finite populations: new perspectives for ESS theory. *Theor. Popul. Biol.* **68**: 19–27.
- Lieberman, E., Hauert, C. & Nowak, M.A. 2005. Evolutionary dynamics on graphs. *Nature* **433**: 312–316.

- Malécot, G. 1975. Heterozygosity and relationship in regularly subdivided populations. *Theor. Popul. Biol.* **8**: 212–241.
- Maruyama, T. 1970. Effective number of alleles in a subdivided population. *Theor. Popul. Biol.* **1**: 273–306.
- Maruyama, T. 1974. A simple proof that certain quantities are independent of the geographical structure of population. *Theor. Popul. Biol.* **5**: 148–154.
- Nagylaki, T. 1976. The decay of genetic variability in geographically structured populations. *Theor. Popul. Biol.* **10**: 70–82.
- Nagylaki, T. 1982. Geographical invariance in population genetics. *J. Theor. Biol.* **99**: 159–172.
- Nagylaki, T. 1983. The robustness of neutral models of geographical variation. *Theor. Popul. Biol.* **24**: 268–294.
- Nakamaru, M., Matsuda, H. & Iwasa, Y. 1997. The evolution of cooperation in a lattice-structured population. *J. Theor. Biol.* **184**: 65–81.
- Nowak, M. 2006. Five rules for the evolution of cooperation. *Science* **314**: 1560–1563.
- Nowak, M.A., Bonhoeffer, S. & May, R.M. 1994. Spatial games and the maintenance of cooperation. *Proc. Natl Acad. Sci. U.S.A.* **91**: 4877–4881.
- Ohtsuki, H. & Nowak, M.A. 2006. Evolutionary games on cycles. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **273**: 2249–2256.
- Ohtsuki, H., Hauert, C., Lieberman, E. & Nowak, M.A. 2006. A simple rule for the evolution of cooperation on graphs and social networks. *Nature* **441**: 502–505.
- Queller, D.C. 1992. Does population viscosity promote kin selection. *Trends Ecol. Evol.* **7**: 322–324.
- Ratnieks, F. 2006. The evolution of cooperation and altruism: the basic conditions are simple and well known. *J. Evol. Biol.* **19**: 1413–1414.
- Reuter, M. & Keller, L. 2001. Sex ratio conflict and worker production in eusocial Hymenoptera. *Am. Nat.* **158**: 166–177.
- Rousset, F. 2003. A minimal derivation of convergence stability measures. *J. Theor. Biol.* **221**: 665–668.
- Rousset, F. 2004. *Genetic Structure and Selection in Subdivided Populations*. Princeton University Press, Princeton, NJ.
- Rousset, F. 2006. Separation of time scales fixation probabilities and convergence to evolutionarily stable states under isolation by distance. *Theor. Popul. Biol.* **69**: 165–179.
- Rousset, F. & Billiard, S. 2000. A theoretical basis for measures of kin selection in subdivided populations: finite populations and localized dispersal. *J. Evol. Biol.* **13**: 814–825.
- Roze, D. & Rousset, F. 2004. The robustness of Hamilton's rule with inbreeding and dominance: kin selection and fixation probabilities under partial sib mating. *Am. Nat.* **164**: 214–231.
- Santos, F.C. & Pacheco, J.M. 2005. Scale-free networks provide a unifying framework for the emergence of cooperation. *Phys. Rev. Lett.* **95**: 098104.
- Taylor, P.D. 1992a. Inclusive fitness in a homogeneous environment. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **240**: 299–302.
- Taylor, P.D. 1992b. Altruism in viscous populations – an inclusive fitness model. *Evol. Ecol.* **6**: 352–356.
- Taylor, P.D. & Frank, S.A. 1996. How to make a kin selection model. *J. Theor. Biol.* **180**: 27–37.
- Taylor, P.D. & Irwin, A.J. 2000. Overlapping generations can promote altruistic behavior. *Evolution* **54**: 1135–1141.
- Watts, D.J. & Strogatz, S.H. 1998. Collective dynamics of small-world networks. *Nature* **393**: 440–442.

Received 5 February 2007; revised 4 July 2007; accepted 4 July 2007

## Appendix A1: Inclusive fitness effect

### Direct fitness method

To derive the explicit expression of the inclusive fitness effect for the life cycle described in the main text, we use the direct fitness method (Taylor & Frank, 1996; Frank, 1998) for finite populations (e.g. Rousset & Billiard, 2000; Rousset, 2004, 2006). In this framework, the inclusive fitness effect of a mutant allele with small phenotypic deviation relative to a resident allele (weak selection) can be evaluated as

$$S = \frac{\partial w}{\partial z_{\bullet}} + \sum_{\mathbf{j}} \frac{\partial w}{\partial z_{\mathbf{j}}} Q_{\mathbf{j}}, \quad (11)$$

where  $Q_{\mathbf{j}}$  denotes the stationary probability that a recipient sampled in a deme at distance  $\mathbf{j}$  from the focal deme bears a copy of a homologous gene (or meme for cultural evolution) drawn in the focal individual. Since  $S$  represents the first-order phenotypic effect of a mutant gene lineage on its fitness, it is sufficient to evaluate the stationary probabilities of genetic identity under a neutral model only (Rousset, 2003, 2004).

The functional form of eqn 11 is convenient for mathematical simplification (see below) but  $S$  can equivalently be expressed in terms of 'relatedness coefficients'  $R_{\mathbf{j}}$  measuring the extent to which two interacting individuals sampled at distance  $\mathbf{j}$  are more likely to share genes (or memes) identical by descent than two individuals sampled at random in the population. Noting that the effects of actors on fitness sum up to zero:  $\partial w / \partial z_{\bullet} + \sum_{\mathbf{j}} \partial w / \partial z_{\mathbf{j}} = 0$  (Rousset & Billiard, 2000; Rousset, 2004), we subtract  $(\partial w / \partial z_{\bullet} + \sum_{\mathbf{j}} \partial w / \partial z_{\mathbf{j}}) \bar{Q}$  from the right-hand side of eqn 11 and divide the whole expression by  $1 - \bar{Q}$ , where  $\bar{Q} = \sum_{\mathbf{j}} Q_{\mathbf{j}} / n_{\mathbf{d}}$  is the average probability of identity between pairs of homologous genes sampled in different individuals.

After simplification, we obtain

$$S \propto \frac{\partial w}{\partial z_{\bullet}} + \sum_{\mathbf{j}} \frac{\partial w}{\partial z_{\mathbf{j}}} R_{\mathbf{j}}, \quad (12)$$

where  $R_{\mathbf{j}} = (Q_{\mathbf{j}} - \bar{Q}) / (1 - \bar{Q})$  measures the extent to which the gene lineages of two individuals sampled at distance  $\mathbf{j}$  coalesce in a more recent past than the gene lineages of two individuals sampled at random from the population. Note that the definition of  $\bar{Q}$  implies that the average relatedness is equal to zero

$$\frac{1}{n_{\mathbf{d}}} \sum_{\mathbf{j}} R_{\mathbf{j}} = 0. \quad (13)$$

### Fitness function

To evaluate explicitly the measure of selection  $S$ , it remains to establish the fitness function  $w$  and the

stationary probabilities of identity (the  $Q_{\mathbf{k}}$ s). In so doing and to relate our formalization directly to previous kin selection models for spatially structured populations, we make one additional assumption for stage (3) of our life cycle described in the main text. Instead of exactly one adult individual dying per unit time, we assume that each adult individual survives with probability  $s$  to the next generation. This allows us to compare the classical kin selection models for the evolution of helping in spatially subdivided populations (e.g. Taylor, 1992a, b; Taylor & Irwin, 2000; Irwin & Taylor, 2001; Rousset, 2004; Lehmann & Keller, 2006) with those of evolutionary graph theory for the ‘death–birth’ and ‘imitation’ processes (Ohtsuki *et al.*, 2006; Ohtsuki & Nowak, 2006). With this assumption, the direct fitness function  $w$  of a focal individual is written as

$$w = s + (1 - s) \sum_{\mathbf{i}} m_{\mathbf{i}} \frac{1 - \sum_{\mathbf{k}} (C_{\mathbf{k}} z_{\bullet} - B_{\mathbf{k}} z_{\mathbf{k}})}{\sum_{\mathbf{j}} m_{\mathbf{i}-\mathbf{j}} \left( 1 + \sum_{\mathbf{k}} B'_{\mathbf{j}-\mathbf{k}} z_{\mathbf{k}}^{\mathbf{R}} \right)} \quad (14)$$

and the subscripts are read on a periodic lattice, e.g. for a circle we have

$$m_{-i} = m_{n_d-i} \text{ and } m_{n_d+i} = m_i.$$

When  $s = 1 - 1/N_T$ , this equation is equivalent to the fitness function presented in the main text and when  $s = 0$  it is equivalent to the fitness function of Rousset (2004, eqn 7.18).

Substituting eqn 14 into eqn 11, evaluating the partial derivatives at the phenotypic value of the defector allele, which does not express any helping behaviour ( $z_{\bullet} = z_0 = \dots = z_{\mathbf{j}} = \dots = 0$ ) and rearranging, we finally obtain

$$S = (1 - s) \left[ - \sum_{\mathbf{k}} C_{\mathbf{k}} + \sum_{\mathbf{k}} B_{\mathbf{k}} Q_{\mathbf{k}} - \sum_{\mathbf{k}} \sum_{\mathbf{i}} \sum_{\mathbf{j}} B'_{\mathbf{j}-\mathbf{k}} m_{\mathbf{i}} m_{\mathbf{i}-\mathbf{j}} Q_{\mathbf{k}}^{\mathbf{R}} \right], \quad (15)$$

where

$$Q_{\mathbf{k}}^{\mathbf{R}} = Q_{\mathbf{k}} \text{ except that } Q_0^{\mathbf{R}} = \frac{1}{N} + \left( \frac{N-1}{N} \right) Q_0. \quad (16)$$

The inclusive fitness effect  $S$  is equal, up to the constant of proportionality  $1 - s$ , to the inclusive fitness effect for spatially subdivided populations established by Taylor (1992b, eqn 3) and Rousset (2004, eqn 7.19), and it consists of three terms. The first term is the fecundity cost for a focal individual expressing the helping allele. The second term is the fecundity benefit received by the focal individual from all other individuals in the population bearing the helping allele. Finally, the third term is the increase in competition faced by the focal individual’s offspring and resulting from him and all other actors in the population expressing the mutant behaviour.

By calculating the corresponding probabilities of identity, eqn 15 allows us to evaluate the selective pressure on the

mutant allele under three different demographic regimes: (1) exactly only individual in the population dies per generation, i.e.  $s = 1 - 1/N_T$  (Moran process, e.g. Ewens, 2004) as assumed by Ohtsuki *et al.* (2006) and Ohtsuki & Nowak (2006); (2) all individuals of the population die in each generation, i.e.  $s = 0$  (Wright–Fisher process, e.g. Ewens, 2004) as assumed by Taylor (1992a, b) and Rousset (2004); and (3) the intermediate regime where each individual has a probability  $s$  of surviving from one generation to the next (Cannings process, e.g. Ewens, 2004) as assumed in Taylor & Irwin (2000), Irwin & Taylor (2001) and Lehmann & Keller (2006). In the latter case, the inclusive fitness effect  $S$  is the expectation over the realizations of the demographic states of the population resulting from the number of adults surviving in each deme being a random variable. In Appendix A2, we present the recurrence equations for the probabilities of identity for these three demographic regimes.

## Appendix A2: Genetic structure of the population

### Overlapping generations: Moran process

From the assumptions that only one individual dies per unit time, we find by applying standard methods (e.g. Maruyama, 1970; Malécot, 1975; Nagylaki, 1976, 1983; Epperson, 1999; Gandon & Rousset, 1999; Epperson, 2003; Rousset, 2004) that the probabilities  $Q_{\mathbf{k}}$  that two individuals sampled at distance  $\mathbf{k}$  bear the same genes satisfy the recursions:

$$Q_{\mathbf{k}}(t + 1) = s_A Q_{\mathbf{k}}(t) + (1 - s_A)(1 - \mu) \sum_{\mathbf{i}} m_{\mathbf{i}} Q_{\mathbf{k}-\mathbf{i}}^{\mathbf{R}}(t) \quad (17)$$

and

$$Q_{\mathbf{k}-\mathbf{i}}^{\mathbf{R}}(t) = Q_{\mathbf{k}-\mathbf{i}}(t) \text{ except that } Q_0^{\mathbf{R}}(t) = \frac{1}{N} + \left( \frac{N-1}{N} \right) Q_0(t), \quad (18)$$

where  $\mu$  is the mutation rate and the subscripts are read on a periodic lattice, e.g. for a circle we have

$$m_{-i} = m_{n_d-i} \text{ and } m_{n_d+i} = m_i.$$

The recursions for the probabilities of identity depend on the probability  $s_A = 1 - 2/N_T$  that two adult individuals sampled at random from the same or from two different demes have survived from one generation to the next ( $N_T \equiv n_d N$  is the total population size). With complementary probability  $1 - s_A$ , one samples an adult individual that has survived and the only new individual of the population. The latter descends from the former with probability  $1/N$ , whenever the new individual originates from the same deme as the parent. In terms of and evolutionary graph theory (Lieberman *et al.*, 2005; Ohtsuki *et al.*, 2006; Ohtsuki & Nowak, 2006),  $m_{\mathbf{l}}$  represents the probability that an individual on a given vertex places its offspring in a vertex at distance  $\mathbf{l}$  (i.e. the

$m_{ij}$ s are the elements of the adjacency matrix of the graph, which is circulant here). At steady state, we have from eqs 17 and 18:

$$Q_k = (1 - \mu) \sum_i m_i Q_{k-i}^R \tag{19}$$

and  $Q_{k-i}^R$  is given by eqn 16. The recursions for the probabilities of identity allow us also to establish a useful formula, which will be helpful in forthcoming calculations. Writing the recursions in matrix form and applying population genetic methods (Rousset, 2004, chapter 3), yields an expression that relates the diversity in a deme ( $1 - Q_0$ ) to total population size  $N_T$  (e.g. Maruyama, 1974; Nagylaki, 1982) by the equality

$$\lim_{\mu \rightarrow 0} \frac{\mu}{1 - Q_0} = \frac{1}{N_T}. \tag{20}$$

This expression is obtained as follows. Write the recursions for the probabilities of identity (eqn 19) in matrix form as

$$\mathbf{Q} = (1 - \mu)\mathbf{F}(\mathbf{Q} + \mathbf{c}), \tag{21}$$

where  $\mathbf{Q}$  is a vector collecting the  $Q_k$ s,  $\mathbf{F}$  is a circulant matrix collecting the  $m_{ij}$ s and  $\mathbf{c} \equiv ((1 - Q_0)/N, 0, 0, \dots)$ . Let  $\mathbf{0}$  and  $\mathbf{1}$  be the column zero and unit vectors, differentiate eqn 21 with respect to  $\mu$  at  $\mu = 0$ , use  $\mathbf{Q}|_{\mu=0} = \mathbf{1}$ ,  $\mathbf{c}|_{\mu=0} = \mathbf{0}$  and  $\mathbf{F}\mathbf{1} = \mathbf{1}$  to obtain

$$\frac{d\mathbf{Q}}{d\mu} = -\mathbf{1} + \mathbf{F} \frac{d\mathbf{Q}}{d\mu} + \mathbf{F} \frac{d\mathbf{c}}{d\mu}, \tag{22}$$

where  $d\mathbf{c}/d\mu = (-dQ_0/d\mu)/N, 0, 0, \dots$ . Premultiplying eqn 22 by the row unit vector  $\mathbf{1}^T$ , using  $\mathbf{1}^T \mathbf{1} = n_d$  and  $\mathbf{1}^T(\mathbf{1} - \mathbf{F}) = \mathbf{0}$  because  $\mathbf{1}^T \mathbf{F} = \mathbf{1}^T$ , we have

$$\frac{dQ_0}{d\mu} = -n_d N. \tag{23}$$

We can now recover eqn 20 by using the definition of the derivative at  $\mu = 0$

$$\frac{dQ_0}{d\mu} = \lim_{\mu \rightarrow 0} \frac{Q_0 - 1}{\mu}. \tag{24}$$

### Nonoverlapping generations: Wright–Fisher process

The recursions for the probabilities of identity for the Wright–Fisher scheme of reproduction are very well known (e.g. Malécot, 1975; Nagylaki, 1976, 1982; Epperson, 1999; Gandon & Rousset, 1999; Rousset & Billiard, 2000; Rousset, 2004), and they satisfy at steady state the equations

$$Q_k = (1 - \mu)^2 \sum_i \sum_j m_i m_{i-j} Q_{k-j}^R, \tag{25}$$

where as before  $Q_{k-j}^R$  is given by eqn 16. The equilibrium diversity in a deme for this model is then given by Rousset (2004, eqn 3.68) and can be expressed as

$$\lim_{\mu \rightarrow 0} \frac{\mu(2 - \mu)}{1 - Q_0} = \frac{1}{N_T}. \tag{26}$$

### Overlapping generations: Cannings process

When each individual has a per generation probability  $s$  of survival, we find that the probabilities of identity satisfy at equilibrium the equations

$$Q_k = s^2 Q_k + 2s(1 - s)(1 - \mu) \sum_i m_i Q_{k-i}^R + (1 - s)^2 (1 - \mu)^2 \sum_i \sum_j m_i m_{i-j} Q_{k-j}^R, \tag{27}$$

where as before  $Q_i^R$  is given by eqn 16. Using the same methods as for the Moran and Wright–Fisher processes, we obtain for the diversity in a deme for this life-history setting

$$\lim_{\mu \rightarrow 0} \frac{\mu(2 - \mu)}{(1 - Q_0)} = \frac{(1 + s)}{N_T}. \tag{28}$$

### Comparing life-history regimes

Comparing eqns 19, 25 and 27 and assuming that  $s \geq 1 - 1/N_T$  inform us that the demographic regime eroding less the probabilities of identity through migration is given by the Moran process (eqn 19). Hence, for a given dispersal distribution resulting in isolation by distance, this is the demographic regime causing the highest relatedness  $R_0$  between locally interacting individuals. This result can also be noted by comparing the steady-state diversity within a deme (compare eqns 20, 26 and 28), which is the lowest under the Moran process (eqn 20).

### Appendix A3: Explicit inclusive fitness effect

We now have the two ingredients, fitness effects and probabilities of identity, that allow us to evaluate the inclusive fitness effect explicitly.

### Overlapping generations: Moran process

Because both dispersal and interaction between individuals are spatially homogeneous and isotropic, the third term of the inclusive fitness effect (eqn 15) is given by

$$\sum_k \sum_i \sum_j B'_{j-k} m_i m_{i-j} Q_k^R = \sum_k \sum_i \sum_j B'_k m_i m_{i-j} Q_{j-k}^R, \tag{29}$$

where the right-hand side of this equality will now be simplified by using eqn 19 twice. This isothermal property of the graphs greatly reduces the number of terms

we need to consider in the derivation of  $S$ . Specifically, we use eqn 19 once under the form  $Q_{i-k} = (1 - \mu) \sum_j m_{i-j} Q_{j-k}^R$  and once under the form  $Q_k = (1 - \mu) \sum_i m_i Q_{i-k} + (1 - \mu) m_k (1 - Q_0)/N$ , which allows us to obtain

$$\begin{aligned} \sum_k \sum_i \sum_j B'_k m_i m_{i-j} Q_{j-k}^R &= \frac{1}{1 - \mu} \sum_k \sum_i B'_k m_i Q_{i-k} \\ &= \frac{1}{1 - \mu} \sum_k B'_k \left( \frac{Q_k}{1 - \mu} - \frac{m_k (1 - Q_0)}{N} \right) \\ &= - \sum_k C_k \left( \frac{Q_0}{(1 - \mu)^2} - \frac{m_0 (1 - Q_0)}{(1 - \mu)N} \right) \\ &\quad + \sum_k B_k \left( \frac{Q_k}{(1 - \mu)^2} - \frac{m_k (1 - Q_0)}{(1 - \mu)N} \right), \end{aligned} \tag{30}$$

where the last line is obtained by using  $B'_k \equiv B_k$  except that  $B'_0 \equiv B_0 - \sum_i C_i$ . Inserting the last equality into the inclusive fitness effect and setting  $s = 1 - 1/N_T$ , we find that it can be written as

$$\begin{aligned} S &= \frac{1}{N_T} \sum_k \left[ -C_k \left( \frac{1 - Q_0}{(1 - \mu)^2} - \frac{\mu + (1 - \mu)\mu}{(1 - \mu)^2} + \frac{m_0 (1 - Q_0)}{(1 - \mu)N} \right) \right. \\ &\quad \left. + B_k \left( \frac{m_k (1 - Q_0)}{(1 - \mu)N} - \frac{Q_k (\mu + (1 - \mu)\mu)}{(1 - \mu)^2} \right) \right]. \end{aligned} \tag{31}$$

Dividing the inclusive fitness effect by  $1 - Q_0$ , using the formula for the stationary diversity within a deme for the Moran process (eqn 20) and taking the low mutation limit ( $\lim_{\mu \rightarrow 0} S/(1 - Q_0)$ ), which results in  $Q_k \rightarrow 1$ , shows that the inclusive fitness effect is proportional to

$$S \propto \sum_k \left[ -C_k \left( 1 + \frac{m_0}{N} - \frac{2}{N_T} \right) + B_k \left( \frac{m_k}{N} - \frac{2}{N_T} \right) \right], \tag{32}$$

which, when  $m_k = 1/n_d$  for all  $k$  (i.e. unstructured population), reduces to eqn 9. Hence, the mutant is selected for when

$$\sum_k \left[ -C_k \left( 1 + \frac{m_0}{N} - \frac{2}{N_T} \right) + B_k \left( \frac{m_k}{N} - \frac{2}{N_T} \right) \right] > 0. \tag{33}$$

**Nonoverlapping generations: Wright–Fisher process**

The inclusive fitness effect when each individual of the population dies in each generation has been analysed by Taylor (1992b) and Rousset (2004, p. 124). Inserting eqn 25 into the right-hand side of eqn 29, we have

$$\sum_k \sum_i \sum_j B'_k m_i m_{i-j} Q_{j-k}^R = \frac{1}{(1 - \mu)^2} \sum_k B'_k Q_k. \tag{34}$$

Inserting this equation into eqn 15 and setting  $s = 0$ , we then have

$$\begin{aligned} S &= \sum_k \left[ -C_k \left( 1 - \frac{Q_0}{(1 - \mu)^2} \right) + B_k Q_k - B_k \frac{Q_k}{(1 - \mu)^2} \right] \\ &= \frac{1}{(1 - \mu)^2} \sum_k \left[ -C_k (1 - Q_0 - \mu(2 - \mu)) - B_k Q_k \mu(2 - \mu) \right]. \end{aligned} \tag{35}$$

Dividing the inclusive fitness effect by  $1 - Q_0$ , using the formula for the stationary diversity within a deme (eqn 26) and taking the low mutation limit ( $\lim_{\mu \rightarrow 0} S/(1 - Q_0)$ ), reveals that the inclusive fitness effect is proportional to

$$S \propto \sum_k \left[ -C_k \left( 1 - \frac{1}{N_T} \right) - \frac{1}{N_T} B_k \right], \tag{36}$$

which is always a net cost (e.g. Rousset, 2004, eqn 7.21) independent of the structure of the population! When migration is random ( $m_k = 1/n_d$  for all  $k$ ), eqn 32 becomes equivalent to eqn 36. Hence, the direction of selection on the mutant for weak selection is the same in panmictic populations, whether there is overlapping generations or not.

**Overlapping generations: Cannings process**

The inclusive fitness effect when each adult individual has a probability  $s$  of surviving per generation has been studied previously under more stringent life-cycle assumptions, that is, for the infinite island model of dispersal by Taylor & Irwin (2000), for the stepping-stone model of dispersal and interactions by Irwin & Taylor (2001) and for an arbitrarily dispersal distribution but with only local interactions by Lehmann & Keller (2006). Inserting eqn 27 into the right-hand side of eqn 29, we have

$$\begin{aligned} \sum_k \sum_i \sum_j B'_k m_i m_{i-j} Q_{j-k}^R &= \sum_k B'_k \left[ \frac{(1 - s^2) Q_k}{(1 - \mu)^2 (1 - s)^2} \right. \\ &\quad \left. - \frac{2s(1 - s)}{(1 - \mu)(1 - s)^2} \sum_i m_i Q_{k-i}^R \right]. \end{aligned} \tag{37}$$

We now expand the double sum appearing in the second term of the right-hand side as

$$\begin{aligned} \sum_k B'_k \sum_i m_i Q_{k-i}^R &= \sum_k B'_k \left[ \sum_i m_i Q_{k-i} + \frac{m_{-k} (1 - Q_0)}{N} \right] \\ &= \sum_k B'_k \left[ \sum_i m_i (Q_{k-i} - Q_0) + Q_0 \right. \\ &\quad \left. + \frac{m_{-k} (1 - Q_0)}{N} \right] \\ &= \sum_k B'_k \left[ Q_0 + \frac{m_k (1 - Q_0)}{N} - Y_k (1 - Q_0) \right], \end{aligned} \tag{38}$$

because  $m_{-k} = m_k$  and

$$Y_{\mathbf{k}} \equiv \sum_{\mathbf{i}} m_{\mathbf{i}} \frac{(Q_0 - Q_{\mathbf{k}-\mathbf{i}})}{(1 - Q_0)}, \quad (39)$$

which is a measure of average diversity. Substituting the last three equations into the inclusive fitness effect (eqn 15), dividing it by  $(1 - Q_0)$  and rearranging we obtain

$$\begin{aligned} \frac{S}{1 - Q_0} = (1 - s) \sum_{\mathbf{k}} \left[ -C_{\mathbf{k}} \left( \frac{1}{\gamma} + \frac{2s(m_0/N - Y_0)}{(1 - s)\sqrt{\gamma}} \right) \right. \\ \left. - \frac{1 - \gamma}{\gamma(1 - Q_0)} - \frac{2s(1 - \sqrt{\gamma})Q_0}{(1 - s)\gamma(1 - Q_0)} \right) \\ + B_{\mathbf{k}} \left( \frac{2s}{(1 - s)} \frac{(Q_0 - Q_{\mathbf{k}})}{(1 - Q_0)} + \frac{2s(m_{\mathbf{k}}/N - Y_{\mathbf{k}})}{(1 - s)\sqrt{\gamma}} \right) \\ \left. - \frac{(1 + s)(1 - \gamma)Q_{\mathbf{k}}}{(1 - s)\gamma(1 - Q_0)} - \frac{2s[1 - (1/\sqrt{\gamma})]Q_0}{(1 - s)(1 - Q_0)} \right], \quad (40) \end{aligned}$$

where  $\gamma \equiv (1 - \mu)^2$ . Taking the low mutation limit ( $\lim_{\mu \rightarrow 0} S/(1 - Q_0)$ ) and using the stationary diversity within a deme for the Cannings process (eqn 28), we have

$$\lim_{\mu \rightarrow 0} \left( \frac{1 - \gamma}{\gamma(1 - Q_0)} + \frac{2s(1 - \sqrt{\gamma})Q_0}{(1 - s)\gamma(1 - Q_0)} \right) = \frac{(1 + s)}{(1 - s)N_T} \quad (41)$$

and

$$\lim_{\mu \rightarrow 0} \left( \frac{(1 + s)(1 - \gamma)Q_{\mathbf{k}}}{(1 - s)\gamma(1 - Q_0)} + \frac{2s[1 - (1/\sqrt{\gamma})]Q_0}{(1 - s)(1 - Q_0)} \right) = \frac{(1 + s)}{(1 - s)N_T}. \quad (42)$$

With these formulae in hand, the inclusive fitness effect finally becomes

$$\begin{aligned} S \propto \sum_{\mathbf{k}} \left[ -C_{\mathbf{k}} \left( 1 + \frac{2s}{(1 - s)} X_0 - \frac{(1 + s)}{(1 - s)N_T} \right) \right. \\ \left. + B_{\mathbf{k}} \left( \frac{2s}{(1 - s)} X_{\mathbf{k}} - \frac{(1 + s)}{(1 - s)N_T} \right) \right], \quad (43) \end{aligned}$$

where

$$X_{\mathbf{k}} \equiv \frac{m_{\mathbf{k}}}{N} + \sum_{\mathbf{i}} m_{\mathbf{i}} \frac{(Q_{\mathbf{k}-\mathbf{i}} - Q_{\mathbf{k}})}{(1 - Q_0)}, \quad (44)$$

which will depend on the shape of the dispersal distribution and must be evaluated in the low mutation limit (when  $\mu \rightarrow 0$ ). Hence, the inclusive fitness effect does not reduce to a simple form without further assumption on the life cycle.

We will now express the  $X_{\mathbf{k}}$ s in terms of the dispersal distribution to subsequently obtain a low migration approximation of the inclusive fitness effect (Nagylaki, 1982; Rousset, 2004). To that aim we use classical results on Fourier analysis and follow similar developments as presented in Rousset (2004, chapter 3). Call  $\psi(\mathbf{z}) \equiv \sum_{\mathbf{i}} m_{\mathbf{i}} e^{i\mathbf{z}}$  the characteristic function (the Fourier transform) of the dispersal distribution, where  $\iota \equiv \sqrt{-1}$ , and let  $\mathcal{Q}(\mathbf{z}) \equiv \sum_{\mathbf{i}} Q_{\mathbf{i}} e^{i\mathbf{z}}$  be the Fourier transform of the  $Q_{\mathbf{i}}$ s. Fourier transforming eqn 27 and rearranging, we find that

$$\begin{aligned} \mathcal{Q}(\mathbf{z}) = s^2 \mathcal{Q}(\mathbf{z}) + 2s(1 - s)(1 - \mu)\psi(\mathbf{z}) \left( \mathcal{Q}(\mathbf{z}) + \frac{(1 - Q_0)}{N} \right) \\ + (1 - s)^2(1 - \mu)^2\psi(\mathbf{z})^2 \left( \mathcal{Q}(\mathbf{z}) + \frac{(1 - Q_0)}{N} \right), \quad (45) \end{aligned}$$

which, once solved for the characteristic function  $\mathcal{Q}(\mathbf{z})$ , yields

$$\mathcal{Q}(\mathbf{z}) = \frac{(1 - Q_0)}{N} \mathcal{F}(\mathbf{z}), \quad (46)$$

where

$$\mathcal{F}(\mathbf{z}) = \frac{(1 - \mu)\psi(\mathbf{z})(2s + (1 - s)(1 - \mu)\psi(\mathbf{z}))}{1 + s(1 - 2(1 - \mu)\psi(\mathbf{z})) - (1 - s)(1 - \mu)^2\psi(\mathbf{z})^2}, \quad (47)$$

which, when  $s \rightarrow 0$ , makes direct contact with the standard formulae (e.g. Malécot, 1975; Nagylaki, 1976; Epperson, 1999; Rousset, 2004). From these equations we can unleash the stationary  $Q_{\mathbf{k}}$  as

$$Q_{\mathbf{k}} = \frac{(1 - Q_0)}{N} \mathcal{L}_{\mathbf{k}}(\mathcal{F}), \quad (48)$$

where  $\mathcal{L}_{\mathbf{k}}(\mathcal{F}) \equiv (1/n_d) \sum_{\mathbf{j}} \mathcal{F}(\mathbf{z}(\mathbf{j})) e^{-i\mathbf{k} \cdot \mathbf{z}(\mathbf{j})}$  is the inverse Fourier transform of  $\mathcal{F}$  at distance  $\mathbf{k}$ .

Noting that  $m_{\mathbf{k}} = \mathcal{L}_{\mathbf{k}}(\psi)$  and using the stationary  $Q_{\mathbf{k}}$ , we can write

$$\begin{aligned} X_{\mathbf{k}} = \frac{1}{N} \left( \mathcal{L}_{\mathbf{k}}(\psi) + \sum_{\mathbf{i}} m_{\mathbf{i}} [\mathcal{L}_{\mathbf{k}-\mathbf{i}}(\mathcal{F}) - \mathcal{L}_{\mathbf{k}}(\mathcal{F})] \right) \\ = \frac{1}{N} \left( \mathcal{L}_{\mathbf{k}}(\psi) + \frac{1}{n_d} \sum_{\mathbf{i}} \sum_{\mathbf{j}} m_{\mathbf{i}} \mathcal{F}(\mathbf{z}(\mathbf{j})) \right. \\ \left. \times [e^{-i(\mathbf{k}-\mathbf{i}) \cdot \mathbf{z}(\mathbf{j})} - e^{-i\mathbf{k} \cdot \mathbf{z}(\mathbf{j})}] \right) \\ = \frac{1}{N} \left( \mathcal{L}_{\mathbf{k}}(\psi) + \frac{1}{n_d} \sum_{\mathbf{i}} \sum_{\mathbf{j}} m_{\mathbf{i}} \mathcal{F}(\mathbf{z}(\mathbf{j})) e^{-i\mathbf{k} \cdot \mathbf{z}(\mathbf{j})} [e^{i\mathbf{i} \cdot \mathbf{z}(\mathbf{j})} - 1] \right) \\ = \frac{1}{N} \left( \mathcal{L}_{\mathbf{k}}(\psi) + \frac{1}{n_d} \sum_{\mathbf{j}} \mathcal{F}(\mathbf{z}(\mathbf{j})) [\psi(\mathbf{z}(\mathbf{j})) - 1] e^{-i\mathbf{k} \cdot \mathbf{z}(\mathbf{j})} \right) \\ = \frac{1}{N} \mathcal{L}_{\mathbf{k}}(\psi + \mathcal{F}\psi - \mathcal{F}). \quad (49) \end{aligned}$$

Substituting eqn 47 into the last equation and taking the low mutation limit ( $\lim_{\mu \rightarrow 0} X_{\mathbf{k}}$ ), we obtain after simplification

$$\lim_{\mu \rightarrow 0} X_{\mathbf{k}} = \frac{1}{N} \mathcal{L}_{\mathbf{k}} \left( \frac{(1 - s)\psi}{1 + s + (1 - s)\psi} \right). \quad (50)$$

Now that the  $X_{\mathbf{k}}$ s are expressed in terms of the characteristic function of the dispersal distribution, we can establish a low migration approximation for these functions. To that aim, we write  $m_0 = (1 - m)$  and  $m_{\mathbf{i}} = mg_{\mathbf{i}}$ , and approximate the  $X_{\mathbf{k}}$ s by letting the migration rate  $m$  go to zero (Rousset, 2004, chapter 3). From these definitions, the characteristic function of dispersal can be

expressed as  $\psi = 1 - m(1 - \sum_{i \neq 0} g_i e^{i \cdot z})$ . Inserting this expression into  $[(1-s)\psi/(1+s+(1-s)\psi)]$  and by Taylor expansion at  $m = 0$ , we get

$$\frac{(1-s)\psi}{1+s+(1-s)\psi} = \frac{(1-s)^2}{4} + \frac{(1-s^2)\psi}{4} + O(m^2). \quad (51)$$

By inverse Fourier transforming this expression and noting that the inverse transform of a constant  $a$  is  $\mathcal{L}_0(a) = 0$  except that  $\mathcal{L}_0(a) = a$ , we finally obtain

$$X_{\mathbf{k}} \approx \frac{(1-s^2)m_{\mathbf{k}}}{4N} \quad (52)$$

except that

$$X_0 \approx \frac{(1-s)^2}{4N} + \frac{(1-s^2)(1-m)}{4N}. \quad (53)$$

These approximations are valid irrespective of the shape of the dispersal distribution, whenever the dispersal rate  $m$  is small.

With these approximations, the inclusive fitness finally becomes

$$\begin{aligned} S \propto & -C_0 + (B_0 - C_0) \left( \frac{s(2-m(1+s))}{2N} - \frac{(1+s)}{(1-s)N_T} \right) \\ & + \sum_{\mathbf{k} \neq 0} \left[ -C_{\mathbf{k}} \left( 1 + \frac{s(2-m(1+s))}{2N} - \frac{(1+s)}{(1-s)N_T} \right) \right. \\ & \left. + B_{\mathbf{k}} \left( \frac{s(1+s)m_{\mathbf{k}}}{2N} - \frac{(1+s)}{(1-s)N_T} \right) \right]. \quad (54) \end{aligned}$$

Letting  $n_d \rightarrow \infty$  in this equation and then  $s \rightarrow 1$  allow us to recover eqn 32 for  $n_d \rightarrow \infty$ .