

# The evolution of social discounting in hierarchically clustered populations

LAURENT LEHMANN\* and FRANÇOIS ROUSSET<sup>†</sup>

\*Department of Ecology and Evolution, University of Lausanne, UNIL-Sorge, 1015 Lausanne, Switzerland, <sup>†</sup>Université Montpellier II, CNRS, UMR 5554, Institut des Sciences de l'Évolution Montpellier, France

## Abstract

The expression of a social behaviour may affect the fitness of actors and recipients living in the present and in the future of the population. When there is a risk that a future reward will not be experienced in such a context, the value of that reward should be discounted; but by how much? Here, we evaluate social discount rates for delayed fitness rewards to group of recipients living at different positions in both space and time than the actor in a hierarchically clustered population. This is a population where individuals are grouped into families, families into villages, villages into clans, and so on, possibly *ad infinitum*. The group-wide fitness effects are assumed to either increase or decrease the fecundity or the survival of recipients and can be arbitrarily extended in space and time. We find that actions changing the survival of individuals living in the future are generally more strongly discounted than fecundity-changing actions for all future times and that the value of future rewards increases as individuals live longer. We also find that delayed fitness effects may not only be discounted by a constant factor per unit delay (exponential discounting), but that, as soon as there is localized dispersal in a population, discounting per unit delay is likely to fall rapidly for small delays and then slowly for longer delays (hyperbolic discounting). As dispersal tends to be localized in natural populations, our results suggest that evolution is likely to favour individuals that express present-biased behaviours and that may be time-inconsistent with respect to their group-wide effects.

**Keywords:** hyperbolic discounting, relatedness, social discounting, spatially structured population

Received 2 March 2011; revision received 28 May 2011; accepted 7 June 2011

Why should I care about posterity? What's posterity ever done for me?

Groucho Marx

## Introduction

Understanding how individuals trade-off between immediate and future rewards to self and to other individuals in a population is one of the classical problem in evolutionary biology and economics (Arrow & Levin 2009). Plants and animals trade-off between present

investment into reproduction and delayed investment into reproduction by temporarily allocating resources to growth (León 1976; Stearns 1992); parents trade-off between present investment into their own survival and investment into the growth of their offspring, which will reproduce only after a delay (Trivers 1974; Becker & Barro 1988); and more generally individuals across all levels of biological organization trade-off between their own short-term interests and the long-term interests of the common good (Hardin 1968; Maynard Smith & Szathmary 1995).

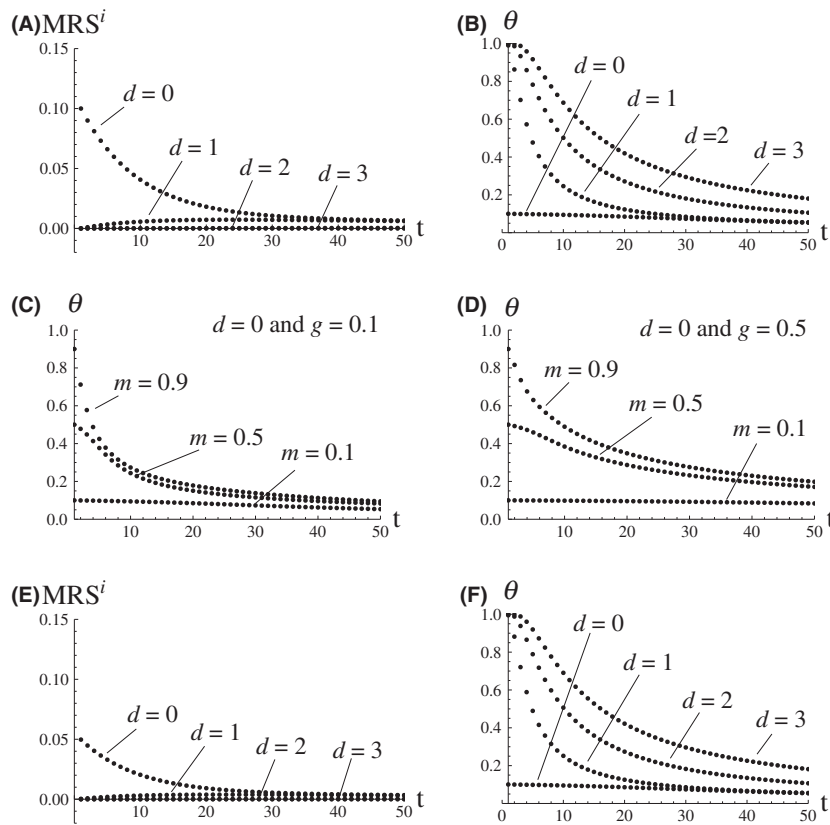
When faced with the choice between an immediate and a future reward to self of equal magnitude, organisms typically prefer the immediate reward and

Correspondence: Laurent Lehmann, Fax: +41 21 692 4165; E-mail: laurent.lehmann@unil.ch

therefore discount the future (Tobin & Logue 1994). Humans and other animals not only discount the future, but they do so in a way where discounting generally falls rapidly per unit of delay for small delays, but then falls more slowly for longer delays (e.g. Ainslie 1975; Laibson 1995; Green & Myerson 1996). Such a decreasing rate of discounting per unit of delay is called hyperbolic discounting (e.g. Rogers 1994; Laibson 1995; Green & Myerson 1996; Sozou 1998), and it causes individuals to have a discount rate that increases as the time preceding the reward becomes shorter. That is, individuals become more impatient when the reward becomes more imminent, in which case they are said to be present-biased. Hyperbolic discounting may even result in time-preference reversal effects (e.g. Sozou 1998, Fig. 1), which occurs, for instance, if early in the year one tries to put money aside for Christmas, but as time goes by, the money is spent on summer vacations (Dasgupta & Maskin 2005).

With hyperbolic discounting, individuals can be described not only as being present-biased but also as being time-inconsistent, a characteristic that has even been interpreted as nonrational (Ainslie 1975; but see the discussion of Sozou 1998).

What kind of discount rates are favoured by evolution? Life history theory predicts that the discounting of a future reward to self should increase with a decrease in the survival probability to (or fertility at) the time when the reward is obtained, as there is an increased risk that the reward does not translate into additional offspring produced later in life (León 1976; Charlesworth 1980; Stearns 1992). Discounting of a future reward to self should also increase with an increase in the growth rate of the population, as offspring produced earlier in life will have proportionally more descendants in the population than offspring produced later in life when the population is growing (León 1976; Charlesworth 1980; Stearns 1992). Inclusive fitness the-



**Fig. 1** Marginal rate of substitution  $MRS_{k,t}^i$  (given by eqn 6) and discount rate  $\theta_t$  for semelparous populations ( $s \rightarrow 0$ ) with dispersal distribution given by  $p_k = (1-g)g^{k-1}/(1-g^H)$  with  $H = 6$ ,  $n_d = N^6$ , and  $N = 10$ . In panels (A)–(B), the quantities are graphed for fecundity effects as a function of  $t$  for  $d(0, \mathbf{k}) = 0$ ,  $d(0, \mathbf{k}) = 1$ ,  $d(0, \mathbf{k}) = 2$ , and  $d(0, \mathbf{k}) = 3$  and the remaining parameter values are given by  $m = 0.1$ ,  $g = 0.1$ . The curves for  $MRS_{k,t}^i$  when  $d(0, \mathbf{k}) = 2$  and  $d(0, \mathbf{k}) = 3$  are indistinguishable, and in both cases, the substitution rate is very low. As the asymptotic value of the substitution rate is zero ( $MRS_{k,\infty}^i = 0$ ), the curves for  $d(0, \mathbf{k}) = 2$  and  $d(0, \mathbf{k}) = 3$  will decrease after having increased. In panels (C)–(D), the discount rate  $\theta_t$  for future rewards to the focal group ( $d(0, \mathbf{k}) = 0$ ) is graphed for fecundity effects for different values of  $g$  and  $m$  (calculated from  $MRS_{0,t}^i$ ). In panels (E)–(F),  $MRS_{k,t}^i$  and discount rate  $\theta_t$  are graphed for survival effects for exactly the same parameter values as in the first row of the figure.

ory further emphasizes that a future reward to offspring should also be discounted according to the relatedness between actor and recipient, as offspring have a lower chance than the parent to transmit its genes to future generations (Hamilton 1964). Taking all these factors into account, Rogers (1994) showed that, for a panmictic population of humans, individual discounting of a future reward should peak at intermediate age. Other models, which do not include demographic processes, but focus instead on the exogenous uncertainty that individuals have over the risk that a future reward will be realized, have also shown that discounting is unlikely to be constant over time and may be hyperbolic (e.g. Kagel *et al.* 1986; Sozou 1998; Stephens 2002; Dasgupta & Maskin 2005, see Henly *et al.* 2007 for an experimental account).

Individuals not only may discount future rewards to self, or to their offspring, but are also likely to discount future rewards to other individuals living in their group or in other groups from the population. Social discounting involves applying a diminishing weight to group-wide future benefits (Sozou 2009). As most natural populations tend to be spatially structured (Clobert *et al.* 2001; Balloux & Lugon-Moulin 2002), most behaviours expressed by one individual are likely to have a fitness consequence on others, be it from the same or from a different group, in the present or in future times. Natural selection is then likely to have equipped humans and other species with social discount rates.

How natural selection shapes the social discount rate in a group-structured population with random migration and iteroparous reproduction was analysed by Sozou (2009). In this case, social discounting is exponential with the value of future rewards decreasing at a constant rate, which is given by the product of the migration and death rates of individuals living within groups. Evolution favours nonunit (incomplete) social discounting in a group-structured population because the value to an actor of a future reward to its local group is governed by the extent to which beneficiaries are genetically related to the actor (Hamilton 1964, 1970, 1971). In a group-structured population, limited migration and small group size induce individuals sampled from the same group in the same or in different generations to be related, as their genes are more likely to coalesce in a more recent common ancestor than are individuals sampled at random from the population.

Natural populations rarely follow the idealized random patterns of migration of group-structured populations, where migrants choose at random a group from the population. Rather, migrants can preferentially move to neighbouring groups rather than randomly in the landscape, a feature accounted for in models of isolation by distance where individuals live on a regular lattice

and where relatedness decreases as the spatial distance between individuals increases (e.g. Malécot 1973, 1975; Sawyer 1976; Nagylaki 1982). Migrants may also preferentially move to groups located at a small hierarchical distance in a hierarchically clustered population (Sawyer & Felsenstein 1983), where individuals are grouped into families, which are grouped into villages, which are grouped into clans, and so on. A hierarchical structure will also generate a pattern of isolation by distance because the relatedness between individuals taken from two different groups is likely to decrease as the hierarchical distance between the groups increases, as individuals taken at random from larger levels of clustering (from higher hierarchical distances) are less likely to share a recent common ancestor than are individuals taken from lower levels of clustering. Such spatial structure may describe the organization of human societies to some extent at least (Johnson & Earle 1987), but multi-level social structures also occur in insects (Bourke & Franks 1995; Holzer *et al.* 2009) and mammals such as shrews (Fontanillas *et al.* 2004), bats (Kerth *et al.* 2011), baboons (Schreier & Swedell 2009) and elephants (Wittemyer *et al.* 2005). Whether a population is hierarchically clustered (Sawyer & Felsenstein 1983) or follows the more standard models of isolation by distance (Malécot 1975), social discounting may not only be temporal but may also depend on the location of the groups where individuals are sampled and may thus be a function of the migration distribution.

In this study, we evaluate marginal rates of substitution in fitness between present and delayed rewards to other individuals living in hierarchically clustered populations with overlapping generations. These substitution rates provide the selective pressures on intertemporal fitness effects and allow us to determine the social discount rates that are favoured by evolution when individuals affect the fecundity or the survival of recipients living in their group, in other groups from the population, and in both present and future times. Our model incorporates phenotypic effect on recipients that can be arbitrarily extended in space and time. It extends previous models with similar contours but that had different population structures and where phenotypic effects were either not extended in space and time, or which did not consider overlapping generations, or which did not include both fecundity and survival-changing actions (Rogers 1990; Taylor 1992; Taylor & Irwin 2000; Rousset 2004; Taylor *et al.* 2007; Lehmann 2007, 2008; Grafen & Archetti 2008; Johnstone & Cant 2008; Sozou 2009). Although we frame our model in terms of hierarchically clustered populations, which is somewhat easier to analyse mathematically, our main results carry over to the more standard models of isolation by distance of population genetics.

This article is organized as follows. Section ‘Hierarchically clustered populations’ defines a hierarchically clustered population and the life cycle of the individuals in that population. Section ‘Phenotypic effects and preferences’ defines the behaviour of individuals and the marginal rates of substitution in fitness and introduces a social discount rate. Section ‘Evolutionary equilibrium under the joint action of natural selection and genetic drift’ summarizes the population genetic method we use to evaluate the discount rate that is favoured by evolution. Section ‘Results’ presents the new results and discusses their relation to previous models, and Section ‘Discussion’ summarizes these results and highlights their biological relevance.

### Hierarchically clustered populations

We consider a population consisting of a finite number  $n_d$  of groups, each with a constant number  $N$  of adult individuals that have the following life cycle (a list of symbols is given in Table 1). (i) Each of the  $N$  adults in a group produces a large number of juveniles and survives to the next time period with probability  $s$ . (ii) Each juvenile either remains philopatric with probability  $(1-m)$  or disperses with complementary probability  $m$  to another group. (iii) In each group, juveniles compete for vacant breeding spots ( $sN$  on average, with binomial variation) so that exactly  $N$  individuals reach adulthood in each group.

We assume that this population is structured according to the hierarchically clustered group model of Sawyer & Felsenstein (1983). In this model, groups can be thought of as families that are grouped into villages of  $n_1$  families, which, in turn, are grouped into clans of  $n_2$  villages, which are grouped into tribes of  $n_3$  clans, and so on up to  $H$  levels of clustering. The total number of groups in the population can then be written as

$$n_d = n_1 n_2 n_3 \cdots n_H, \quad (1)$$

where  $n_H$  is the number of clusters in the highest level of clustering. When  $H = 1$ , this reduces to the island model of dispersal (Wright 1931).

Let the vector  $\mathbf{i} = (i_1, i_2, i_3, \dots, i_H)$  represent the  $i_1$ th group ( $0 \leq i_1 < n_1$ ) in the  $i_2$ th village ( $0 \leq i_2 < n_2$ ) in the  $i_3$ th clan ( $0 \leq i_3 < n_3$ ), ..., so that the space of groups in the population is given by  $S = \{(i_1, i_2, i_3, \dots, i_H) : 0 \leq i_j < n_j\}$ . We denote by  $d(\mathbf{i}, \mathbf{j})$  the hierarchical distance between groups  $\mathbf{i}$  and  $\mathbf{j}$ , which is defined as the lowest level of clustering necessary to separate an individual randomly sampled in group  $\mathbf{i}$  from that of an individual randomly sampled in group  $\mathbf{j}$ . In particular,  $d(\mathbf{i}, \mathbf{i}) = 0$  means that the two individuals are sampled from the same group (or family);  $d(\mathbf{i}, \mathbf{j}) = 1$  means that the two

**Table 1** List of symbols

Symbol	Definition
$N$	Group size.
$n_d$	Number of groups in the population.
$s$	Survival probability of an individual from one time period to the next.
$m$	Migration rate out of the natal group.
$\mu$	Mutation rate.
$\mathbf{i}$	Position of a group in the space of groups; that is, $\mathbf{i} = (i_1, i_2, i_3, \dots, i_H)$ represents the $i_1$ th group ( $0 \leq i_1 < n_1$ ) in the $i_2$ th village ( $0 \leq i_2 < n_2$ ) in the $i_3$ th clan ( $0 \leq i_3 < n_3$ ), ...
$\mathbf{0}$	Position in the space of groups of the focal group [ $\mathbf{0} = (0, 0, \dots, 0)$ ], which is a representative group of the population.
$d(\mathbf{i}, \mathbf{j})$	Hierarchical distance between group $\mathbf{i}$ and $\mathbf{j}$ .
$p_k$	Probability that an individual migrates to hierarchical distance $k$ .
$n_k$	Number of clusters at the $k$ -th level of clustering.
$r_k$	Total number of groups at hierarchical distance $k$ from the disperser's group.
$C$	Fecundity (or survival) cost of expressing the mutant allele.
$B_{\mathbf{k}, t}$	Fecundity (or the survival) benefit shared among all individuals in group $\mathbf{k}, t$ and resulting from an individual expressing the mutant allele in the focal group in a focal time period (at $\mathbf{k} = \mathbf{0}$ and $t = 0$ ).
$Q_{\mathbf{k}, t}^R$	Probability that an individual randomly sampled from group $\mathbf{k}, t$ carries an allele identical by descent with that from an individual sampled from the focal group.
$P_{\mathbf{k}, t}$	Probability that a line of descent from an individual residing in the focal group at period $t = 0$ will be in group $\mathbf{k}$ at $t$ time periods later
$MRS_{\mathbf{k}, t}^i$	Marginal substitution rate between present rewards to self at $t = 0$ and delayed reward to $\mathbf{k}, t$ group members.
$MRS_{\mathbf{k}, t}^S$	Marginal substitution rate between present rewards to all focal group members at $t = 0$ and delayed reward to $\mathbf{k}, t$ group members.
$\theta_l$	Discount rate of rewards to group members living at hierarchical distance $k$ .
$\tau, \tau_{\mathbf{k}, t}$	Transmission coefficients of mutant genes to the next time period.
$\psi_l$	$l$ -th eigenvalue of the Markov migration matrix.

individuals are sampled in different groups within the same village;  $d(\mathbf{i}, \mathbf{j}) = 2$  that they are sampled from two different villages in the same clan. Likewise,  $d(\mathbf{i}, \mathbf{j}) = 3$  means that the two individuals are sampled from two different clans within the same tribe. More formally, the hierarchical distance between group  $\mathbf{i}$  and  $\mathbf{j}$  is defined as  $d(\mathbf{i}, \mathbf{j}) = 0$  and  $d(\mathbf{i}, \mathbf{j}) = k$  if  $i_k \neq j_k$  and  $i_l = j_l$  for  $l > k$  (Sawyer & Felsenstein 1983, eqn 2.2).

We denote by  $m(\mathbf{i}, \mathbf{j})$  the probability that an individual migrates from group  $\mathbf{i}$  to group  $\mathbf{j}$ . This probability is

assumed to depend only on the relative hierarchical distance between groups. Provided we define  $\mathbf{j} \pm \mathbf{i}$  as  $j_k \pm i_k$  modulo  $n_k$  for all  $k$ , we can then write  $m(\mathbf{i}, \mathbf{j}) = m(0, \mathbf{j} - \mathbf{i})$ . The step distribution  $m_i \equiv m(0, \mathbf{i})$  is then spatially homogeneous and can be written as

$$m_{\mathbf{j}} = \frac{p_k}{r_k} \quad (2)$$

for  $d(0, \mathbf{j}) = k$  (Sawyer & Felsenstein 1983), where  $p_k$  is the probability that an individual migrates to hierarchical distance  $k$  ( $\sum_{k=0}^H p_k = 1$  and with  $p_0 = (1-m)$  being the probability that the individual remains in its group),  $r_0 = 1$ , and  $r_k = n_1 n_2 n_3 \dots (n_k - 1)$  is the total number of groups at hierarchical distance  $k$  from the disperser's group, which gives the number of groups the disperser may settle in.

We assume that  $0 < m < 1$  and  $p_k > 0$  for all  $k$  so that an immortal line of descent of an individual may possibly visit any group in the population. The Markov transition matrix determined by the step distribution  $m_{\mathbf{j}}$  (the migration matrix) can then be seen to define an aperiodic random walk of this line of descent in the space of groups ( $S$ ), a feature that will be exploited in our analysis.

### Phenotypic effects and preferences

We assume that the individuals in the population express a genetically determined social behaviour (phenotype), which may affect the vital rates (here fecundity or survival) of the actor and that of other individuals in the population, the recipients of the actor's behaviour. In order to describe the phenotypic effects of the expression of the behaviour on various classes of recipients, we focus on a representative individual in the population, called the focal individual. This individual lives in a focal group, which is taken without loss of generality to be the group located at  $\mathbf{0} = (0, 0, \dots, 0)$  in the space of colonies  $S$ .

We assume that the focal individual expresses a behaviour that may increase the fecundity (or the survival) of the whole set of individuals living in group  $\mathbf{k}$  at  $t$  time periods in the future by  $B_{\mathbf{k}, t}$ , where by a time period we mean one iteration of the life cycle described previously. That is,  $B_{\mathbf{k}, t}$  is the total fecundity (or the survival) benefit shared among all individuals in group  $\mathbf{k}, t$  and resulting from the focal individual expressing its behaviour in a focal time period. This extended phenotypic effect is assumed to come at some fecundity (or survival) cost  $C$  to the focal individual and may either involve an active transfer of resources between individuals that are in contact or result from a modification of the physical environment in future time periods (e.g.

niche construction, Odling-Smee *et al.* 2003) without actor and recipient being ever in direct contact.

In order to ascertain how evolution shapes individuals to trade-off between the present effect  $C$  on self and the delayed group-wide effect  $B_{\mathbf{k}, t}$ , we introduce the ratio

$$\text{MRS}_{\mathbf{k}, t}^i \equiv \frac{C}{B_{\mathbf{k}, t}}, \quad (3)$$

which is defined as the number of units of its own fecundity (or survival) that a focal individual is willing to exchange for one unit of fecundity (or survival) accruing to the whole set of individuals living in group  $\mathbf{k}$  at  $t$  time periods in the future. Additionally, we also want to ascertain how evolution shapes individuals to trade-off between the present effect  $B_{0,0}$  on all focal group members in the present and delayed effect  $B_{\mathbf{k}, t}$ . To that aim, we introduce the ratio

$$\text{MRS}_{\mathbf{k}, t}^s \equiv -\frac{B_{0,0}}{B_{\mathbf{k}, t}}, \quad (4)$$

which is the number of units of fecundity (or survival) to the focal individual's group the focal individual is willing to exchange for one unit of fecundity (or survival) accruing to the whole set of individuals living in group  $\mathbf{k}, t$  (hence  $C$  has been replaced by  $-B_{0,0}$ , where the minus sign reflects the fact that we are looking at a fitness decrease).

Both  $\text{MRS}_{\mathbf{k}, t}^i$  and  $\text{MRS}_{\mathbf{k}, t}^s$  are marginal rates of substitution in fitness (Hansson & Stuart 1990; Rogers 1994). As these substitution rates represent cost-to-benefit ratios of expressing behaviours, they are traditionally used to ascertain how evolution allocates optimally resources to competing ends (e.g. Maynard-Smith 1982; Frank 1998). But the substitution rates can also be used to evaluate discount rates. Allocation of resources between the present and some time horizon  $t$  can be determined by the long-term discount rate, which is the rate per time step at which rewards are discounted such as to yield the cost-benefit ratio over  $t$  time steps (e.g. Rogers 1994; Hirshleifer *et al.* 2006). Evaluating the discount rate is useful as it allows one to conveniently capture the trade-off between present and delayed rewards, and to determine whether discounting is constant over time or hyperbolic (Rogers 1994; Green & Myerson 1996; Laibson 1995; Sozou 1998).

We define the discount rate  $\theta_l$  from the marginal substitution rate  $\text{MRS}_{\mathbf{k}, t}^s$  as

$$\text{MRS}_{\mathbf{k}, t}^s = (1 - \theta_l)^t \Leftrightarrow \theta_l = 1 - \left( \text{MRS}_{\mathbf{k}, t}^s \right)^{\frac{1}{t}} \quad (5)$$

for  $d(0, \mathbf{k}) = l$ . Hence,  $\theta_l$  gives the rate per time step at which a benefit to the group located at hierarchical dis-

tance  $l$  from the focal group should be discounted such as to yield the marginal substitution rate over  $t$  time steps, and where by definition  $MRS_{0,0}^s = 1$ . The discount rate  $\theta_l$  also captures discounting of phenotypic effects on groups located at different hierarchical distances because rewards to these groups will be discounted differently if the cost-to-benefit ratio (marginal substitution rate) of affecting the fitness of the group members differs. For instance, lower substitution rates with increasing hierarchical distance will result in higher discounting.

Equation (5) allows one to evaluate the discount rate from fitness effects (Rogers 1994; Sozou 2009), and the main aim of this study is it to evaluate  $MRS_{k,t}^i$  and the discount rate  $\theta_l^1$  that are favoured by evolution explicitly in terms of the demographic parameters  $N$ ,  $s$  and  $m_j$ .

### Evolutionary equilibrium under the joint action of natural selection and genetic drift

In order to evaluate the substitution rates ( $MRS_{k,t}^i$ ,  $MRS_{k,t}^s$ ) and the discount rate  $\theta_l$ , we assume that the individuals in the population are haploid and bear a single genetic locus coding for the behaviour under interest and that only two different alleles can simultaneously segregate in the population: a mutant allele, whose carriers express the behaviour with effects  $-C$  on self and  $B_{k,t}$  on others for evaluating  $MRS_{k,t}^i$  (or effects  $-B_{0,0}$  on self and  $B_{k,t}$  on others for evaluating  $MRS_{k,t}^s$ ), and a resident allele, whose carriers receives the fecundity or survival benefits but pay no cost. We also make the standard simplifying assumption of weak selection (the  $C$  and  $B_{k,t}$ 's are of small order  $\delta$ ), without which it would very difficult to obtain explicit analytical expressions of  $MRS_{k,t}^i$  and  $MRS_{k,t}^s$  in structured populations.

At an evolutionary equilibrium, the fixation probability of a single mutant allele resulting in phenotypic effects  $C$  and  $B_{k,t}$  must be equivalent to the fixation probability of a single resident allele. The fixation probability captures the overall effects of both genetic drift and natural selection on gene frequency change, from the appearance as a single copy to the eventual fixation or loss of the mutant from the population (Foster & Young 1990; Rousset & Billiard 2000; Fudenberg & Imhof 2006).

We use population genetic techniques in order to evaluate the weak selection effects of the mutant allele

on its fixation probability and find that the marginal rates of substitution in fitness can be expressed at an evolutionary equilibrium as

$$\begin{aligned} MRS_{k,t}^i &= \frac{\tau_{k,t}}{\tau_{\bullet}} \\ MRS_{k,t}^s &= \frac{\tau_{k,t}}{\tau_{0,0}}, \end{aligned} \quad (6)$$

where the coefficients  $\tau_{\bullet}$ ,  $\tau_{0,0}$  and  $\tau_{k,t}$  will depend on the demographic parameters  $N$ ,  $s$ ,  $m_j$ , and whether fitness effects are fecundity or survival-changing, but not on  $C_{k,t}$  and  $B_{k,t}$ . A detailed justification of eqn 6 is given in Appendix I. Equation 6 entails that  $MRS_{0,0}^s = 1$  and it also illustrates that  $\theta_l$  can be used to discount the future rewards at a direct cost to self because  $MRS_{k,t}^i = MRS_{0,0}^i MRS_{k,t}^s$  so that the rate of decline with delay of  $MRS_{k,t}^i$  is the same as that of  $MRS_{k,t}^s$ .

The coefficients  $\tau_{\bullet}$  and  $\tau_{k,t}$  can be thought of as standardized transmission rates of mutant gene copies over one time period and the right member of eqn 6 as scaled relatedness coefficients where the effect of local competition has been included (Queller 1994; Grafen & Archetti 2008). Namely, the substitution rates can be thought of as above-average transmission rates of replicate copies of the mutant allele by individuals in group  $k,t$  relative to the transmission rate of the mutant by the focal individual (or average group member for  $MRS_{k,t}^s$ ) living in the present. For instance, for fecundity effects, we find that we can write

$$\tau_{\bullet} = -\lim_{\mu \rightarrow 0} \frac{(1-s)}{1-Q_{0,0}} \left[ 1 - \sum_i \sum_j m_i m_{i-j} Q_{j,0}^R \right], \quad (7)$$

where  $Q_{j,0}^R$  ( $Q_{j,0}$ ) is the probability that a gene sampled in the focal individual and a homologous gene randomly sampled with replacement (without replacement) in group  $j$  are identical by descent, and  $\mu$  is the mutation rate from one allele to another and the mutation model is taken here to be the infinite allele model (Kimura & Crow 1964) (see Appendix I for details).

The factor  $(1-s)/(1-Q_{0,0})$  in eqn 7 is always positive and thus does not affect the forthcoming results. The first term in square brackets, '1', in eqn 7 can be interpreted as the direct fitness cost to a focal individual from expressing a mutant allele that reduces its fecundity by one unit, which thus results in a loss of mutant alleles sent into the next time period, while the second term as the indirect fitness benefit stemming from the decrease in competition faced by relatives of the focal individual when it decreases its fecundity by one unit. This competition term depends on the probability  $\sum_i m_i m_{i-j}$  that an offspring of the focal individual com-

<sup>1</sup> The discount rate  $\theta_l$  can be interpreted as a discrete analogue of the average evolutionary discount rate of Rogers (1994, eqn 15). The average being over the instantaneous discount rate (or short-term discount rate), which gives the rate at which a future reward is discounted over a single given time step (Laibson 1995; Sozou 1998).

petes against an offspring produced in group  $j$ , which is related to the focal individual by  $Q_{j,0}^R$ .

The coefficient  $\tau_{k,t}$  is

$$\tau_{k,t} = \lim_{\mu \rightarrow 0} \frac{(1-s)}{1-Q_{0,0}} \left[ Q_{k,t}^R - \sum_i \sum_j m_i m_{i-j} Q_{j-k,t}^R \right], \quad (8)$$

where  $Q_{k,t}^R$  can be interpreted as the probability that a gene sampled in the focal individual and a homologous gene randomly sampled in group  $k$  at  $t$  time periods later are identical by descent. The first term in square brackets can be interpreted as the indirect fitness benefit accruing to a focal individual from increasing the fecundity of recipients in group  $k,t$  by a unit amount and the second term is the indirect fitness cost stemming from increasing the competition faced by relatives of the focal individual living in other groups and resulting from increasing the fecundity of recipients in group  $k,t$ . The transmission coefficients for the survival effects take a similar form to eqns 7–8 (see eqns A.13–A.14 in Appendix D), and this form emphasizes a general feature of evolution in structured population, which is that a behaviour resulting in an increase in the fecundity or survival of various classes of recipients may also decrease the fitness of the same or other classes of recipients through the concomitant increase in competition (e.g. Grafen 1984; Taylor 1992; Queller 1994; Taylor & Irwin 2000; West *et al.* 2007; Grafen & Archetti 2008; Johnstone & Cant 2008; Sozou 2009).

It is worth recalling at this point that the phenotypic effects  $B_{k,t}$  may either be positive (incrementing the fitness of the recipients), in which case we refer to the behaviour as ‘helping’, or negative, in which case we refer to the behaviour as ‘harming’. Depending on the structure of the population, the coefficients  $\tau_{k,t}$  may turn out to be negative so that selection favours harming. This occurs if individuals from group  $k,t$  are negatively related to the focal individual, in which case the absolute value of  $MRS_{k,t}^i$  is taken in order to evaluate the discount rate  $\theta_i$  defined by eqn 5. Both helping and harming may generally qualify in our model as, respectively, altruism and spite *sensu* Hamilton (1964, 1970). But in some situations, this may not be the case (especially when  $t = 0$ ), since whether a behaviour qualifies as altruism and spite in the above sense depends upon its effect on the fitness of the actor (eqns A.2 and A.7), which may depend in a complex way on the demographic parameters of the model (e.g.  $N$ ,  $s$ ,  $m_j$ ). The explicit calculations of the coefficients  $\tau_{\bullet}$  and  $\tau_{k,t}$  determining the selection pressure on helping and harming is carried out in Appendix I.

Finally, it is also worth mentioning that the phenotypic effect  $B_{k,t}$  are not necessarily the result of a single and

unilateral action taken by the focal individual. On the contrary, it captures a total change in the vital rate of individuals in group  $k,t$  stemming from the focal individual expressing the mutant allele over one iteration of the life cycle. Hence,  $B_{k,t}$  is a net genetic effect and may be the outcome of multimove social interactions (sequence of actions during an individual’s lifespan) involving the interaction between the phenotypic expressions or behavioural response rules of any number of individuals living in present and past generations of the population (e.g.  $N$ -players repeated prisoner’s dilemma game, negotiation game, optimal foraging game, repeated rounds of cultural transmission, ecological public goods game, Lehmann & Rousset 2010, section 7; an explicit calculation of the  $B_{k,t}$  coefficients for an ecological public goods game is given in Lehmann 2008, eqn 17).

## Results

### Marginal substitution rates

*Fecundity effects.* When the phenotypic effects  $C$  and  $B_{k,t}$  are on the fecundity of individuals, we find that

$$MRS_{k,t}^i = \begin{cases} \frac{\delta_{k0} + \mathcal{L}_k(F) - (1+s)/n_d}{N + \mathcal{L}_0(F) - (1+s)/n_d} & \text{if } t = 0 \\ \frac{\mathcal{L}_k(G_t) - (1+s)/n_d}{N + \mathcal{L}_0(F) - (1+s)/n_d} & \text{otherwise,} \end{cases} \quad (9)$$

where  $\delta_{k0}$  is the Kronecker-Delta ( $\delta_{k0} = 1$  if  $k = 0$ , zero otherwise),  $\mathcal{L}_k(D)$  is the inverse Fourier transform of a given function  $D$  at  $k$  (eqn A.31 in Appendix II), and

$$F(\psi) = \frac{2s\psi}{1+s+(1-s)\psi},$$

$$G_t(\psi) = \frac{(1+s)(1+\psi)[s+(1-s)\psi]^t}{1+s+(1-s)\psi}, \quad (10)$$

where  $\psi$  is the characteristic function of the dispersal distribution (eqn A.17). The marginal rate of substitution in fitness  $MRS_{k,t}^i$  is given by eqn 9 with  $N$  replaced by one. The proof of eqn 9 is given in Appendix III, see in particular eqns A.38 and A.41.

*Survival effects.* When the phenotypic effects  $C$  and  $B_{k,t}$  are on the survival of individuals, we find that

$$MRS_{k,t}^i = \begin{cases} \frac{\delta_{k0} + \mathcal{L}_k(F) - (1+s)/(2n_d)}{N + \mathcal{L}_0(F) - (1+s)/(2n_d)} & \text{if } t = 0 \\ \frac{\mathcal{L}_k(G_t) - (1+s)/(2n_d)}{N + \mathcal{L}_0(F) - (1+s)/(2n_d)} & \text{otherwise,} \end{cases} \quad (11)$$

where

$$F(\psi) = -\frac{(1-s)\psi}{1+s+(1-s)\psi},$$

$$G_t(\psi) = \frac{(1+s)[s+(1-s)\psi]^t}{1+s+(1-s)\psi}. \quad (12)$$

The proof of eqn 11 is given in Appendix III, see eqns A.48 and A.51.

*Explicit Fourier transform.* Equations 9 and 11 apply both to hierarchically clustered populations and to the more classical homogeneous isolation-by-distance models of population genetics, where groups of individuals are positioned at the nodes of a regular lattice (e.g. a circular lattice in one dimension and a torus in two dimensions, Maruyama 1970b; Malécot 1975; Sawyer 1976; Nagylaki 1982; Taylor 1992) because in both cases the migration distribution is translationally invariant:  $m(\mathbf{i}, \mathbf{j}) = m(0, \mathbf{j}-\mathbf{i})$  for  $\mathbf{i}, \mathbf{j} \in S$ . The inverse Fourier transform for isolation-by-distance models is available but does not usually reduce to simple expressions, and the analysis turns out to be somewhat easier for hierarchically clustered population. For this latter case, the inverse transform of a given function  $D(\psi)$  (here  $F(\psi)$  or  $G_t(\psi)$ ) takes the form

$$\mathcal{L}_h(D(\psi)) = \frac{D(\psi_0)}{n_d} - \frac{D(\psi_l)}{n_1 n_2 \cdots n_l} + \sum_{j=l+1}^H \frac{D(\psi_j)(n_j-1)}{n_1 n_2 \cdots n_j} \quad (13)$$

for  $d(0, \mathbf{h}) = l$ , where the middle term does not occur if  $l = 0$ . Further,  $\psi_0 = 1$ , and for  $l > 0$

$$\psi_l = \sum_{j=0}^{l-1} p_j - p_l / (n_l - 1), \quad (14)$$

which is the characteristic function of the dispersal distribution evaluated at  $d(0, \mathbf{h}) = l$ . The proof of eqn 13 is given in Appendix V.

#### Common features of the substitution rates

As  $s + (1-s)\psi_l < 1$  for  $l > 0$  if  $s < 1$  and  $G_t(\psi_0) = 1 + s$  for fecundity effects and  $G_t(\psi_0) = (1+s)/2$  for survival effects, we obtain by substituting eqn 13 into eqns 9 and 11 that  $\text{MRS}_{\mathbf{k}, \infty}^i = 0$  and  $\text{MRS}_{\mathbf{k}, \infty}^s = 0$  for both survival and fecundity effects if individuals have a non-zero probability of dying. The present fitness value of an effect on the fecundity or on the survival of any group of individuals in the very long run ( $t \rightarrow \infty$ ) is thus zero, which stems from the fact that the relatedness between a focal individual sampled in the present and any recipient sampled in the very long run from the population must be equal to zero. This can be seen more explicitly by writing, for instance,  $\text{MRS}_{\mathbf{k}, t}^i$  for  $d(0, \mathbf{k}) = l$  and for fecundity effects when  $t > 0$  as

$$\text{MRS}_{\mathbf{k}, t}^i = \frac{\sum_{j=l+1}^H \frac{y_j(1+s)(1+\psi_j)[s+(1-s)\psi_j]^t}{N + \sum_{j=1}^H y_j 2s\psi_j}}{-\frac{y_l(1+s)(1+\psi_l)[s+(1-s)\psi_l]^t}{N + \sum_{j=1}^H y_j 2s\psi_j}}, \quad (15)$$

where  $y_j = (n_j-1)/[(n_1 n_2 \cdots n_j)\{s+(1-s)\psi_j\}]$  and the second term does not occur if  $d(0, \mathbf{k}) = 0$ . Each term of this equation goes to zero as  $t \rightarrow \infty$  if  $s < 1$ , but it also illustrates that the discounting of future rewards involves a sum of exponential functions, so discounting is unlikely to be exponential. Further, as  $\psi_l$  can be close to one for certain dispersal distribution, the temporal discounting can be extremely low. So how does the valuation of future group-wide benefit falls off with delay?

In a general perspective, be it for hierarchically clustered or isolation-by-distance models, the stationary probabilities of identity by descent (the  $Q_{\mathbf{k}, t}$ 's) can be expressed in terms of the powers of the nonunit eigenvalues of the Markov matrix describing the random walk of a gene lineage across space and across age classes (whose  $l$ th eigenvalue is given by  $s + (1-s)\psi_l$  for hierarchically clustered populations, see eqn 14). As can be seen from eqn 15, the substitution rates are then rational functions of such expressions and do not generally reduce to a single rational expression. The discount rate is then variable and can decrease or even transiently increase over time and may thus result in hyperbolic discounting.

In the long run, the substitution rate is dominated by the largest subdominant eigenvalue of the Markov matrix and the discount rate may then be considered approximately constant. If individuals are more likely to migrate to a small hierarchical distance than to a longer one (dispersal is localized) such that  $p_i > p_{i+1}(n_i-1)/(n_{i+1}-1)$ , which implies  $\psi_{i+1} > \psi_i$  for  $i > 0$  from eqn 14, then the subdominant eigenvalue is given by  $\lambda_H = s + (1-s)\psi_H$ . The valuation of future rewards then decreases asymptotically at rate  $1-\lambda_H = (1-s)(1-\psi_H)$  per unit delay, where  $\psi_H = 1-p_H n_H / (n_H-1)$ . In the isolation-by-distance models on an unbounded lattice, an infinite number of subdominant eigenvalues are arbitrarily close to unity (Maruyama 1970b), so that in practice for large lattices, no eigenvalue can be singled out to characterize the long-term behaviour of the discount rate. In that case too, the discount rate can transiently increase over time. Hence, in a geographically structured population, there is no reason to believe that evolution may favour exponential social discounting, where valuation of future rewards falls by a constant factor per unit time.

We also note that the sign of  $\text{MRS}_{\mathbf{k}, t}^i$  is always negative for  $d(0, \mathbf{k}) = H$ , which holds for both fecundity and



survival effects. Hence, an actor is selected to harm an individual from the highest level of clustering. This stems from the fact that such individuals will be negatively related to the actor. Equation 15 shows that the lowest the hierarchical distance between actor and recipient, the more likely an actor should help the recipient, but the spatial discounting of a future reward (or penalty) to other group members depends on the value of the various life history parameters and does not decrease proportionally with the hierarchical distance. Apart from the above observations, eqns 9 and 11 (or eqn 15) remain complicated. It is thus worthwhile to analyse special cases of these equations in order to gain a better intuition about how selection shapes spatial and temporal discounting. This is carried out in the next section.

### Specific cases

*Infinite island model of dispersal.* Here, we assume that when individuals migrate, they choose a group at random from the whole population to settle in, so that migration is panmictic. This situation can be described as  $p_0 = 1 - m$ ,  $p_k = mr_k / (n_d - 1)$  for  $k > 0$  in eqn 2, where  $m$  is the migration rate, which gives  $m_0 = 1 - m$ , and  $m_j = m / (n_d - 1)$ . It is also convenient to assume that  $n = n_1 = \dots = n_H$ , in which case we obtain from eqn 14 that  $\psi_l = 1 - mn_d / (n_d - 1)$  for  $l > 0$ . If we further assume that the number of groups becomes very large (say  $n_d \rightarrow \infty$ ), then the probability of identity by descent between actor and recipient sampled from two different groups vanishes. Only individuals within groups will be related, and it can be checked that  $MRS_{\mathbf{k},t}^i = 0$  and  $MRS_{\mathbf{k},t}^s = 0$  in the limit for all  $\mathbf{k}$  except  $\mathbf{k} = \mathbf{0}$  [ $D(\psi_t)$  is a constant in eqn 13, in which case the whole expression is equal to zero for  $\mathbf{k} \neq \mathbf{0}$  and so are eqns 9 and 11]. For fecundity effects, we then find that for all  $t$

$$MRS_{\mathbf{0},t}^i = \frac{(1+s)(2-m)[s+(1-s)(1-m)]^t}{N[2-m(1-s)]+2(1-m)s}, \quad (16)$$

while for survival effects, we have

$$MRS_{\mathbf{0},t}^i = \frac{(1+s)[s+(1-s)(1-m)]^t}{N[2-m(1-s)]-(1-m)(1-s)}. \quad (17)$$

In the presence of complete migration, ( $m = 1$ ), both eqns 16 and 17 reduce to  $MRS_{\mathbf{0},t}^i = s^t / N$ . Here, an actor is then unrelated to group members living  $t$  time periods in the future except if it has survived until that stage (probability  $s^t$ ) in which case it may benefit from its own behaviour expressed  $t$  time periods earlier by magnitude  $1/N$ . We then have  $MRS_{\mathbf{0},t}^s = s^t$ , and the dis-

count rate is then simply given by the probability of dying  $\theta_t = (1-s)$  (Rogers 1994; Sozou 2009).

More generally, comparing eqns 16–17 shows that the marginal rate of substitution in fitness is higher under fecundity than survival effects. This stems from the fact that by reducing the probability that group members die, fewer breeding spots are vacated and available to the offspring of the focal individual that do not migrate. By contrast, by increasing the fecundity of group members, the focal individual decreases the chance that its offspring settle in an available breeding spot only if they do not disperse and compete against nondispersers. The effect on local competition of providing a fitness reward to group members is thus stronger under survival effects, a feature first noted by Taylor & Irwin (2000) for a model where  $t = 0$ , in which case eqns 16–17 exactly reduce to their result, and their observation about competition extends to hierarchically and isolation-by-distance structured populations.

Setting  $N = 1$  in eqns 16–17 gives  $MRS_{\mathbf{0},t}^s = [s + (1-s)(1-m)]^t$  for both models, which in turn yields the discount rate  $\theta_t = m(1-s)$ . Future reward to the local group should thus be discounted at a constant rate per unit delay, which is given by the product of the probability of migration and death per time period. This agrees with the result of Sozou (2009) established under a continuous time model with fecundity effects and shows that his result holds for both fecundity and survival effects.

*Semelparous reproduction.* We now assume that all individuals have a zero survival probability ( $s \rightarrow 0$ ), in which case individuals reproduce only once (they are semelparous) and the underlying reproductive scheme corresponds to the standard Wright–Fisher model of population genetics (Ewens 2004). Letting  $s \rightarrow 0$  in eqn 9, we have  $F = 0$  and  $G_t = \psi^t$ , which produces for all  $t$

$$MRS_{\mathbf{k},t}^i = \frac{n_d P_{\mathbf{k},t} - 1}{n_d N - 1}, \quad (18)$$

where  $P_{\mathbf{k},t} = \mathcal{L}_{\mathbf{k}}(\psi^t)$  can be interpreted as the probability that a line of descent from an individual residing in the focal group will be in group  $\mathbf{k}$  at  $t$  time periods in the future ( $\mathcal{L}_{\mathbf{k}}(\psi^t)$  is the inverse transform of the  $t$ -fold convolution of the characteristic function of the dispersal distribution). Equation 18 is equivalent to results found for the evolution of extended phenotypes under a standard model of isolation by distance (Lehmann 2008, eqns 10 and A.21).

As  $\sum_{\mathbf{k}} P_{\mathbf{k},t} = 1$  for all  $t$  and  $P_{\mathbf{0},0} = 1$ , we have  $P_{\mathbf{k},0} = 0$  for  $\mathbf{k} \neq \mathbf{0}$ , which illustrates that selection may favour

helping of the individuals from the focal group in the time period where the behaviour is expressed, while selection may favour harming individuals living in any other group, be it from the same village, clan or tribe. Similar results were observed under a standard isolation-by-distance pattern of population structure, where individuals living in groups at any other lattice point than the focal group should be harmed by the focal individual (as implied by Rousset 2004, eqn 7.21).

When dispersal is preferentially between hierarchically closer levels, so that  $p_i > p_{i+1}$  for all  $i$ ,  $P_{0,t}$  is a decreasing function of  $t$  (see eqn A.66). But  $P_{k,t}$  is likely to first increase over time, then decrease, before approaching the uniform distribution asymptotically:  $P_{k,\infty} = 1/n_d$  (see eqn A.65), in which case  $MRS_{k,\infty}^i = 0$ . This suggests that under localized dispersal in a hierarchically clustered population, the present value of a future reward to the local group is likely to decrease as time passes by, while the present value of a future penalty to the neighbouring groups may increase as well as decrease with  $t$ . If  $n_d P_{k,t} - 1 > 0$  for some  $t$ , then helping instead of harming should be favoured. More generally, for an arbitrary dispersal distribution or for localized dispersal on a regular lattice,  $P_{0,t}$  may also transiently increase over time, but it is not clear whether this is realistic.

In order to be able to evaluate  $P_{k,t}$  quantitatively, we assume that the dispersal distribution is given by a truncated geometric distribution:  $p_0 = 1 - m$  and  $p_k = m(1-g)g^{k-1}/(1-g^H)$  for  $k > 0$ , where the parameter  $g$  (varying between zero and one) allows one to investigate a continuum of spatial structures, ranging from the case where individuals tend to migrate only to the nearest cluster when  $g \rightarrow 0$  or tend to choose a level of clustering at random when  $g \rightarrow 1$ , in which case  $p_k \rightarrow m/H$ . Figure 1 shows that for a population with six levels of clustering, where  $n_k = n$  for all  $k$ , and  $n = 10$  and  $N = 10$  ( $10^7$  individuals in the population), individuals from the same group and village should be helped, while the substitution rate  $MRS_{k,t}^i$  (eqn 6) is vanishingly low for  $d(\mathbf{0}, \mathbf{k}) > 1$ . Panel (B) of Fig. 1 also illustrates that the discount rate  $\theta_l$  of a future reward to the local group decreases very slowly and is approximately constant. But panel (B) of Fig. 1 also illustrates that the discount rate of a future reward to individuals from other groups actually falls rapidly for small delays, but then slowly for longer delays, namely discounting is hyperbolic (Rogers 1994; Green & Myerson 1996; Sozou 1998). Finally, the peak of the bump in the curve describing the discount rate for  $d(\mathbf{0}, \mathbf{k}) = 2$  and  $d(\mathbf{0}, \mathbf{k}) = 3$  in panel (A) of Fig. 1 delineates the region where individuals should be helped from that where they should be harmed. That is, for  $t$  smaller than the peak of the bump (located here at  $t = 1$ ), individuals should be harmed but the gains of harming are extremely small

because  $\theta_l$  is high, and for  $t$  larger than the peak of the bump, individuals should be helped (see Fig. 2 here below where this is more apparent).

In panels (A)–(B) of Fig. 1, the migration rate is assumed to be small ( $m = 0.1$ ) and that dispersal is localized ( $g = 0.1$ ). But when the migration rate becomes larger and dispersal remains localized, the discount rate  $\theta_l$  of a future reward to the focal group is no longer approximately constant and becomes hyperbolic [panels (C)–(D) of Fig. 1]. That an increase in migration results in a variable discount rate can be read out of the eigenvalues of the migration matrix. Assuming a large number of groups and a large number of levels of clustering ( $n_d$  and  $H$  large), the  $l$ -th eigenvalue is approximately  $\psi_l = 1 - mg^{l-1}$  (see eqn A.60 of Appendix V). When  $m$  and  $g$  are small, the different eigenvalues will approximately take the same value and thus dictate a similar rate of decay on  $MRS_{0,t}^i$  (eqn 15), which decreases proportionally to this value. But as  $m$  increases, the difference between the different values the eigenvalues take becomes stronger and so  $MRS_{0,t}^i$  no longer decays constantly.

For survival effects, we were unable to find a compact expression for  $MRS_{k,t}^i$  because when  $s \rightarrow 0$  in eqn 11, we have  $F = -\psi/(1 + \psi)$  and  $G_t = \psi^t/(1 + \psi)$ , which does not simplify easily. Nevertheless, we see that for  $t=0$ , each individual, except those of the focal group, should be harmed and more so than under fecundity effects. As was observed for fecundity effects, as time goes by the present value of a future reward to the local community decreases, while that to other groups may increase [panels (A)–(B) of Fig. 2]. The figure also shows that discounting will generally not be exponential when the migration rate becomes larger and dispersal remains localized [panels (C)–(D) of Fig. 2], which is the same as that observed above for semelparous populations. Likewise, the value of future rewards will be higher under fecundity that survival effects [compare panels (A) and (E) of Fig. 1], and the bump in the curves for  $d(\mathbf{0}, \mathbf{k}) = 2$  and  $d(\mathbf{0}, \mathbf{k}) = 3$  again delineates regions of helping and harming, which also shows that harming is more likely to be selected for under survival than fecundity effects, holding everything else constant.

*Iteroparous reproduction.* We now allow individuals to survive from one time period to the next. When  $s$  takes positive values and dispersal is localized, the expression for  $MRS_{k,t}^i$  remains complicated. By increasing the survival probability  $s$ , the relatedness between group members tends to increase, and we thus expect that  $MRS_{k,t}^i$  will be higher than in the previous section, everything else being held constant. This is indeed the case and is illustrated in panels (A)–(B) of Fig. 2. As was observed

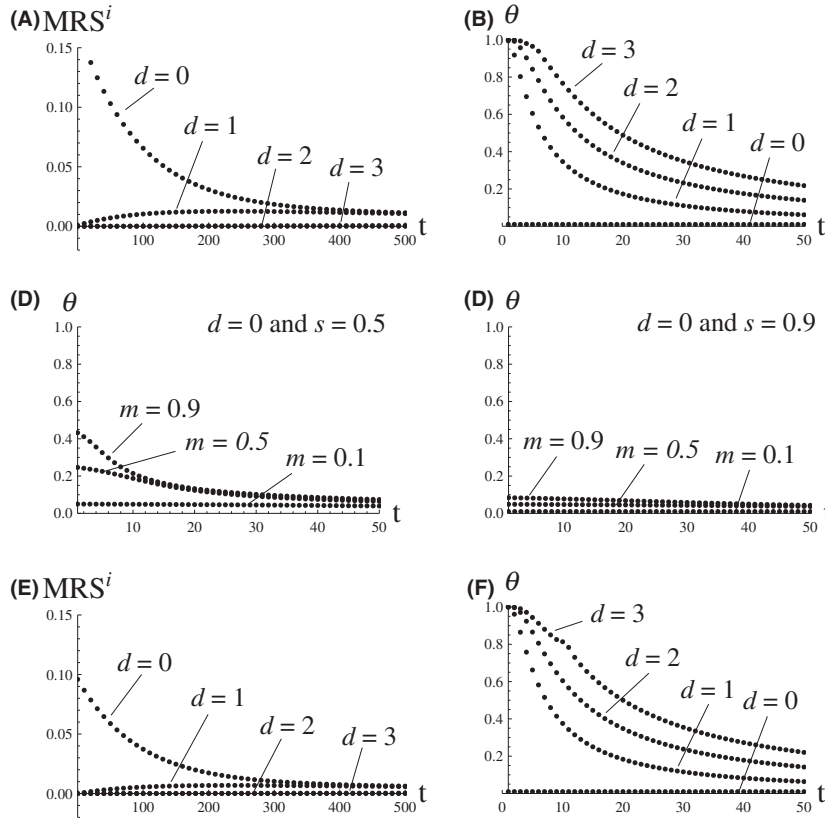


Fig. 2 Same as Fig. 1, but for iteroparous populations with  $s = 0.9$  except in panels (A)–(B), where  $\theta_t$  for future rewards to the focal group ( $d(0, \mathbf{k}) = 0$ ) is graphed for fecundity effects for different values of  $s$  and  $m$ .

in the previous section ( $s \rightarrow 0$ ), when the migration rate becomes larger and dispersal remains localized, the discount rate  $\theta_t$  of a future reward to the focal group becomes hyperbolic [panels (C)–(D) of Fig. 2], but the effect is mitigated by survival because as  $s$  increases, the eigenvalues of the transition matrix become more similar to each other ( $\lambda_t = s + (1-s)\psi_t$ ). Panels (E)–(F) of Fig. 2 show that future rewards will be more valuable under fecundity than survival effects, which was also observed in the previous section. Finally, comparing Figs 1 and 2 illustrates that, everything else being equal, the discount rate per unit delay is likely to be larger for iteroparous than for semelparous populations.

Letting  $s \rightarrow 1$  in eqn 9 (which gives  $F(\psi) = \psi$  and  $G_t(\psi) = 1 + \psi$ ), we find that for fecundity effects and for all  $t$ , we have

$$MRS_{k,t}^i = \frac{n_d(\delta_{k0} + m_k) - 2}{n_d(N + m_0) - 2}, \tag{19}$$

which is directly determined by the migration distribution  $m_k$  and is independent of  $t$  so that there is no temporal discounting. In the limit as  $s$  approaches 1, at most one individual dies per time interval, and then one can formulate the process over time intervals in

which exactly one individual dies, which produces the Moran model of reproduction (Ewens 2004). Equation 19 actually agrees exactly with the qualitative results found for the Moran process under a standard isolation-by-distance population structure and when phenotypic effects are intra-temporal and only on nearest-neighbours (Ohtsuki & Nowak 2006; Taylor *et al.* 2007), and when the phenotypic effects are arbitrarily extended in space (Lehmann *et al.* 2007, eqn 8). Hence, eqn 19 suggests that for the Moran process, extended phenotypic effects in time should not be discounted when the number of individuals in the population becomes infinitely large, which may merit further investigations. Finally, if we assume that  $n_k = n$  for all  $k$ , then we have  $n_d m_k \approx n^{H-k} m p_k$  so that individuals living at hierarchical distance  $k$  should be helped if  $n^{H-k} m p_k > 2$ , and harmed otherwise.

For survival effects, we find that when  $s \rightarrow 1$  in eqn 11, we have

$$MRS_{k,t}^i = \frac{n_d \delta_{k0} - 1}{n_d N - 1}, \tag{20}$$

which is independent not only of time but also of the spatial structure. In this case, individuals from every

group in the population should be harmed except those from the focal group. This again agrees exactly and generalizes known results found under the Moran process and going under the heading of the 'BD protocol' (Ohtsuki & Nowak 2006; Taylor *et al.* 2007), which can be biologically interpreted as representing a situation with phenotypic effects on survival (Lehmann & Rousset 2010, section 4c).

*Small migration.* We now assume that individuals migrate only with small probability out of their natal group (weak migration) but let all other life history parameters vary freely and set  $m_0 = 1 - m$  in the dispersal distribution (eqn 2), so that the dispersal probability to any other cluster is proportional to  $m$ . We then perform a first-order Taylor expansion of the characteristic function of dispersal around  $m = 0$  and neglect terms of order  $O(m^2)$  in this function, which gives for fecundity effects

$$\text{MRS}_{\mathbf{k},t}^i \simeq \begin{cases} \frac{1}{x} [n_d(2 - \{s + 2(1-s)t\}m) - 2] & \text{if } \mathbf{k} = \mathbf{0} \\ \frac{1}{x} [n_d\{s + (1-s)2t\}m_{\mathbf{k}} - 2] & \text{otherwise,} \end{cases} \quad (21)$$

where

$$x = n_d \left[ \frac{2(N+s)}{1+s} - ms \right] - 2. \quad (22)$$

The proof of eqn 21 is given in Appendix IV. This equation shows that for small migration rates, an increase in survival decreases the value of groups living in the future of the population except the focal group, whose value actually increases with higher survival.

Under small migration rates, we find for survival effects that

$$\text{MRS}_{\mathbf{k},t}^i \simeq \begin{cases} \frac{1}{x} [n_d\{2 - (1-s)(2t-1)m\} - 2] & \text{if } \mathbf{k} = \mathbf{0} \\ \frac{1}{x} [n_d(1-s)(2t-1)m_{\mathbf{k}} - 2] & \text{otherwise,} \end{cases} \quad (23)$$

where

$$y = n_d \left[ \frac{4(N-1)}{1+s} + m(1-s) + 2 \right] - 2. \quad (24)$$

The proof of eqn 23 is also given in Appendix IV. This equation illustrates that, as for fecundity effects, an increase in survival decreases the value of groups living in the future of the population except the focal group. Comparing eqns 21–23 also shows that for  $\mathbf{k} \neq \mathbf{0}$ , the numerator of eqn 21 is larger than that in eqn 23 because  $s+(1-s)2t > (1-s)(2t-1)$ . Hence, everything else being equal, this shows in accordance with the previous results that the present value of a future reward is stronger under fecundity than survival effects.

If we assume that the total number of groups becomes very large ( $n_d \rightarrow \infty$ ) and that  $s = 0$ , then we have from eqn 2 that  $\text{MRS}_{\mathbf{0},t}^i = (1 - tm)/N$  and  $\text{MRS}_{\mathbf{k},t}^i = tmp_{\mathbf{k}}/\{n_1 n_2 \cdots (n_k - 1)N\}$ . This illustrates that the value of a future reward to a group at distance  $\mathbf{k}, t$  from the focal group is discounted according to the total number of individuals in the cluster.

## Discussion

### Summary of the results

In this study, we have derived social discount rates, which are favoured by evolution in hierarchically clustered populations, where individuals can have semelparous or iteroparous life histories. Social discounting involves applying a diminishing weight to group-wide future benefits (Sozou 2009). We have evaluated discount rates for phenotypic effects that can be arbitrarily extended in space and time, from here to infinity. Our analysis, which also applies qualitatively to the more standard models of isolation-by-distance types of population structure, where groups are arranged on a regular lattice (e.g. a circle in one dimension, a torus in two dimensions), has revealed four main features of how evolution is likely to shape social discounting in spatially structured populations.

First, the present value of a future reward to the local group should be discounted by an approximately constantly falling factor under both survival and fecundity effects when dispersal is localized and migration is small [panel (B) of Fig. 1]. This results in social discounting that tends to be exponential, which was shown to be the rule in the island model of dispersal (Sozou 2009). When dispersal becomes stronger but remains localized and survival is not too large, a future reward to the local group should be discounted by a factor falling rapidly for small delays but then slowly for longer delays [panels (C)–(D) of Fig. 1]. This results in social discounting that is hyperbolic (Rogers 1994; Green & Myerson 1996; Sozou 1998). We find that a future reward to individuals from other groups, such as those from the same village, clan or tribe, should generally be discounted with declining rates [panel (A) of Fig. 1 and extensive numerical investigations]. Our analysis thus suggests that evolution will generally not favour exponential social discounting when it occurs in spatially structured populations and may favour hyperbolic discounting.

That evolution may favour a discount rate that is nonconstant as a function of delay is message number one of our results. We expect this to be a general conclusion, not specific to our model, as it stems from the fact that the decay of relatedness per unit delay is unli-

kely to be exponential under localized dispersal. In such situations, relatedness is a sum of at least as many exponential functions as there are distinct dispersal probabilities  $m_i$  to different hierarchical or spatial distances, and even under the stepping-stone model of dispersal on a lattice, a large number of different eigenvalues affect relatedness (Maruyama 1970a, b). This suggests that nonconstant discounting should be a generic feature of many demographic scenarios under localized dispersal, and exponential discounting is likely to be an exception. This result should apply not only to the long-term discount rate  $\theta$ , but also to the short-term discount rate (instantaneous discount rate), which gives the rate of at which a future reward is discounted over a single given time step (Laibson 1995; Sozou 1998; numerical explorations not shown suggest that the instantaneous discount rate is as variable as  $\theta$  in our model).

The second feature that emerges from our analysis is that the present value of a future group-wide reward is higher for all delays when the reward affects the fecundity of individuals living in the future rather than when it affects their survival [compare panel (A) to panel (E) in Fig. 1 and Fig. 2]. This is a consequence of selection being stronger on fecundity than on survival enhancing actions (see section 'Infinite island model of dispersal' and Taylor & Irwin 2000). But the discount rate per unit delay is approximately the same in the two situations [compare panel (B) to panel (F) in Fig. 1 and Fig. 2], which suggests that the change in the valuation of future rewards follows the same trend under both fecundity and survival-changing behaviours. In our examples, we have also assumed that the expression of the social behaviour affects the same vital rate of actors and recipients, but it may be the case that the behaviour affects the survival of actors and the fecundity of recipients (or vice versa). This can be accounted for by using the appropriate transmission coefficients in the substitution rates (e.g. using eqn A.13 for  $\tau_\bullet$  and eqn 8 for  $\tau_{k,t}$  in  $MRS_{k,t}^i$  of eqn 6), and it would not result in a qualitative difference for the time dynamics of the social discount rate as only the effect on recipients is time dependent.

The third feature that emerges from our analysis is that the present value of a future reward tends to be higher for all delays when the survival of individuals from one time period to the next increases [compare panels (A) in Figs 1 and 2 and panels (E) in those two figures]. This follows from the fact that relatedness between actor and recipient is higher when individuals remain throughout their lifespan in the group they have settled. This was already observed in the absence of phenotypic effects extending in time (Taylor & Irwin 2000; Irwin & Taylor 2001), and our analysis shows that

it holds more generally, but different results are expected when adult individuals can also disperse (Lehmann & Rousset 2010).

The fourth feature illustrated by our model is that individuals taken in groups at different hierarchical distance from each other will be related only if the total number of individuals in the cluster encompassing those groups is not too large. The present value of a future reward to the individuals from a given level of clustering then decreases proportionally according to the total number of individuals in that cluster (section 'Small migration'). Hence, evolution will favour incomplete (nonunit) spatial discounting only if clusters encompassing actor and recipients of the behaviour are not too large, which is actually a standard prediction of inclusive fitness theory (Keller 1999; Bourke 2011). When the total population is of finite size, individuals taken at the largest hierarchical distance will be negatively related to the actor and so should be harmed instead of helped or treated indifferently (see eqn 15), thus always inhibiting the evolution of benevolent relations between individuals taken from different clusters of the largest level of clustering, which is again a standard prediction (Hamilton 1970; Hirshleifer 1998).

### *Biological implications*

That spatial discounting increases rapidly with the hierarchical distance between groups is another way of saying that the selective pressure on helping decreases in that case. Hierarchically clustered population may have been the form of social organization of early human populations (Johnson & Earle 1987), and individuals may then have affected the fitness of others from the same village, clan or tribe. Although the selective pressure on social behaviours affecting individuals from other levels of clustering may be small if, for instance, the number of clans within tribes is large, it may still have played a role by tipping the balance in favour of better outcomes in conditions of selection between alternative equilibria (Binmore *et al.* 1995; Binmore 2005), or by initiating the emergence of forms of cooperation between individuals from different clans that initially need some interactions occurring between relatives such as reciprocity (Axelrod & Hamilton 1981), be it of the direct, indirect or generalized form.

Hierarchically clustered populations occur in other mammals as well (e.g. Fontanillas *et al.* 2004; Kerth *et al.* 2011; Wittemyer *et al.* 2005; Schreier & Swedell 2009). It is usually the lowest level of clustering that functions as the unified social unit, while interactions among individuals from higher levels of clustering tend to be more temporary and less cohesive, because altruism between them is less likely to occur (Keller 1999;

Bourke 2011). In Hamadryas baboons, the most cohesive unit is the one-male unit, which consists of one male and several females with their offspring. Several such units may come together at night as the increased number of individuals in the cluster may provide enhanced protection to predators (Wittemyer *et al.* 2005). In elephants, the merging of multiple families into a higher social level structure has been suggested to increase anti-predatory behaviour and better resource/territorial defence (Schreier & Swedell 2009). While the indirect fitness effects stemming from changing the fitness of individuals from second or third level of clustering may be small, the higher levels of clustering may nevertheless be organized along kinship lines because relatedness should not be equal to zero between individuals from different clusters, unless the total number of individuals in a cluster is very large.

But these results on how individuals trade-off between fitness effect to self and to individuals from other groups in the population mainly parallel relatedness distributions observed under isolation-by-distance models (Maruyama 1970a; Malécot 1975; Taylor & Irwin 2000) and expected from previous multilevel kin selection arguments (Keller 1999; Bourke 2011). By contrast, our findings concerning the form of temporal discounting are more relevant. Here, our results suggest that hyperbolic social discounting may be selected for as soon as there is localized dispersal. Hyperbolic discounting causes individuals to be present-biased and even time-inconsistent (e.g. Sozou 1998, Fig. 1). Our results suggest that these characteristics do not only apply to individual discounting, which has been repeatedly observed in animals (e.g. Ainslie 1975; Laibson 1995; Green & Myerson 1996), but may also apply to social discounting, thereby causing individuals to be present-biased with regard to the consequences of their actions on their local or neighbouring groups. Because a constant discount rate is favoured under the island model of dispersal, where individuals do not express any temporal bias (see eqns 16–17 and Sozou 2009), a main prediction of our model is that the more the dispersal is localized, the more present-biased individuals are likely to be. This illustrates that demography and migration patterns are likely to tune the cognitive biases of the individuals in a population (here impulsivity and time-inconsistency).

Whether evolution has really equipped organisms with nonunit social discounting rates depends on the extent to which the extended phenotypic effects (the  $B_{k,t}$ 's) can be extended in time. If extended phenotypic effects do not exist in natural populations to begin with, then there is no need to discount the future. Extended phenotypic effects may, for instance, result from individuals refraining of consuming present resources,

which may increase resource abundance for the actor and its group members at later time points. But the phenotypic effects  $B_{k,t}$  may also extend beyond the actor's lifespan. Plants may change the availability of nutrients in the soil of later generations and parasite may affect the bodily conditions of their host, which may change the fitness of future generations of parasites (e.g. inceptive perturbations, Odling-Smee *et al.* 2003, Table 2.1). In humans, the constructions of settlements, the provision of agricultural field or technological innovations may potentially last beyond the lifespan of the actor and result in multigenerational phenotypic effects.

The extent to which phenotypic effects extend over multiple time periods in natural populations would be interesting to document, as the presence of such effects may strengthen the hypothesis that evolution has equipped individuals with incomplete (nonunit) social discount rates. Our analysis suggests that whether communal constructions will rapidly fall apart or survive for a longer time, possibly beyond the constructor's lifespan, is likely to correlate with the relatedness of individuals sampled at different time points in the same group, which could be measured by using molecular markers. For instance, the higher the relatedness between generations in the focal group, the stronger the selection pressure on investment into behaviours increasing the fitness of individuals living in downstream generations (eqns A.12 and A.14 with  $\mathbf{k} = \mathbf{0}$ ). Our model also predicts that when faced with the choice between an immediate reward to self and a future reward to all group members (like in a public goods game), and everything else being the same, individuals living in populations exhibiting isolation by distance should be more impatient than those living in populations where dispersal is random.

#### *Limitations of the model*

It is worthwhile to point out that our model of transfer of resources between individuals under isolation by distance has at least three main limitations that merit further study. The first is that we have evaluated fitness effects for average phenotypic effects, where individuals express their behaviour regardless of their age. But it may be more realistic to consider that individuals express behaviours conditional on their age, as the population structure will change with their age and so the valuation of rewards given to others will change with age as well. Understanding the evolution of time preference in this context will require the evaluation of context-specific relatedness coefficients (e.g. Johnstone & Cant 2010). The second limitation is that individuals do not senesce (or age) as the survival probability is constant throughout life. Ageing affects individual dis-

counting (Rogers 1994; Sozou & Seymour 2003), and this may affect social discounting as well. The third limitation is that we assumed constant population size. A model allowing for an explicit demography may also be more appropriate for understanding the evolution of time preferences. In an expanding population, which is the case for the human population since its inception, individual discounting of the future may be increased (Rogers 1994) and this may again affect social discounting. Taking all these features into account may allow for a better understanding of how the demographic background of a population affects the evolution of an individual's valuation of its future effects on the common goods of the population.

### Acknowledgements

We thank Sébastien Lion, Nicolas Perrin, Jörgen Weibull and a reviewer for very helpful comments on this paper, which markedly improved the presentation of our model and results.

### References

- Ainslie G (1975) Specious reward: a behavioural theory of impulsiveness and impulse control. *Psychological Bulletin* **82**, 463–496.
- Arrow KJ, Levin SA (2009) Intergenerational resource transfers with random offspring numbers. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 13702–13706.
- Axelrod R, Hamilton WD (1981) The evolution of cooperation. *Science* **211**, 1390–1396.
- Balloux, F, Lugon-Moulin N (2002) The estimation of population differentiation with microsatellite markers. *Molecular Ecology* **11**, 155–165.
- Becker GS, Barro RJ (1988) A reformulation of the economic theory of fertility. *The Quarterly Journal of Economics* **103**, 1–25.
- Binmore K (2005) *Natural Justice*. Oxford University Press, Oxford.
- Binmore K, Samuelson L, Vaughan R (1995) Musical chairs: modeling noisy evolution. *Games and Economic Behavior* **11**, 1–35.
- Bourke A (2011) *Principles of Social Evolution*. Oxford University Press, Oxford.
- Bourke A, Franks N (1995) *Social Evolution in Ants*. Princeton University Press, Princeton, New Jersey.
- Charlesworth B (1980) *Evolution in Age-Structured Populations*. Cambridge University Press, Cambridge, Massachusetts.
- Clobert J, Danchin E, Dhondt AA, Nichols JD (eds) (2001) *Dispersal*. Oxford University Press, Oxford.
- Dasgupta P, Maskin E (2005) Uncertainty and hyperbolic discounting. *The American Economic Review* **95**, 1290–1291.
- Ewens WJ (2004) *Mathematical Population Genetics*. Springer-Verlag, New York.
- Fontanillas P, Petit E, Perrin N (2004) Estimating sex-specific dispersal rates with autosomal markers in hierarchically structured populations. *Evolution* **58**, 886–894.
- Foster D, Young HP (1990) Stochastic evolutionary game dynamics. *Theoretical Population Biology* **38**, 219–232.
- Frank SA (1998) *Foundations of Social Evolution*. Princeton University Press, Princeton, New Jersey.
- Fudenberg D, Imhof LA (2006) Imitation processes with small mutations. *Journal of Economic Theory* **131**, 251–262.
- Gandon S, Rousset F (1999) Evolution of stepping-stone dispersal rates. *Proceedings of the Royal Society of London Series B-Biological Sciences* **221**, 2507–2513.
- Grafen A (1984) Natural selection, kin selection and group selection. In: *Behavioral Ecology: An Evolutionary Approach* (eds Krebs JR, Davies N), pp. 62–84. Blackwell Scientific Publications, Oxford.
- Grafen A, Archetti M (2008) Natural selection of altruism in inelastic viscous homogeneous populations. *Journal of Theoretical Biology* **252**, 694–710.
- Green L, Myerson J (1996) Exponential versus hyperbolic discounting of delayed outcomes: risk and waiting time. *American Zoologist* **36**, 496–505.
- Hamilton WD (1964) The genetical evolution of social behaviour, 1. *Journal of Theoretical Biology* **7**, 1–16.
- Hamilton WD (1970) Selfish and spiteful behavior in an evolutionary model. *Nature* **228**, 1218–1220.
- Hamilton WD (1971) Selection of selfish and altruistic behaviour in some extreme models. In: *Man and Beast: Comparative Social Behavior* (eds Eisenberg J, Dillon W), pp. 59–91. Smithsonian Institution Press, Washington, District of Columbia.
- Hansson I, Stuart C (1990) Malthusian selection of preferences. *The American Economic Review* **80**, 529–544.
- Hardin G (1968) The tragedy of the commons. *Science* **162**, 1243–1248.
- Henly SE, Ostdiek A, Blackwell E, Knutie S, Dunlap AS, Stephens DW (2007) The discounting-by-interruption hypothesis: model and experiments. *Behavioral Ecology* **19**, 154–162.
- Hirshleifer J (1998) The bioeconomic causes of war. *Managerial and Decision Economics* **19**, 457–66.
- Hirshleifer J, Glazer A, Hirshleifer D (2006) *Price Theory and Applications*, 7th edn. Cambridge University Press, Cambridge, Massachusetts.
- Holzer B, Keller L, and Chapuisat M (2009) Genetic clusters and sex-biased gene flow in a unicolonial Formica ant. *BMC Evolutionary Biology* **9**, 1–11.
- Irwin AJ, Taylor PD (2001) Evolution of altruism in stepping-stone populations with overlapping generations. *Theoretical Population Biology* **60**, 315–325.
- Johnson AW, Earle T (1987) *The Evolution of Human Societies*. Stanford University Press, Stanford, California.
- Johnstone RA, Cant MA (2008) Sex differences in dispersal and the evolution of helping and harming. *American Naturalist* **172**, 318–330.
- Johnstone RA, Cant MA (2010) The evolution of menopause in cetaceans and humans: the role of demography. *Proceedings of the Royal Society B-Biological Sciences* **277**, 3765–3771.
- Kagel J, Green L, Caraco T (1986) When foragers discount the future: constraint or adaptation? *Animal Behaviour* **34**, 271–283.
- Keller L (1999) *Levels of Selection in Evolution*. Princeton University Press, Princeton, New Jersey.
- Kerth G, Perony N, Sweitzer F (2011) Bats are able to maintain long-term social relationships despite the high fission–fusion

- dynamics of their groups. *Proceedings of the Royal Society of London Series B-Biological Sciences*. In press.
- Kimura M, Crow JF (1964) The number of alleles that can be maintained in a finite population. *Genetics* **49**, 725–738.
- Laiabson D (1995) Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics* **112**, 443–447.
- Lehmann L (2007) The evolution of trans-generational altruism: kin selection meets niche construction. *Journal of Evolutionary Biology* **20**, 181–189.
- Lehmann L (2008) The adaptive dynamics of niche constructing traits in spatially subdivided populations: evolving posthumous extended phenotypes. *Evolution* **62**, 549–566.
- Lehmann L (2010) Space-time relatedness and Hamilton's rule for long-lasting behaviors in viscous populations. *American Naturalist* **175**, 136–143.
- Lehmann L, Rousset F (2009) Perturbation expansions of multilocus fixation probabilities for frequency-dependent selection with applications to the Hill-Robertson effect and to the joint evolution of helping and punishment. *Theoretical Population Biology* **76**, 35–51.
- Lehmann L, Rousset F (2010) How life-history and demography promote or inhibit the evolution of helping behaviors. *Philosophical Transactions of the Royal Society B* **365**, 2599–2617.
- Lehmann L, Keller L, Sumpter D (2007) The evolution of helping and harming on graphs: the return of the inclusive fitness effect. *Journal of Evolutionary Biology* **20**, 2284–2295.
- León JA (1976) Life histories as adaptive strategies. *Journal of Theoretical Biology* **60**, 301–335.
- Lessard S, Ladret V (2007) The probability of fixation of a single mutant in an exchangeable selection model. *Journal of Mathematical Biology* **54**, 721–744.
- Malécot G (1973) Génétique des populations diploïdes naturelle dans le cas d'un seul locus. III-Parenté, mutations et migration. *Annale de Génétique et de Sélection Animale* **5**, 333–361.
- Malécot G (1975) Heterozygosity and relationship in regularly subdivided populations. *Theoretical Population Biology* **8**, 212–241.
- Maruyama T (1970a) Effective number of alleles in a subdivided population. *Theoretical Population Biology* **1**, 273–306.
- Maruyama T (1970b) Stepping stone models of finite length. *Advanced Applied Probability* **2**, 229–258.
- Maynard-Smith J (1982) *Evolution and the Theory of Games*. Cambridge University Press, Cambridge, Massachusetts.
- Maynard Smith J, Szathmari E (1995) *The Major Transitions in Evolution*. Oxford University Press, Oxford.
- Nagylaki T (1982) Geographical invariance in population genetics. *Journal of Theoretical Biology* **99**, 159–172.
- Odling-Smee FJ, Laland KL, Feldman MW (2003) *Niche Construction*. Princeton University Press, Princeton, New Jersey.
- Ohtsuki H, Nowak MA (2006) Evolutionary games on cycles. *Proceedings of the Royal Society of London Series B-Biological Sciences* **273**, 2249–2256.
- Queller DC (1994) Genetic relatedness in viscous populations. *Evolutionary Ecology* **8**, 70–73.
- Rogers AR (1990) Group selection by selective emigration: the effects of migration and kin structure. *American Naturalist* **135**, 398–413.
- Rogers AR (1994) Evolution of time preference by natural selection. *The American Economic Review* **84**, 460–481.
- 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, New Jersey.
- Rousset F, Billiard S (2000) A theoretical basis for measures of kin selection in subdivided populations: finite populations and localized dispersal. *Journal of Evolutionary Biology* **13**, 814–825.
- Sawyer S (1976) Results for the stepping stone model for migration in population genetics. *Annals of Probability* **4**, 699–728.
- Sawyer S, Felsenstein J (1983) Isolation by distance in a hierarchically clustered population. *Journal of Applied Probability* **20**, 1–10.
- Schreier AL, Swedell L (2009) The fourth level of social structure in a multi-level society: ecological and social functions of clans in hamadryas baboons. *American Journal of Primatology* **71**, 948–955.
- Sozou PD (1998) On hyperbolic discounting and uncertain hazard rates. *Proceedings of the Royal Society of London Series B-Biological Sciences* **265**, 2015–2020.
- Sozou PD (2009) Individual and social discounting in a viscous population. *Proceedings of the Royal Society of London Series B-Biological Sciences* **276**, 2955–2962.
- Sozou PD, Seymour RM (2003) Augmented discounting: interaction between ageing and time-preference behaviour. *Proceedings of the Royal Society B-Biological Sciences* **270**, 1047–1053.
- Stearns S (1992) *The Evolution of Life Histories*. Oxford University Press, Oxford.
- Stephens DW (2002) Discrimination, discounting and impulsivity: a role for an informational constraint. *Proceedings of the Royal Society B-Biological Sciences* **357**, 1527–1537.
- Taylor PD (1992) Inclusive fitness in a homogeneous environment. *Proceedings of the Royal Society of London Series B-Biological Sciences* **240**, 299–302.
- Taylor PD, Frank SA (1996) How to make a kin selection model. *Journal of Theoretical Biology* **180**, 27–37.
- Taylor PD, Irwin AJ (2000) Overlapping generations can promote altruistic behavior. *Evolution* **54**, 1135–1141.
- Taylor PD, Day T, Wild G (2007) Evolution of cooperation in a finite homogeneous graph. *Nature* **447**, 469–472.
- Tobin H, Logue A (1994) Self-control across species (*Columba livia*, *Homo sapiens*, and *Rattus norvegicus*). *Journal of Comparative Psychology* **108**, 126–133.
- Trivers RL (1974) Parent-offspring conflict. *American Zoologist* **14**, 249–264.
- West SA, Griffin AS, Gardner A (2007) Evolutionary explanations for cooperation. *Current Biology* **17**, 661–672.
- Wittemyer G, Douglas-Hamilton I, Getz WM (2005) The socioecology of elephants: analysis of the processes creating multitiered social structures. *Animal Behaviour* **69**, 1357–1371.
- Wright S (1931) Evolution in Mendelian populations. *Genetics* **16**, 97–159.

---

L.L. is a researcher at the University of Lausanne (Switzerland). His research focuses on understanding the selective pressure on social behaviors such as altruism and spite and



the ontogeny of these behaviors during an individual's lifespan. F.R. is a CNRS researcher at the University of Montpellier II. His research investigates Emile Cioran's theory that spite is the foundation of social life.

## Appendix I: method

### *Inclusive effects on fixation probabilities*

In this appendix, we describe how to obtain eqn 6 of the main text by focusing on the fixation probability of the mutant allele. To that aim, we first note that the average effect of a class of actors on the fitness of individuals bearing the mutant allele depends on the probability that both the actor and recipient bear the mutant allele, hence that their gene lineages have a common ancestor in a single ancestral individual carrying the mutant allele. Under weak selection, the cumulative effects on allele frequency change of actor–recipient interactions over generations until the loss or fixation of a single initial mutant allele can be expressed in terms of the fitness differential induced by the interaction weighted by the average coalescence time (i.e. the average time to the first common ancestor) of pair of genes taken in actor and recipients. The overall effect of selection can then be expressed in terms of fitness differential weighted ratios of average coalescence times for different pairs of genes (Rousset 2003; for more rigorous arguments see also Lessard & Ladret 2007; Lehmann & Rousset 2009). Actor–recipient interactions do not need to be limited between individuals living in the same or adjacent generations but may involve a multigenerational gap between the expression of the behaviour by actors and their phenotypic effect on the fitness of recipients, in which case, even for an arbitrary number of time steps between the generation of the actor and that of the recipient, selection can still be expressed in terms of average coalescence times (Lehmann 2007).

The coalescence times are purely genealogical properties, not depending on any mutation process, but can be recovered as low-mutation limit values of measures of the genetic structure of populations in models with genetic mutation (Rousset & Billiard 2000; Rousset 2004). For effective computations in the hierarchically clustered population model presented in the main text, the perturbation of the fixation probability of a single mutant allele is conveniently written as

$$\phi = \lim_{\mu \rightarrow 0} \frac{1}{1 - Q_{0,0}} \left[ \frac{\partial w}{\partial z_{\bullet}} + \sum_{t=0}^{\infty} \sum_{\mathbf{k}} \frac{\partial w}{\partial z_{\mathbf{k},t}} Q_{\mathbf{k},t} \right], \quad (\text{A.1})$$

(Lehmann 2007, eqn A.10)<sup>2</sup>, where  $\mu$  is the mutation rate from one allele to another. The function  $w$  gives the fitness of a focal individual, which is defined as the expected number of its descendants (possibly including itself) after one full iteration of the life cycle of the organism (Hamilton 1964), and it depends on the focal individual's phenotype  $z_{\bullet}$  and potentially on the phenotype of every individual living in the population in present and past times, where  $z_{\mathbf{k},t}$  is the average phenotype of individuals living in group  $\mathbf{k}$  at  $t$  time periods in the past. The partial derivatives are the fitness differentials with  $\partial w / \partial z_{\bullet}$  being the effect on its fitness of the focal individual expressing its phenotype and  $\partial w / \partial z_{\mathbf{k},t}$  is the effect of all neighbours separated by distance  $\mathbf{k},t$  from the focal group. Finally,  $Q_{\mathbf{k},t}$  is the stationary probability that a gene sampled in the focal individual is identical with a homologous gene sampled in an individual chosen at random without replacement from group  $\mathbf{k}$  at  $t$  time periods in the past. Under the present model, the phenotypes can be thought of as the average frequency of the mutant allele among the classes of actors under consideration and are evaluated at  $z_{\bullet} = \dots = z_{\mathbf{k},t} = 0$ ; that is, in a population monomorphic for the resident allele that does not express the  $C$  and  $B_{j,t}$  perturbations, but more general interpretations in terms of evolving traits are possible (Rousset & Billiard 2000; Lehmann 2008).

### *Direct fitness function*

The direct fitness function (Taylor & Frank 1996; Frank 1998) for the hierarchically clustered population model can be written as

$$w = s + (1 - s) \sum_{\mathbf{i}} m_{\mathbf{i}} \frac{1 + v_{\bullet}}{\sum_{\mathbf{j}} m_{\mathbf{i}-\mathbf{j}} (1 + v_{\mathbf{j}})}, \quad (\text{A.2})$$

where  $1 + v_{\bullet}$  is the fecundity of a focal individual (strictly speaking the fecundity of an individual relative

<sup>2</sup> Selection favours the mutant if  $\phi > 0$ , which, owing to the property that the effects on the focal individual's fitness of all individuals in a generation in the population must sum up to zero (Rousset & Billiard 2000), can also be recast under the form  $\sum_{\mathbf{i}} \sum_{\mathbf{k}} R_{\mathbf{k},t} b_{\mathbf{k},t} - c > 0$  (Lehmann 2010). This is Hamilton's rule, where  $-c \equiv \partial w / \partial z_{\bullet}$ ,  $b_{\mathbf{k},t} \equiv \partial w / \partial z_{\mathbf{k},t}$ , and  $R_{\mathbf{k},t} = \lim_{\mu \rightarrow 0} (Q_{\mathbf{k},t} - \bar{Q}_t) / (1 - \bar{Q}_t)$ , with  $\bar{Q}_t = \sum_{\mathbf{k}} Q_{\mathbf{k},t} / n_d$  being the average probability of identity between two homologous alleles sampled in two distinct individuals living at  $t$  generations apart. This relatedness coefficient  $R_{\mathbf{k},t}$  provides a measure of the extent to which an individual sampled in group  $\mathbf{k},t$  is more (or less) likely to transmit a mutant allele to the next generation than is a randomly sampled individual from  $t$ , relative to the extent to which the focal individual is more likely to transmit the mutant allele to the next generation than is another individual sampled in its generation. But owing to the zero sum property of fitness effects, one can also choose to write relatedness as  $R_{\mathbf{k},t} = \lim_{\mu \rightarrow 0} (Q_{\mathbf{k},t} - Q_{0,0}) / (1 - Q_{0,0})$  so that relatedness coefficients are defined relative to the probability of identity  $Q_{0,0}$  in the focal deme, which facilitates empirical estimations (see Lehmann & Rousset 2010, section 'Results' for a discussion)

to that of a resident in a monomorphic population), and  $v_\bullet$  is the increase in the fecundity of the focal individual stemming from individuals in the population carrying the mutant allele. The fitness function  $w$  can be understood by noting that a fraction  $m_i(1+v_\bullet)$  of the focal individual's offspring enter in competition in group  $i$  with a fraction  $\sum_j m_{i-j}(1+v_j)$  of the total number of offspring produced in the population, where  $v_j$  is the increase in average relative fecundity of individuals in group  $j$ . Equation A.2 has exactly the same form as that found under the more standard isolation-by-distance models of population structure (e.g. Gandon & Rousset 1999, eqn 1, Lehmann 2008, eqn 5).

This formulation of the fitness function will allow us to evaluate the two substitution rates defined by eqns 3–4, but we will write the fecundities on a case-by-case basis. In order to capture the case where the focal individual trades off its own fecundity with that of individuals from group  $\mathbf{k},t$  (eqn 3), we need to evaluate the evolutionary dynamics of a mutant allele having phenotypic effect  $-C$  on the actor and  $B_{\mathbf{k},t}$  on group  $\mathbf{k},t$  members. We then write

$$v_\bullet = -Cz_\bullet + B_{\mathbf{k},t}z_{\mathbf{k},t}^R, \quad (\text{A.3})$$

where  $C > 0$ ,  $B_{\mathbf{k},t} > 0$ , and  $z_{\mathbf{k},t}^R = z_{\mathbf{k},t}$  for all  $\mathbf{k}$  and  $t$  except that  $z_{0,0}^R = z_\bullet/N + z_{0,0}(N-1)/N$  because the focal individual is in proportion  $1/N$  in the focal group. For the case where the focal individual trades-off the average fecundity of the focal group in the focal generation with that of individuals from group  $\mathbf{k},t$ , we need to evaluate the evolutionary dynamics of a mutant having phenotypic effect  $B_{0,0}$  on the focal group and effect  $B_{\mathbf{k},t}$  on group  $\mathbf{k},t$  with either  $\mathbf{k} > 0$  or  $t > 0$ , which gives

$$v_\bullet = B_{0,0}z_{0,t}^R + B_{\mathbf{k},t}z_{\mathbf{k},t}^R. \quad (\text{A.4})$$

Although both eqns A.3 and A.4 describe situations where the mutant alleles have effects on two classes of recipients, one may more generally be interested in assessing selection on alleles having an effect on an arbitrary number of recipients in which case one can set  $v_\bullet = -Cz_\bullet + \sum_{t=0}^{\infty} \sum_{\mathbf{k}} B_{\mathbf{k},t}z_{\mathbf{k},t}^R$ . The expressions for  $v_j$  corresponding to eqns A.3 and A.4 are respectively

$$v_j = -Cz_{j,0}^R + B_{\mathbf{k},t}z_{j-\mathbf{k},t}^R, \quad (\text{A.5})$$

and

$$v_j = B_{0,0}z_{j,0}^R + B_{\mathbf{k},t}z_{j-\mathbf{k},t}^R, \quad (\text{A.6})$$

which can be changed to  $v_j = -Cz_{j,0}^R + \sum_{t=0}^{\infty} \sum_{\mathbf{k}} B_{\mathbf{k},t}z_{j-\mathbf{k},t}^R$  if actors have an effect on an arbitrary

number of recipients living at different positions in space and time.

For survival effects, the direct fitness function can be written as

$$w = s(1+v_\bullet) + \sum_i m_i[1-s(1+v_i)], \quad (\text{A.7})$$

which can be understood by noting that a focal individual survives with probability  $s(1+v_\bullet)$  to the next time period and that a fraction  $m_i$  of its offspring (no effect on fecundity) enters in competition in group  $i$ , where there is an average fraction  $1-s(1+v_i)$  of vacated breeding spots. Note that at  $z_\bullet = \dots = z_{\mathbf{k},t} = \dots = 0$ , we have  $w = 1$  in both eqns A.2 and A.7.

Substituting eqns A.3 and A.5 into eqn A.2 (or into eqn A.7), which is then inserted into eqn A.1, and evaluating the partial derivatives show that  $\phi = 0$  can be written as

$$B_{\mathbf{k},t}\tau_{\mathbf{k},t} - C\tau_\bullet = 0, \quad (\text{A.8})$$

and substituting eqns A.4 and A.6 into eqn A.2 (or into eqn A.7) and evaluating  $\phi = 0$  show that it can be written as

$$B_{\mathbf{k},t}\tau_{\mathbf{k},t} + B_{0,0}\tau_{0,0} = 0. \quad (\text{A.9})$$

The coefficients  $\tau_\bullet$  and  $\tau_{\mathbf{k},t}$  in these equations will depend on the demographic parameters  $N$ ,  $s$ ,  $m_j$  and on whether phenotypic effects are fecundity or survival enhancing, but not on  $C_{\mathbf{k},t}$  and  $B_{\mathbf{k},t}$ . Substituting eqns A.8 and A.9 into, respectively, eqns 3–4 of the main text yields eqn 6.

### Fecundity effects

The coefficient  $\tau_\bullet$  in eqn A.8 for fecundity effects can be directly obtained by substituting eqns A.2, A.3, and A.5 into eqn A.1 and then evaluating  $\tau_\bullet = \phi/C$  by setting  $B_{0,0} = 0$  and  $B_{\mathbf{k},0} = 0$ . This gives

$$\tau_\bullet = -\lim_{\mu \rightarrow 0} \frac{(1-s)}{1-Q_{0,0}} \left[ 1 - \sum_i \sum_j m_i m_{i-j} Q_{j,0}^R \right], \quad (\text{eqn A.10})$$

with

$$Q_{j,t}^R = Q_{j,t} \text{ for all } j \text{ and } t \text{ except } Q_{0,0}^R = \frac{1}{N} + \left( \frac{N-1}{N} \right) Q_{0,0}, \quad (\text{A.11})$$

where  $Q_{j,t}^R$  is the probability that two individuals randomly sampled at 'distance  $j,t$ ' carry a homologous gene identical by descent so that sampling with replacement occurs only if the two individuals are sampled from the same group.

The coefficients  $\tau_{k,t}$  in eqns A.8–A.9 for fecundity effects are obtained by substituting eqns A.2, A.3 and A.5 into eqn A.1 and then evaluating  $\tau_{k,t} = \phi/B_{k,t}$  by setting  $C = 0$  and  $B_{j,h} = 0$  except for  $j = k$  and  $h = t$ . This gives

$$\tau_{k,t} = \lim_{\mu \rightarrow 0} \frac{(1-s)}{1-Q_{0,0}} \left[ Q_{k,t}^R - \sum_i \sum_j m_i m_{i-j} Q_{j-k,t}^R \right], \quad (A.12)$$

where  $Q_{j-k,t}^R$  is the probability of identity by descent between two homologous genes sampled in two individuals living at  $k$  steps apart in the population and  $t$  time periods of interval.

### Survival effects

For survival effects, the coefficient  $\tau_{\bullet}$  in eqn A.8 is obtained by substituting eqns A.3, A.5, and A.7 into eqn A.1 and then proceeding as to obtain eqn A.10. This gives

$$\tau_{\bullet} = - \lim_{\mu \rightarrow 0} \frac{s}{1-Q_{0,0}} \left[ 1 - \sum_j m_j Q_{j,0}^R \right], \quad (A.13)$$

where the first term in square brackets, ‘1’, can be interpreted as the direct fitness cost to a focal individual from expressing a mutant allele that reduces its survival by a unit amount, while the second term as the indirect fitness benefit stemming from the decrease in competition faced by relatives of the focal individual when it decreases its survival. This reduction in competition depends on the probability  $m_j$  that an offspring of the focal individual competes for vacant spots in group  $j$ , which are increased by an amount  $sQ_{j,0}^R$ .

The coefficient  $\tau_{k,t}$  for survival effects is given by

$$\tau_{k,t} = \lim_{\mu \rightarrow 0} \frac{s}{1-Q_{0,0}} \left[ Q_{k,t}^R - \sum_j m_j Q_{j-k,t}^R \right]. \quad (A.14)$$

## Appendix II: IBD probabilities

In this appendix, we evaluate the explicit expressions for the stationary IBD probabilities necessary to evaluate the coefficients  $\tau_{\bullet}$  and  $\tau_{k,t}$  under the life cycle described in the main text. In doing so, we use standard population genetic methods for homogeneous populations (e.g. Malécot 1975; Nagylaki 1982; Rousset 2004) and use classical results on Fourier analysis. All of these results used in this appendix and the methods used to construct recurrence equations for the  $Q_{k,t}$  probabilities are detailed in Rousset (2004), and the appendix summarizes and extends results derived previously (Malécot 1973; Lehmann *et al.* 2007; Grafen & Archetti 2008; Lehmann 2008).

### Recursions for spatial IBD probabilities

Following the life cycle assumptions described in the main text, the probability of identity between a pair of homologous genes sampled in two different adult individuals living at distance  $k$  of each other in the same time period satisfies at steady state the recursion

$$Q_{k,0} = s^2 Q_{k,0} + 2s(1-s)\sqrt{\gamma} \sum_i m_i Q_{k-i,0}^R + (1-s)^2 \gamma \sum_i \sum_j m_i m_{i-j} Q_{k-j,0}^R, \quad (A.15)$$

where  $\gamma = (1-\mu)^2$ . This equation can be understood by noting that with probability  $s^2$ , two surviving (adult) individuals are sampled, in which case the two individuals carry homologous genes that are identical with probability  $Q_{k,0}$ ; with probability  $2s(1-s)$ , a newborn and an adult are sampled, in which case the two individuals carry homologous genes that are identical with probability  $\sqrt{\gamma} Q_{k-i,0}^R$  when the newborn has migrated  $i$  steps. Finally, with probability  $(1-s)^2$ , two newborns are sampled, in which case with probability  $m_i m_{i-j}$  they carry homologous genes that are identical with probability  $\gamma Q_{k-j,0}^R$  (Lehmann *et al.* 2007, eqn 27).

Note that different assumptions about the occurrence of mutation events in the life cycle may alter the nature of the mutation factors in eqn A.15, but would not affect our qualitative conclusions (differences in fixation probabilities would be only of order of the inverse to the total population size). Further, in order to match the assumptions about mutation leading to the derivation of eqn A.1, all components of the right member of eqn A.15 should be multiplied by  $\gamma$ , rather than by different powers of  $\gamma$ , but eqn A.15 is biologically more realistic and using this form, rather than that with the same powers in  $\gamma$ , does not affect the qualitative and quantitative results reported below. For these reasons, we use eqn A.15 throughout this Appendix and inserting eqn A.11 under the form  $Q_{0,0}^R = Q_{0,0} + (1-Q_{0,0})/N$  allows us to write eqn A.15 as

$$Q_{k,0} = s^2 Q_{k,0} + 2s(1-s)\sqrt{\gamma} \left[ \sum_i m_i Q_{k-i,0} + m_k \frac{(1-Q_{0,0})}{N} \right] + (1-s)^2 \gamma \left[ \sum_i \sum_j m_i m_{i-j} Q_{k-j,0} + \sum_i m_i m_{i-k} \frac{(1-Q_{0,0})}{N} \right]. \quad (A.16)$$

We write

$$\psi(\theta) \equiv \sum_{k \in S} m_k e^{k \cdot \theta} \quad (A.17)$$

the Fourier transform of the dispersal distribution (characteristic function), where  $\iota \equiv \sqrt{-1}$ ,  $S = \{(k_1, k_2, k_3, \dots, k_H) : 0 \leq k_i < n_i\}$ . We write  $\psi_{\mathbf{h}}$  its value in  $\boldsymbol{\theta}(\mathbf{h}) \equiv 2\pi(h_1/n_1, h_2/n_2, \dots, h_H/n_H)$  (Sawyer & Felsenstein 1983, eqn 2.6). Call  $\varrho_0(\mathbf{h}) \equiv \sum_{\mathbf{k} \in S} Q_{\mathbf{k},0} e^{i\mathbf{k} \cdot \boldsymbol{\theta}(\mathbf{h})}$  the Fourier transform of the IBD probabilities between pairs of genes sampled in the same time period. Multiplying both sides of eqn A.16 by  $e^{i\mathbf{k} \cdot \boldsymbol{\theta}(\mathbf{h})}$ , which can be written as  $e^{i\boldsymbol{\theta}(\mathbf{h})} e^{i(\mathbf{j}-\mathbf{i}) \cdot \boldsymbol{\theta}(\mathbf{h})} e^{i(\mathbf{k}-\mathbf{j}) \cdot \boldsymbol{\theta}(\mathbf{h})}$ , and summing over the space of all groups produces the recursion

$$\begin{aligned} \varrho_0(\mathbf{h}) &= s^2 \varrho_0(\mathbf{h}) + 2s(1-s)\sqrt{\gamma}\psi_{\mathbf{h}} \left( \varrho_0(\mathbf{h}) + \frac{1-Q_{0,0}}{N} \right) \\ &\quad + (1-s)^2 \gamma \psi_{\mathbf{h}}^2 \left( \varrho_0(\mathbf{h}) + \frac{1-Q_{0,0}}{N} \right), \end{aligned} \quad (\text{A.18})$$

which can be solved for  $\varrho_0(\mathbf{h})$  to give

$$\varrho_0(\mathbf{h}) = \frac{(1-Q_{0,0})}{N} X_{\mathbf{h}}, \quad (\text{A.19})$$

where

$$X_{\mathbf{h}} = \frac{2s\sqrt{\gamma}\psi_{\mathbf{h}} + (1-s)\gamma\psi_{\mathbf{h}}^2}{1+s - [2s\sqrt{\gamma}\psi_{\mathbf{h}} + (1-s)\gamma\psi_{\mathbf{h}}^2]}. \quad (\text{A.20})$$

### Recursions for space-time IBD probabilities

The probability of IBD between a gene sampled in a focal individual residing in a focal group (situated at position  $\mathbf{0}$ ) and a homologous gene sampled from an individual chosen at random in group  $\mathbf{k}$  at  $t$  time periods earlier satisfies for  $t \geq 2$  the recursion

$$Q_{\mathbf{k},t} = sQ_{\mathbf{k},t-1} + (1-s)\sqrt{\gamma} \sum_{\mathbf{i}} m_{\mathbf{i}} Q_{\mathbf{k}-\mathbf{i},t-1}. \quad (\text{A.21})$$

The right member can be understood by noting that with probability  $s$ , the focal individual was already alive one time period earlier, in which case the IBD probability between the focal individual's gene and the gene sampled  $t$  time periods earlier is  $Q_{\mathbf{k},t-1}$ . With probability  $1-s$  the focal individual is a newborn, in which case it has migrated from group  $\mathbf{i}$  with probability  $m_{\mathbf{i}}$  and the IBD probability between the focal individual's gene and the gene sampled  $t$  time periods earlier is  $Q_{\mathbf{k}-\mathbf{i},t-1}$ .

For  $t = 1$  and  $\mathbf{k} \neq \mathbf{0}$ , one has

$$Q_{\mathbf{k},1} = sQ_{\mathbf{k},0} + (1-s)\sqrt{\gamma} \left( \sum_{\mathbf{i}} m_{\mathbf{i}} Q_{\mathbf{k}-\mathbf{i},0} + m_{\mathbf{k}} \frac{(1-Q_{0,0})}{N} \right), \quad (\text{A.22})$$

where the second term accounts for the fact that the individual sampled at random one time period earlier

might be the parent of the focal individual. Finally, for  $t = 1$  and  $\mathbf{k} = \mathbf{0}$

$$\begin{aligned} Q_{0,1} &= s \left( Q_{0,0} + \frac{1-Q_{0,0}}{N} \right) \\ &\quad + (1-s)\sqrt{\gamma} \left( \sum_{\mathbf{i}} m_{\mathbf{i}} Q_{\mathbf{k}-\mathbf{i},0} + m_{\mathbf{k}} \frac{(1-Q_{0,0})}{N} \right), \end{aligned} \quad (\text{A.23})$$

where the first term accounts for the fact that the individual sampled at random one time period earlier might be the focal individual itself. Combining the last two equations produces for  $t = 1$  and all  $\mathbf{k}$

$$\begin{aligned} Q_{\mathbf{k},1} &= s \left( Q_{\mathbf{k},0} + \delta_{\mathbf{k}\mathbf{0}} \frac{1-Q_{0,0}}{N} \right) \\ &\quad + (1-s)\sqrt{\gamma} \left( \sum_{\mathbf{i}} m_{\mathbf{i}} Q_{\mathbf{k}-\mathbf{i},0} + m_{\mathbf{k}} \frac{(1-Q_{0,0})}{N} \right), \end{aligned} \quad (\text{A.24})$$

where  $\delta_{\mathbf{k}\mathbf{0}}$  is the Kronecker-Delta.

Call  $\varrho_t(\mathbf{h}) \equiv \sum_{\mathbf{k} \in S} Q_{\mathbf{k},t} e^{i\mathbf{k} \cdot \boldsymbol{\theta}(\mathbf{h})}$  the characteristic function of the probabilities of identity between pairs of genes sampled at  $t$  time periods of interval. Multiplying both sides of eqn A.21 by  $e^{i\mathbf{k} \cdot \boldsymbol{\theta}(\mathbf{h})}$  and summing over the space of all groups gives the recursion

$$\varrho_t(\mathbf{h}) = (s + (1-s)\sqrt{\gamma}\psi_{\mathbf{h}}) \varrho_{t-1}(\mathbf{h}), \quad (\text{A.25})$$

whose solution is  $\varrho_t(\mathbf{h}) = (s + (1-s)\sqrt{\gamma}\psi_{\mathbf{h}})^{t-1} \varrho_1(\mathbf{h})$ . The  $\varrho_1(\mathbf{h})$  term is obtained by multiplying both sides of eqn A.24 by  $e^{i\mathbf{k} \cdot \boldsymbol{\theta}(\mathbf{h})}$ , summing over the space of all groups, noting that  $\sum_{\mathbf{k}} \delta_{\mathbf{k}\mathbf{0}} e^{i\mathbf{k} \cdot \boldsymbol{\theta}(\mathbf{h})} = 1$  if  $\mathbf{k} = \mathbf{0}$ , zero otherwise, which gives

$$\varrho_1(\mathbf{h}) = (s + (1-s)\sqrt{\gamma}\psi_{\mathbf{h}}) \left( \varrho_0(\mathbf{h}) + \frac{1-Q_{0,0}}{N} \right), \quad (\text{A.26})$$

whereby

$$\varrho_t(\mathbf{h}) = (s + (1-s)\sqrt{\gamma}\psi_{\mathbf{h}})^t \left( \varrho_0(\mathbf{h}) + \frac{1-Q_{0,0}}{N} \right). \quad (\text{A.27})$$

Substituting eqn A.19 into eqn A.27 yields

$$\varrho_t(\mathbf{h}) = \frac{(1-Q_{0,0})}{N} Y_{\mathbf{h},t}, \quad (\text{A.28})$$

where

$$Y_{\mathbf{h},t} = \frac{(1+s)(s + (1-s)\sqrt{\gamma}\psi_{\mathbf{h}})^t}{1+s - [2s\sqrt{\gamma}\psi_{\mathbf{h}} + (1-s)\gamma\psi_{\mathbf{h}}^2]}. \quad (\text{A.29})$$

### Stationary IBD probabilities

From eqns A.19 and A.28, we finally have that the stationary IBD probability  $Q_{\mathbf{k},t}$  can be obtained as

$$Q_{\mathbf{k},t} = \frac{(1 - Q_{0,0})}{N} \mathcal{L}_{\mathbf{k}}(D_{\mathbf{h},t}), \quad (\text{A.30})$$

where

$$\mathcal{L}_{\mathbf{k}}(D_{\mathbf{h},t}) \equiv \frac{1}{n_d} \sum_{\mathbf{h} \in S} D_{\mathbf{h},t} e^{-i\mathbf{k} \cdot \theta(\mathbf{h})} \quad (\text{A.31})$$

is the inverse Fourier transform at distance  $\mathbf{k}$  of the function  $D_{\mathbf{h},t}$ , which is given by eqn A.20 for  $t = 0$ , eqn A.29 otherwise.

### Appendix III: explicit selective effects

#### Fecundity effects

*Transmission coefficients.* Substituting eqn A.11 into eqns A.10 and A.12 gives

$$\tau_{\bullet} = -\lim_{\mu \rightarrow 0} \frac{(1-s)}{1-Q_{0,0}} \left[ 1 - \sum_{\mathbf{i}} \sum_{\mathbf{j}} m_{\mathbf{i}} m_{\mathbf{i}-\mathbf{j}} Q_{\mathbf{j},0} - \frac{(1-Q_{0,0})}{N} \sum_{\mathbf{i}} m_{\mathbf{i}}^2 \right]$$

$$\tau_{0,0} = \lim_{\mu \rightarrow 0} \frac{(1-s)}{1-Q_{0,0}} \left[ \frac{(1-Q_{0,0})}{N} + Q_{0,0} - \sum_{\mathbf{i}} \sum_{\mathbf{j}} m_{\mathbf{i}} m_{\mathbf{i}-\mathbf{j}} Q_{\mathbf{j},0} - \frac{(1-Q_{0,0})}{N} \sum_{\mathbf{i}} m_{\mathbf{i}}^2 \right]$$

$$\tau_{\mathbf{k},0} = \lim_{\mu \rightarrow 0} \frac{(1-s)}{1-Q_{0,0}} \left[ Q_{\mathbf{k},0} - \sum_{\mathbf{i}} \sum_{\mathbf{j}} m_{\mathbf{i}} m_{\mathbf{i}-\mathbf{j}} Q_{\mathbf{j}-\mathbf{k},0} - \frac{(1-Q_{0,0})}{N} \sum_{\mathbf{i}} m_{\mathbf{i}} m_{\mathbf{i}-\mathbf{k}} \right] \text{ if } \mathbf{k} > \mathbf{0}$$

$$\tau_{\mathbf{k},t} = \lim_{\mu \rightarrow 0} \frac{(1-s)}{1-Q_{0,0}} \left[ Q_{\mathbf{k},t} - \sum_{\mathbf{i}} \sum_{\mathbf{j}} m_{\mathbf{i}} m_{\mathbf{i}-\mathbf{j}} Q_{\mathbf{j}-\mathbf{k},t} \right] \text{ if } \mathbf{k} > \mathbf{0}, t > 0. \quad (\text{A.32})$$

Using eqn A.31, we will simplify the above equations by using

$$\begin{aligned} Q_{\mathbf{k},t} - \sum_{\mathbf{i}} \sum_{\mathbf{j}} m_{\mathbf{i}} m_{\mathbf{i}-\mathbf{j}} Q_{\mathbf{j}-\mathbf{k},t} &= \frac{(1-Q_{0,0})}{N n_d} \sum_{\mathbf{h}} D_{\mathbf{h},t} \left[ e^{-i\mathbf{k} \cdot \theta(\mathbf{h})} \right. \\ &\quad \left. - \sum_{\mathbf{i}} \sum_{\mathbf{j}} m_{\mathbf{i}} m_{\mathbf{i}-\mathbf{j}} e^{-i(\mathbf{k}-\mathbf{j}) \cdot \theta(\mathbf{h})} \right] \\ &= \frac{(1-Q_{0,0})}{N n_d} \sum_{\mathbf{h}} D_{\mathbf{h},t} (1 - \psi_{\mathbf{h}}^2) e^{-i\mathbf{k} \cdot \theta(\mathbf{h})}, \quad (\text{A.33}) \end{aligned}$$

$\sum_{\mathbf{i}} m_{\mathbf{i}}^2 = \frac{1}{n_d} \sum_{\mathbf{h}} \psi_{\mathbf{h}}^2$ , and  $\sum_{\mathbf{i}} m_{\mathbf{i}} m_{\mathbf{i}-\mathbf{k}} = \frac{1}{n_d} \sum_{\mathbf{h}} \psi_{\mathbf{h}}^2 e^{-i\mathbf{k} \cdot \theta(\mathbf{h})}$ . Substituting these identities into eqn A.32 and using eqns A.20 and A.29 gives

$$\begin{aligned} \tau_{\bullet} &= -(1-s) \left[ 1 + \frac{1}{N n_d} \sum_{\mathbf{h}} \lim_{\mu \rightarrow 0} X_{\mathbf{h}} (1 - \psi_{\mathbf{h}}^2) - \frac{1}{N n_d} \sum_{\mathbf{h}} \psi_{\mathbf{h}}^2 \right] \\ \tau_{0,0} &= (1-s) \left[ \frac{1}{N} + \frac{1}{N n_d} \sum_{\mathbf{h}} \lim_{\mu \rightarrow 0} X_{\mathbf{h}} (1 - \psi_{\mathbf{h}}^2) - \frac{1}{N n_d} \sum_{\mathbf{h}} \psi_{\mathbf{h}}^2 \right] \\ \tau_{\mathbf{k},0} &= (1-s) \left[ \frac{1}{N n_d} \sum_{\mathbf{h}} \lim_{\mu \rightarrow 0} X_{\mathbf{h}} (1 - \psi_{\mathbf{h}}^2) e^{-i\mathbf{k} \cdot \theta(\mathbf{h})} - \frac{1}{N n_d} \sum_{\mathbf{h}} \psi_{\mathbf{h}}^2 e^{-i\mathbf{k} \cdot \theta(\mathbf{h})} \right] \text{ if } \mathbf{k} > \mathbf{0} \\ \tau_{\mathbf{k},t} &= (1-s) \left[ \frac{1}{N n_d} \sum_{\mathbf{h}} \lim_{\mu \rightarrow 0} Y_{\mathbf{h},t} (1 - \psi_{\mathbf{h}}^2) e^{-i\mathbf{k} \cdot \theta(\mathbf{h})} \right] \text{ if } \mathbf{k} > \mathbf{0}, t > 0. \quad (\text{A.34}) \end{aligned}$$

*Intra-temporal effects.* In order to evaluate  $\tau_{\bullet}$ ,  $\tau_{0,0}$  and  $\tau_{\mathbf{k},0}$ , we need the function  $\lim_{\mu \rightarrow 0} X_{\mathbf{h}} (1 - \psi_{\mathbf{h}}^2)$ . For  $\mathbf{h} \neq \mathbf{0}$ , this is from eqn A.20

$$Y_{\mathbf{h}} \equiv \frac{\psi_{\mathbf{h}} (1 + \psi_{\mathbf{h}}) [2s + (1-s)\psi_{\mathbf{h}}]}{1 + s + (1-s)\psi_{\mathbf{h}}}. \quad (\text{A.35})$$

For  $\mathbf{h} = \mathbf{0}$ ,  $Y_{\mathbf{0}} = 1 + s$  because  $\psi_{\mathbf{0}} = 1$ , while  $\lim_{\mu \rightarrow 0} X_{\mathbf{0}} (1 - \psi_{\mathbf{0}}^2) = 0$ . Hence

$$\frac{1}{n_d} \sum_{\mathbf{h}} \lim_{\mu \rightarrow 0} X_{\mathbf{h}} (1 - \psi_{\mathbf{h}}^2) e^{-i\mathbf{k} \cdot \theta(\mathbf{h})} = \frac{1}{n_d} \sum_{\mathbf{h}} Y_{\mathbf{h}} e^{-i\mathbf{k} \cdot \theta(\mathbf{h})} - \frac{1+s}{n_d}. \quad (\text{A.36})$$

Inserting this equation into eqn A.32 and using the identity

$$F_{\mathbf{h}} \equiv Y_{\mathbf{h}} - \psi_{\mathbf{h}}^2 = \frac{2s\psi_{\mathbf{h}}}{1 + [s + (1-s)\psi_{\mathbf{h}}]} \quad (\text{A.37})$$

gives

$$\tau_{\bullet} = -\frac{(1-s)}{N} \left[ N + \mathcal{L}_0(F_{\mathbf{h}}) - \frac{1+s}{n_d} \right]$$

$$\tau_{0,0} = \frac{(1-s)}{N} \left[ 1 + \mathcal{L}_0(F_{\mathbf{h}}) - \frac{1+s}{n_d} \right]$$

$$\tau_{\mathbf{k},0} = \frac{(1-s)}{N} \left[ \mathcal{L}_{\mathbf{k}}(F_{\mathbf{h}}) - \frac{1+s}{n_d} \right] \text{ if } \mathbf{k} > \mathbf{0}. \quad (\text{A.38})$$

*Inter-temporal effects.* In order to evaluate  $\tau_{\mathbf{k},t}$ , we need  $\lim_{\mu \rightarrow 0} Y_{\mathbf{h},t}(1 - \psi_{\mathbf{h}}^2)$ . For  $\mathbf{h} \neq \mathbf{0}$ , this is from eqn A.29

$$G_{\mathbf{h},t} \equiv \frac{(1+s)(1 + \psi_{\mathbf{h}}[s + (1-s)\psi_{\mathbf{h}}]^t)}{1 + s + (1-s)\psi_{\mathbf{h}}}. \quad (\text{A.39})$$

For  $\mathbf{h} = \mathbf{0}$ , we have  $G_{0,t} = 1 + s$ , while  $\lim_{\mu \rightarrow 0} Y_{0,t}(1 - \psi_0^2) = 0$ , whereby

$$\frac{1}{n_d} \sum_{\mathbf{h}} \lim_{\mu \rightarrow 0} Y_{\mathbf{h},t}(1 - \psi_{\mathbf{h}}^2)e^{-\mathbf{k} \cdot \theta(\mathbf{h})} = \frac{1}{n_d} \sum_{\mathbf{h}} G_{\mathbf{h},t}e^{-\mathbf{k} \cdot \theta(\mathbf{h})} - \frac{1+s}{n_d}, \quad (\text{A.40})$$

which, on substitution into eqn A.32, produces

$$\tau_{\mathbf{k},t} = \frac{(1-s)}{N} \left[ \mathcal{L}_{\mathbf{k}}(G_{\mathbf{h},t}) - \frac{1+s}{n_d} \right] \text{ if } \mathbf{k} > \mathbf{0}, t > 0. \quad (\text{A.41})$$

### Survival effects

*Transmission coefficients.* Substituting eqn A.11 into eqns A.13–A.14 gives

$$\tau_{\bullet} = -\lim_{\mu \rightarrow 0} \frac{s}{1 - Q_{0,0}} \left[ 1 - \sum_{\mathbf{j}} m_{\mathbf{j}} Q_{\mathbf{j},0} - m_0 \frac{(1 - Q_{0,0})}{N} \right]$$

$$\tau_{0,0} = \lim_{\mu \rightarrow 0} \frac{s}{1 - Q_{0,0}} \left[ \frac{(1 - Q_{0,0})}{N} + Q_{0,0} - \sum_{\mathbf{j}} m_{\mathbf{j}} Q_{\mathbf{j},0} - m_0 \frac{(1 - Q_{0,0})}{N} \right]$$

$$\tau_{\mathbf{k},0} = \lim_{\mu \rightarrow 0} \frac{s}{1 - Q_{0,0}} \left[ Q_{\mathbf{k},0} - \sum_{\mathbf{j}} m_{\mathbf{j}} Q_{\mathbf{j}-\mathbf{k},0} - m_{\mathbf{k}} \frac{(1 - Q_{0,0})}{N} \right] \text{ if } \mathbf{k} > \mathbf{0}$$

$$\tau_{\mathbf{k},t} = \lim_{\mu \rightarrow 0} \frac{s}{1 - Q_{0,0}} \left[ Q_{\mathbf{k},t} - \sum_{\mathbf{j}} m_{\mathbf{j}} Q_{\mathbf{j}-\mathbf{k},t} \right] \text{ if } \mathbf{k} > \mathbf{0}, t > 0. \quad (\text{A.42})$$

These equations will be simplified by using eqn A.31 and

$$\begin{aligned} Q_{\mathbf{k},t} - \sum_{\mathbf{j}} m_{\mathbf{j}} Q_{\mathbf{j}-\mathbf{k},t} &= \frac{(1 - Q_{0,0})}{Nn_d} \sum_{\mathbf{h}} D_{\mathbf{h},t} \left[ e^{-\mathbf{k} \cdot \theta(\mathbf{h})} - \sum_{\mathbf{j}} m_{\mathbf{j}} e^{-\mathbf{j} \cdot \theta(\mathbf{h})} \right] \\ &= \frac{(1 - Q_{0,0})}{Nn_d} \sum_{\mathbf{h}} D_{\mathbf{h},t} (1 - \psi_{\mathbf{h}}) e^{-\mathbf{k} \cdot \theta(\mathbf{h})}, \end{aligned} \quad (\text{A.43})$$

$m_0 = \frac{1}{n_d} \sum_{\mathbf{h}} \psi_{\mathbf{h}}$ , and  $m_{\mathbf{k}} = \frac{1}{n_d} \sum_{\mathbf{h}} \psi_{\mathbf{h}} e^{-\mathbf{k} \cdot \theta(\mathbf{h})}$ . Substituting these identities, eqns A.20, and A.29 into eqn A.42 gives

$$\tau_{\bullet} = -s \left[ 1 + \frac{1}{Nn_d} \sum_{\mathbf{h}} \lim_{\mu \rightarrow 0} X_{\mathbf{h}} (1 - \psi_{\mathbf{h}}) - \frac{1}{Nn_d} \sum_{\mathbf{h}} \psi_{\mathbf{h}} \right]$$

$$\tau_{0,0} = s \left[ \frac{1}{N} + \frac{1}{Nn_d} \sum_{\mathbf{h}} \lim_{\mu \rightarrow 0} X_{\mathbf{h}} (1 - \psi_{\mathbf{h}}) - \frac{1}{Nn_d} \sum_{\mathbf{h}} \psi_{\mathbf{h}} \right]$$

$$\begin{aligned} \tau_{\mathbf{k},0} &= s \left[ \frac{1}{Nn_d} \sum_{\mathbf{h}} \lim_{\mu \rightarrow 0} X_{\mathbf{h}} (1 - \psi_{\mathbf{h}}) e^{-\mathbf{k} \cdot \theta(\mathbf{h})} - \frac{1}{Nn_d} \sum_{\mathbf{h}} \psi_{\mathbf{h}} e^{-\mathbf{k} \cdot \theta(\mathbf{h})} \right] \text{ if } \mathbf{k} > \mathbf{0} \end{aligned}$$

$$\tau_{\mathbf{k},t} = s \left[ \frac{1}{Nn_d} \sum_{\mathbf{h}} \lim_{\mu \rightarrow 0} Y_{\mathbf{h},t} (1 - \psi_{\mathbf{h}}) e^{-\mathbf{k} \cdot \theta(\mathbf{h})} \right] \text{ if } \mathbf{k} > \mathbf{0}, t > 0. \quad (\text{A.44})$$

*Intra-temporal effects.* In order to evaluate  $\tau_{\bullet}$ ,  $\tau_{0,0}$ , and  $\tau_{\mathbf{k},0}$  under survival effects, we need the function  $\lim_{\mu \rightarrow 0} X_{\mathbf{h}}(1 - \psi_{\mathbf{h}})$ . For  $\mathbf{h} \neq \mathbf{0}$ , this is

$$Y_{\mathbf{h}} \equiv \frac{\psi_{\mathbf{h}}[2s + (1-s)\psi_{\mathbf{h}}]}{1 + s + (1-s)\psi_{\mathbf{h}}}. \quad (\text{A.45})$$

For  $\mathbf{h} = \mathbf{0}$ ,  $Y_0 = (1+s)/2$ , while  $\lim_{\mu \rightarrow 0} X_0(1 - \psi_0) = 0$ , hence

$$\sum_{\mathbf{h}} \lim_{\mu \rightarrow 0} X_{\mathbf{h}} (1 - \psi_{\mathbf{h}}) e^{-ik \cdot \theta(\mathbf{h})} = \frac{1}{n_d} \sum_{\mathbf{h}} Y_{\mathbf{h}} e^{-ik \cdot \theta(\mathbf{h})} - \frac{1+s}{2n_d}. \quad (\text{A.46})$$

Inserting this equation into eqn A.44 and using the identity

$$F_{\mathbf{h}} \equiv \frac{\psi_{\mathbf{h}} [2s + (1-s)\psi_{\mathbf{h}}]}{1 + [s + (1-s)\psi_{\mathbf{h}}]} - \psi_{\mathbf{h}} = -\frac{(1-s)\psi_{\mathbf{h}}}{1 + s + (1-s)\psi_{\mathbf{h}}} \quad (\text{A.47})$$

gives

$$\tau_{\bullet} = -\frac{s}{N} \left[ N + \mathcal{L}_0(F_{\mathbf{h}}) - \frac{1+s}{2n_d} \right]$$

$$\tau_{0,0} = \frac{s}{N} \left[ 1 + \mathcal{L}_0(F_{\mathbf{h}}) - \frac{1+s}{2n_d} \right]$$

$$\tau_{\mathbf{k},0} = -\frac{s}{N} \left[ \mathcal{L}_{\mathbf{k}}(F_{\mathbf{h}}) - \frac{1+s}{2n_d} \right] \text{ if } \mathbf{k} > \mathbf{0}. \quad (\text{A.48})$$

*Inter-temporal effects.* In order to evaluate  $\tau_{\mathbf{k},t}$ , we need the function  $\lim_{\mu \rightarrow 0} Y_{\mathbf{h},t}(1-\psi_{\mathbf{h}})$ . For  $\mathbf{h} \neq \mathbf{0}$  this is

$$G_{\mathbf{h},t} \equiv \frac{(1+s)[s + (1-s)\psi_{\mathbf{h}}]^t}{1 + s + (1-s)\psi_{\mathbf{h}}}. \quad (\text{A.49})$$

For  $\mathbf{h} = \mathbf{0}$ ,  $G_{0,t} = (1+s)/2$ , while  $\lim_{\mu \rightarrow 0} Y_{0,t}(1-\psi_0) = 0$ , hence

$$\sum_{\mathbf{h}} \lim_{\mu \rightarrow 0} Y_{\mathbf{h}} (1 - \psi_{\mathbf{h}}) e^{-ik \cdot \theta(\mathbf{h})} = \frac{1}{n_d} \sum_{\mathbf{h}} G_{\mathbf{h},t} e^{-ik \cdot \theta(\mathbf{h})} - \frac{1+s}{2n_d}, \quad (\text{A.50})$$

which, on substitution into eqn A.44, produces

$$\tau_{\mathbf{k},t} = \frac{s}{N} \left[ \mathcal{L}_{\mathbf{k}}(G_{\mathbf{h},t}) - \frac{1+s}{2n_d} \right] \text{ if } \mathbf{k} > \mathbf{0}, t > 0. \quad (\text{A.51})$$

#### Appendix IV: small migration approximation

The expressions for the inverse Fourier transform of  $F_{\mathbf{h}}$  (eqn A.37 for fecundity effects and eqn A.47 for survival effects) and  $G_{\mathbf{h},t}$  (eqn A.39 for fecundity effects and eqn A.49 for survival effects) are complicated and will be approximated in this appendix by assuming that the migration rate  $m$  is very small (weak migration approximation). We can then write  $m_0 = (1-m)$  and  $m_i = mg_i$  (Rousset 2004, chapter 3). From these definitions, the characteristic function of the dispersal distribution can be expressed as  $\psi = 1-mx$ , where  $x = 1 - \sum_{i \neq 0} g_i e^{i \cdot \mathbf{h}}$ .

#### Fecundity effects

Inserting  $\psi = 1-mx$  into  $F_{\mathbf{h}}$  (eqn A.37) and  $G_{\mathbf{h},t}$  (eqn A.39) and Taylor expanding around  $m = 0$  gives

$$F_{\mathbf{h}} = \frac{s(1-s + (1+s)\psi)}{2} + O(m^2)$$

$$G_{\mathbf{h},t} = (1+s) \left[ 1 + \frac{\{s + 2(1-s)t\}(\psi - 1)}{2} \right] + O(m^2). \quad (\text{A.52})$$

Noting that the inverse transform of a constant  $a$  is  $\mathcal{L}_{\mathbf{k}}(a) = 0$  except that  $\mathcal{L}_0(a) = a$ , we obtain after rearrangements from eqns A.38 and A.41 the approximations

$$\tau_{\bullet} = \frac{(1-s)}{N} \left[ N + \frac{s\{2 - (1+s)m\}}{2} - \frac{1+s}{n_d} \right]$$

$$\tau_{0,t} = \frac{(1-s)}{N} \left[ 1 + \frac{s\{2 - (1+s)m\} - 2m(1-s^2)t}{2} - \frac{1+s}{n_d} \right]$$

$$\tau_{\mathbf{k},t} = \frac{(1-s)}{N} \left[ \frac{(1+s)\{s + 2(1-s)t\}m_{\mathbf{k}}}{2} - \frac{1+s}{n_d} \right] \text{ if } \mathbf{k} > \mathbf{0}. \quad (\text{A.53})$$

Substituting these expressions into eqn 6 yields eqn 22 of the main text.

#### Survival effects

Inserting  $\psi = 1-mx$  into  $F_{\mathbf{h}}$  and  $G_{\mathbf{h},t}$  for survival effects (respectively eqns A.47 and A.49) and Taylor expanding around  $m = 0$  gives

$$F_{\mathbf{h}} = -\frac{(1-s)(1-s + (1+s)\psi)}{4} + O(m^2)$$

$$G_{\mathbf{h},t} = (1+s) \left[ \frac{3 - \psi + \{s + 2(1-s)t\}(\psi - 1)}{4} \right] + O(m^2). \quad (\text{A.54})$$

Inverse transforming these expressions and inserting into eqns A.48 and A.51 gives after rearrangements

$$\tau_{\bullet} = \frac{s}{N} \left[ N - \frac{(1-s)\{1-s + (1+s)(1-m)\}}{4} - \frac{1+s}{2n_d} \right]$$

$$\tau_{0,t} = \frac{s}{N} \left[ 1 - \frac{(1-s)\{1-s + (1+s)(1-m + 2tm)\}}{4} - \frac{1+s}{2n_d} \right]$$

$$\tau_{k,t} = \frac{s}{N} \left[ \frac{\{s + 2(1-s)t - 1\}m_k}{4} - \frac{1+s}{2n_d} \right] \text{ if } \mathbf{k} > \mathbf{0}. \quad (\text{A.55})$$

Substituting these expressions into eqn 6 yields eqn 22 of the main text.

### Appendix V: Fourier transforms for hierarchically clustered populations

We here recall some results of Sawyer & Felsenstein (1983) about Fourier transforms in the hierarchically clustered model that we will then apply. Fourier methods also more classically apply to more standard models of isolation by distance, where individuals are more likely to disperse locally, for instance to adjacent patches (Malécot 1975).

The characteristic function of the dispersal distribution can be written

$$\psi_{\mathbf{h}} = \sum_{k=0}^H \sum_{\mathbf{a} \in A(k)} \frac{p_k}{r_k} \exp \left[ 2\pi i \sum_{j=1}^k a_j h_j / n_j \right], \quad (\text{A.56})$$

where  $A(0) = \{0\}$  and  $A(k) = \{\mathbf{a} \in S: a_k \neq 0, a_{k+1} = \dots = a_H = 0\}$  for  $k > 0$  (Sawyer & Felsenstein 1983, p. 4), so that the number of elements in the set  $A(k)$  is  $|A(0)| = 1$  and  $|A(k)| = n_1 n_2 \dots (n_k - 1) = r_k$  for  $k > 0$ .

If  $h_1 = h_2 = \dots = h_{l-1} = 0$  and  $h_l \neq 0$ , then

$$\sum_{\mathbf{a} \in A(k)} e^{2\pi i \sum_{j=1}^k a_j h_j / n_j} = \begin{cases} \sum_{\mathbf{a} \in A(k)} e^0 = |A(k)| & \text{if } 0 \leq k < l \\ \sum_{\mathbf{a} \in A(l)} e^{2\pi i a_l h_l / n_l} = -n_1 n_2 \dots n_{l-1} & \text{if } k = l \\ \sum_{\mathbf{a} \in A(k)} e^{2\pi i \sum_{j=1}^k a_j h_j / n_j} = 0 & \text{if } l < k \leq H, \end{cases} \quad (\text{A.57})$$

where the second line is obtained by noting that  $\sum_{a_l=1}^{n_l-1} e^{2\pi i a_l h_l / n_l} = -1$  can be factored out of the sum and third line is obtained by noting that  $\sum_{a_l=0}^{n_l-1} e^{2\pi i a_l h_l / n_l} = 0$  can be factored out of the sum. Thereby  $\psi_{\mathbf{0}} = 1$  for  $l = 0$  ( $d(\mathbf{0}, \mathbf{h}) = 0$ ) and for  $d(\mathbf{0}, \mathbf{h}) = l > 0$

$$\psi_{\mathbf{h}} \equiv \psi_l = p_0 + p_1 + \dots + p_{l-1} - p_l / (n_l - 1) \quad (\text{A.58})$$

(Sawyer & Felsenstein 1983, eqn 2.7). Relative compact explicit expressions for  $\psi_l$  can be found for specific dispersal distribution. For instance, when dispersal is panmictic  $p_0 = 1 - m$  and  $p_k = m r_k / (n_d - 1)$  for  $k > 0$  (Sawyer

& Felsenstein 1983). On substitution of these equations into A.58 and rearranging produces

$$\psi_l = 1 - m n_d / (n_d - 1). \quad (\text{A.59})$$

One may also assume that the dispersal distribution is given by a truncated geometric distribution:  $p_0 = 1 - m$  and  $p_k = m(1-g)g^{k-1} / (1-g^H)$  for  $k > 0$  and that  $n_l = n$  for all  $l$ . This gives

$$\psi_l = 1 - \frac{m[(n-q)(q^l - q^{1+H}) + (1-q)q^{1+H}]}{(n-1)q(1-q^H)}, \quad (\text{A.60})$$

which, when  $n$  and  $H$  are very large (say  $n \rightarrow \infty$  and  $H \rightarrow \infty$ ), reduces to  $\psi_l = 1 - m q^{l-1}$ .

Equation A.58 will prove especially useful when evaluating inverse Fourier transforms of any function  $D(\psi_{\mathbf{b}})$  of the characteristic distribution  $\psi_{\mathbf{b}}$ . Let  $\mathcal{L}_{\mathbf{h}}(D) \equiv \frac{1}{n_d} \sum_{\mathbf{b} \in S} D(\psi_{\mathbf{b}}) e^{-i \mathbf{h} \cdot \theta(\mathbf{b})}$  be the inverse Fourier transform of  $D$  at  $\mathbf{b}$ . This equation can be written as

$$\mathcal{L}_{\mathbf{h}}(D) = \frac{1}{n_d} \sum_{k=0}^H \sum_{\mathbf{b} \in B(k)} D(\psi_{\mathbf{b}}) \exp \left[ -2\pi i \sum_{j=k}^H b_j h_j / n_j \right], \quad (\text{A.61})$$

where  $B(H) = \{0\}$  and  $B(k) = \{\mathbf{b} \in S: b_1 = \dots = b_2 = \dots = b_{k-1} = 0, b_k \neq 0\}$ . The number of elements in the set  $B(k)$  are  $|B(0)| = 1$  and  $|B(k)| = (n_k - 1) n_{k+1} \dots n_H$  for  $k > 0$ . The set  $B(k)$  corresponds precisely to the conditions leading to eqn A.58, so that

$$\mathcal{L}_{\mathbf{h}}(D) = \frac{1}{n_d} \sum_{k=0}^H D(\psi_k) \sum_{\mathbf{b} \in B(k)} \exp \left[ -2\pi i \sum_{j=k}^H b_j h_j / n_j \right]. \quad (\text{A.62})$$

If the hierarchical distance of group  $\mathbf{h}$  relative to a focal group is  $d(\mathbf{h}, \mathbf{0}) = l$ , then  $h_l \neq 0$  and  $h_{l+1} = h_{l+2} = \dots = h_H = 0$ . Therefore

$$\sum_{\mathbf{b} \in B(k)} e^{-2\pi i \sum_{j=k}^H b_j h_j / n_j} = \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{if } 0 < k < l \\ \sum_{\mathbf{b} \in B(l)} e^{-2\pi i b_l h_l / n_l} = -n_{l+1} n_{l+2} \dots n_H & \text{if } k = l \\ \sum_{\mathbf{b} \in B(k)} e^0 = |B(k)| & \text{if } l < k \leq H, \end{cases} \quad (\text{A.63})$$

where the second line is obtained by noting that  $\sum_{a_l=0}^{n_l-1} e^{-2\pi i a_l h_l / n_l} = 0$  can be factored out of the sum. With this, the Fourier transform becomes



$$\mathcal{L}_{\mathbf{h}}(D) = \frac{D(\psi_0)}{n_d} - \frac{D(\psi_l)}{n_1 n_2 \cdots n_l} + \sum_{j=l+1}^H \frac{D(\psi_j)(n_j - 1)}{n_1 n_2 \cdots n_j} \quad (\text{A.64})$$

for  $d(\mathbf{h}, 0) = l$ , where the middle term does not occur if  $l = 0$  and  $\psi_l$  is given by eqn A.58.

For instance, when  $D(\psi) = \psi^t$ ,  $\mathcal{L}_{\mathbf{h}}(\psi^t)$  gives the probability  $P_{\mathbf{h}, t}$  that a line of descent from an individual residing in the focal group will be in a group at hierarchical distance  $d(\mathbf{h}, 0)$  at  $t$  time periods in the future (see section ‘Semelparous reproduction’ of the main text).

For  $d(\mathbf{h}, 0) = l > 0$ , this is

$$P_{\mathbf{h}, t} = \frac{1}{n_d} - \frac{\left[ \sum_{i=0}^{l-1} p_i - p_l / (n_l - 1) \right]^t}{n_1 n_2 \cdots n_l} + \sum_{j=l+1}^H \frac{\left[ \sum_{i=0}^{j-1} p_i - p_j / (n_j - 1) \right]^t (n_j - 1)}{n_1 n_2 \cdots n_j}, \quad (\text{A.65})$$

which is equal to zero if  $t = 0$ , while

$$P_{0, t} = \frac{1}{n_d} + \sum_{j=1}^H \frac{\left[ \sum_{i=0}^{j-1} p_i - p_j / (n_j - 1) \right]^t (n_j - 1)}{n_1 n_2 \cdots n_j}, \quad (\text{A.66})$$

which is equal to one if  $t = 0$ . From eqn A.66, we can see that  $P_{0, t}$  is a decreasing function of  $t$  if  $p_i > p_{i+1}$  for all  $i$ , as it then involves only positive terms in the sum. By contrast, under the same conditions,  $P_{\mathbf{h}, t}$  for  $\mathbf{h} \neq 0$  can first increase, reach a value exceeding  $1/n_d$  and then decrease before approaching the asymptotic value of  $1/n_d$ . This is more likely to be the case for small hierarchical distances  $l$  as the last term in eqn A.65 (the sum) is then more likely to dominate the negative term.