JOURNAL OF Evolutionary Biology eseb European Society for Evolutionary Biology

doi: 10.1111/jeb.12462

SHORT COMMUNICATION

The robustness of the weak selection approximation for the evolution of altruism against strong selection

C. MULLON & L. LEHMANN

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

Keywords:

evolution of altruism;
strong selection;
limited dispersal;
weak selection.

Abstract

The weak selection approximation of population genetics has made possible the analysis of social evolution under a considerable variety of biological scenarios. Despite its extensive usage, the accuracy of weak selection in predicting the emergence of altruism under limited dispersal when selection intensity increases remains unclear. Here, we derive the condition for the spread of an altruistic mutant in the infinite island model of dispersal under a Moran reproductive process and arbitrary strength of selection. The simplicity of the model allows us to compare weak and strong selection regimes analytically. Our results demonstrate that the weak selection approximation is robust to moderate increases in selection intensity and therefore provides a good approximation to understand the invasion of altruism in spatially structured population. In particular, we find that the weak selection approximation is excellent even if selection is very strong, when either migration is much stronger than selection or when patches are large. Importantly, we emphasize that the weak selection approximation provides the ideal condition for the invasion of altruism, and increasing selection intensity will impede the emergence of altruism. We discuss that this should also hold for more complicated life cycles and for culturally transmitted altruism. Using the weak selection approximation is therefore unlikely to miss out on any demographic scenario that lead to the evolution of altruism under limited dispersal.

Introduction

Since Eshel (1972)'s seminal paper, mathematical modelling of the spread of alleles coding for altruistic behaviours in patch structured populations under limited dispersal has been the focus of intense research (see Frank, 1998 and Rousset, 2004 for general accounts). In spite of a rich literature and significant progress in this domain, characterizing the invasion condition of altruism under general demographic scenarios for arbitrary strengths of selection remains difficult, principally due to the frequency-dependent nature of selection on social behaviours. Deriving the invasion condition for an altruistic mutant allele requires taking into account the interplay between selection and local genetic drift on the local fluctuations of allele frequencies. This in turn requires tracking the distribution of mutant alleles within and across patches. The difficulty of this task means that studies so far have relied on simplifying assumptions to reach interpretable results.

Early work on the evolution of alleles coding for altruism in spatially structured population (Eshel, 1972; Aoki, 1982; Motro, 1982) is based on the classical island model of dispersal (Wright, 1931), with constant population size and haploid Wright–Fisher reproduction. Owing to the simplicity of their demographic assumptions, these studies were able to derive the recursion equations that describe the full distribution of allele frequencies in the population across generations under arbitrary strength of selection and thus characterize the kinship structure conducive to altruism. This allowed to gain insight into the spread of altruistic mutants when rare, and their stability when frequent,

Correspondence: Laurent Lehmann, Department of Ecology and Evolution, University of Lausanne, 1004 Lausanne, Switzerland.
Tel.: +4121 692 4183; fax: +4121 692 4165;
e-mail: laurent.lehmann@unil.ch

albeit in a qualitative manner because the mutant distribution for the exact process could not be determined. Generalizing this approach to take into account even the most basic biological features, such as diploidy or dioecy, has proven even more challenging. The resulting dynamical systems rapidly become intractable, and its complexity has notably led to confusions about the selection pressure on dispersal (Frank, 1998, pp. 117–120).

Later work on the evolution of altruism has approximated the invasion condition by assuming that selection is weak (e.g. Taylor, 1992a; Frank, 1998), which for instance occurs when difference in fitness between competing types is small (Nagylaki, 1992; Wild & Traulsen, 2007). Instead of having to track the full mutant distribution, studies that assume weak selection only need to evaluate the first and second moment of this distribution. Despite this simplification, the weak selection approximation still captures frequency-dependent interactions and the effect of local genetic drift using relatedness coefficients between patch members, which are conveniently calculated in the absence of selection (i.e. 'pedigree' or 'neutral' relatedness, Rousset, 2004). Relatedness quantifies the kinship structure of the population and thus how much a mutant is more likely to interact with another mutant than with a randomly sampled type from the population. In the island model under haploid reproduction, relatedness is straightforwardly obtained as the probability that a pair of homologous genes sampled in different individuals from the same patch are identical-by-descent.

By opening the door to tractable analysis of selection on social behaviour, the weak selection approximation has allowed for the study of altruism evolution in patch structured populations under considerably more realistic scenarios. This has led to an extensive and consistent analytical literature disentangling the role of various demographic, ecological, environmental, behavioural and genetic features in the evolution of altruistic helping under limited dispersal (e.g. Aoki, 1982; Rogers, 1990; Taylor, 1992a,b; van Baalen & Rand, 1998; Frank, 1998; Taylor & Irwin, 2000; Lehmann & Perrin, 2002; Le Galliard *et al.*, 2003; Roze & Rousset, 2004; Gardner & West, 2006; Lehmann *et al.*, 2006; Ohtsuki *et al.*, 2006; Grafen, 2007a,b; Lehmann & Balloux, 2007; Lion & van Baalen, 2007; Rousset & Roze, 2007; Rousset & Roze, 2007; Alizon & Taylor, 2008; Johnstone & Cant, 2008; Sozou, 2009; Wild & Fernandes, 2009; Gardner, 2010; Lion & Gandon, 2010; Ohtsuki, 2010; Van Dyken, 2010; Akcay & Van Cleve, 2012; Bao & Wild, 2012; Johnstone *et al.*, 2012; Kuijper & Johnstone, 2012; Ohtsuki, 2012; Rodrigues & Gardner, 2012; Taylor & Maciejewski, 2012; Van Dyken & Wade, 2012; Yeh & Gardner, 2012). In general, conditions such as a low-dispersal probability, small patch size, overlapping generations or high reproductive variance, which increase relatedness between patch members and therefore the tendency for altruists to interact with other altruists, facilitate the evolution of altruism.

In spite of the significant number of studies using the weak selection approximation, it remains unclear how robust the approximation is against increased selection intensity. Analytical work assessing the accuracy of the approximation has been limited to well-mixed populations (i.e. no limited dispersal, Antal *et al.*, 2009; Wu *et al.*, 2013) and showed that its accuracy depends on the number of alleles present in the population. Under limited dispersal, the accuracy of the weak selection approximation has only been evaluated by individually based simulations for various social behaviours (e.g. Pen, 2000; Leturque & Rousset, 2002; Le Galliard *et al.*, 2003; Roze & Rousset, 2003; Guillaume & Perrin, 2006; Lehmann *et al.*, 2007b; Rousset & Roze, 2007; Lion & Gandon, 2010), and it has shown that the approximation works well when the intensity of selection on a mutant is of the order of 10^{-2}. However, no systematic analysis of how the condition for the spread of altruism varies with the selection intensity exists for structured populations.

To evaluate how sensitive the conclusions based on the weak selection approximation are to increased selection intensity, we study in this article the invasion condition of an altruistic mutant under various selection intensities. To that aim, we revisit the simplest demographic scenario for a patch structured population. We use the island model of dispersal but with a Moran, rather than a Wright–Fisher reproductive process, which allows us to characterize the invasion condition analytically for arbitrary selection strength. The rest of this article is organized as follows. First, we derive the general invasion condition of a mutant in the infinite island model of dispersal. Then, we compare the weak and strong selection regime for the spread of an altruistic mutant arising as a single copy. Finally, we discuss our results in connection with the broader literature on the evolution of altruism in patch structured populations.

Model

Biological assumptions

We model the spread of an allele in a structured population made up of an infinite number of patches, each of constant size N. The life cycle is as follows: (i) adult individuals interact socially with each other within patches and receive pay-offs from these interactions; (ii) each individual produces a very large number of offspring proportional to their pay-offs; (iii) each offspring either disperses to another patch with probability m or remains in its natal patch with complementary probability $1-m$; and (iv) on each patch, a randomly sampled adult dies and offspring compete for the vacated breeding spot. This life cycle corresponds to

Wright's (1931) island model of dispersal with a Moran reproductive process (Ewens, 2004).

Individuals are haploid, and two alleles segregate in the population, a mutant (A) and a resident (a). When an individual expresses the mutant allele, its pay-off obtained during social interactions is reduced by a fixed cost C, while the $N-1$ patch neighbours receive a benefit B. An individual expressing the resident allele pays no cost, but reaps the benefits produced by neighbouring mutants. Then, in a patch with i mutants, the pay-offs received by a mutant and by a resident, respectively, are

$$\begin{aligned}\pi_{\mathrm{A},i} &= B\frac{(i-1)}{N-1} - C\\ \pi_{\mathrm{a},i} &= B\frac{i}{N-1}.\end{aligned} \tag{1}$$

Invasion fitness

Whether an allele that has arisen in a single copy by mutation in a patch will invade or go extinct can be answered by evaluating the growth rate of the mutant lineage when rare in the population (Metz & Gyllenberg, 2001; Ajar, 2003). We denote this quantity by ρ and call it the invasion fitness of the mutant allele. To derive ρ, we first observe that as the mutant is initially rare, no mutant can immigrate into the focal patch (the patch where the mutant has appeared). Therefore, the change in the number $X(t) \in \{0,1,2,\ldots,N\}$ of mutants in the focal patch at time t, and descending from the founding mutant at $t = 0$ ($X(0) = 1$), can be described by a time homogeneous Markov chain, with transition probability denoted by $p_{ij} = \Pr(X(t+1) = j|X(t) = i)$. In the absence of mutant immigration and with $m > 0$, we have $p_{00} = 1$, and extinction of the mutant lineage in the focal patch is the only absorbing state of the Markov chain.

We can obtain the growth rate of the mutant by first calculating the expected number of successful emigrant mutants that are produced in the focal patch in the time between the appearance of the mutant and the extinction of its lineage (Chesson, 1984; Metz & Gyllenberg, 2001; Ajar, 2003; Massol *et al.*, 2009). To do so, we denote by $p_{t,i} = \Pr(X(t) = i) = \sum_k p_{ki}\Pr(X(t-1) = k)$ the probability there are i mutants at time t, and by $\bar{t}_i = \sum_{t=0}^{\infty} p_{t,i}$ the mean number of time steps during which there are i mutants before lineage extinction, also called the sojourn time in state i. Then, with e_i as the expected number of successful emigrants produced by a single mutant on the focal patch in which there are i mutants (and conditional on the rest of the population being monomorphic for the resident), invasion fitness is given by:

$$\rho = \sum_{i=1}^{N} i e_i \bar{t}_i - 1, \tag{2}$$

where the first term is the expected number of successful emigrants produced by the mutant lineage, from the origin of the founding mutant until the local extinction of the lineage (Chesson, 1984; Metz & Gyllenberg, 2001; Ajar, 2003; Massol *et al.*, 2009).

The sign of invasion fitness, ρ, indicates the direction of selection on the mutant. In a monomorphic population, that is when the mutant has no phenotypic effect on fitness and there is no selection pressure acting upon it, $\rho = 0$. This can be seen from eqn (2) by considering that a lineage founded by a single individual, which eventually goes extinct locally, must produce on average one emigrant for the population size to remain constant ($\sum_{i=1}^{N} i e_i \bar{t}_i = 1$). But a mutant lineage that produces on average more than one future emigrants, $\rho > 0$, will invade. Conversely, a mutant lineage that produces on average less than one future emigrants, $\rho < 0$, will die out.

In Appendix A, we show that invasion fitness of the mutant can be equivalently expressed as:

$$\rho = \sum_{i=1}^{N} (w_i - 1) i \bar{t}_i, \tag{3}$$

where w_i is the fitness of a mutant carrier (the expected total number of adult offspring produced by a mutant carrier) when there are i mutants in the focal patch. Individual fitness is given by the sum of the expected number of adult offspring that successfully emigrate (e_i), with those that successfully establish in the focal patch, i.e. the philopatric component of fitness, that we write ϕ_i:

$$w_i = \phi_i + e_i. \tag{4}$$

Equation (3) turns out to be sometimes more practical than the one given by eqn (3) to evaluate invasion fitness explicitly. In addition, eqn (3) shows immediately that in the absence of selection, that is when each individual has the same fitness ($w_i = 1$), the invasion fitness reduces to $\rho = 0$. This result had so far only been reached numerically (Metz & Gyllenberg, 2001, p. 502), or for the Wright–Fisher process by involved computations (Ajar, 2003, eq. 34).

Fitness

To derive the invasion fitness of the altruistic mutant (eqn 4), we first calculate the components of individual fitness, ϕ_i and e_i (eqn 5). To do so, we write the relative number of offspring produced by a mutant and a resident in a patch with i mutants, respectively, as

$$\begin{aligned}f_{\mathrm{A},i} &= 1 + \delta\pi_{\mathrm{A},i}\\ f_{\mathrm{a},i} &= 1 + \delta\pi_{\mathrm{a},i},\end{aligned} \tag{5}$$

where $\delta \in [0,1]$ is a parameter that tunes the strength of selection, that is the extent to which pay-offs affect reproductive output. Then, for a Moran process, we have:

$$\begin{aligned}\phi_i &= \frac{N-1}{N} + \frac{1}{N}\frac{(1-m)f_{\mathrm{A},i}}{(1-m)\big(i\, f_{\mathrm{A},i}/N + (N-i)f_{\mathrm{a},i}/N\big) + m\, f_{\mathrm{a},0}}\\ e_i &= \frac{m\, f_{\mathrm{A},i}}{N\, f_{\mathrm{a},0}},\end{aligned} \tag{6}$$

in agreement with standard expressions for fitness in the island model (e.g. Gandon, 1999).

Local distribution of mutants

To compute ρ (eqn 3), we also need to calculate the sojourn time $\bar{t}_i$ in state i, that is in our model, the expected number of generations during which i altruists are present. This is found by considering the Markov chain that describes local lineage size $X(t)$. As reproduction follows a Moran process, the associated Markov chain describes a so-called birth–death process (e.g. Karlin & Taylor, 1975; Grimmett & Stirzaker, 2001), whose transition probabilities are:

$$p_{ij} = \begin{cases} b_i, & \text{if} \quad j = i+1 \text{ ('birth' of a mutant)}\\ d_i, & \text{if} \quad j = i-1 \text{ ('death' of a mutant)}\\ 1-(b_i+d_i) & \text{if} \quad j = i \text{ ('no change')}\\ 0 & \text{otherwise.}\end{cases} \tag{7}$$

Standard results on birth–death processes (e.g. Ewens, 2004, eq. 2.160) show that when the initial state of the chain is $X(0) = 1$, $\bar{t}_i$ is given by:

$$\bar{t}_i = \frac{1}{d_1}\prod_{k=1}^{i-1}\frac{b_k}{d_{k+1}}, \tag{8}$$

where, given our assumptions, the birth and death probabilities are given by:

$$\begin{aligned} b_i &= \frac{(N-i)}{N}\frac{(1-m)i f_{\mathrm{A},i}}{(1-m)\big(i\, f_{\mathrm{A},i} + (N-i)f_{\mathrm{a},i}\big) + mN\, f_{\mathrm{a},0}}\\ d_i &= \frac{i}{N}\frac{(1-m)(N-i)f_{\mathrm{a},i} + mN f_{\mathrm{a},0}}{(1-m)\big(i\, f_{\mathrm{A},i} + (N-i)f_{\mathrm{a},i}\big) + mN\, f_{\mathrm{a},0}}.\end{aligned} \tag{9}$$

With eqns (8)–(9), we have all the elements necessary to evaluate invasion fitness under a Moran process. While it is possible to use eqns (2)–(3) to evaluate the invasion fitness for different demographic scenarios, computing the sojourn times $\bar{t}_i$ will generally be much more complicated. Even for a simple Wright–Fisher model (where each individual on a patch dies after reproduction), finding out $\bar{t}_i$ rapidly becomes computationally expensive with patch size N because all the transition probabilities p_{ij} are nonzero, which means that calculating $\bar{t}_i$ requires inverting a nonsparse $N \times N$ transient matrix (the transition matrix without the absorbing states). In contrast, most of the transitions probabilities are zero for the Moran process (eqn 7). The transient matrix of the Markov chain in this case can easily be inverted, leading to eqn (8), which we use to compute the invasion fitness for arbitrary patch size.

Results

Weak selection

We first look at the condition for the spread of altruism when selection is weak. If individual fitness w_i (eqn 4) is Taylor expanded around $\delta = 0$, we show in Appendix B that:

$$\rho = \delta\frac{1-m+N}{1+m(N-1)}\left(-C + \frac{1-m}{(1-m)+N}B\right). \tag{10}$$

From eqn (10), the cost-to-benefit ratio C/B under which selection favours the spread of altruism ($\rho > 0$) must satisfy:

$$\frac{C}{B} < \frac{(1-m)}{N+(1-m)} \tag{11}$$

(Fig. 1), which is consistent with previous results reached by assuming weak selection from the outset (eqn A-10 of Taylor & Irwin (2000) with $s \rightarrow 1$, eqn 8 of Lehmann *et al.* (2007a) for an infinite island model of dispersal). Eqn (11) displays the classical result that small patches and weak migration favour the spread of altruism (e.g. Eshel, 1972; Taylor & Irwin, 2000). This occurs because in those cases, individuals within patches are related: they are more likely to carry homologous genes identical-by-descent from a common recent ancestor than are two individuals sampled at random from the population. For the altruistic allele, the common recent ancestor is the founder of the

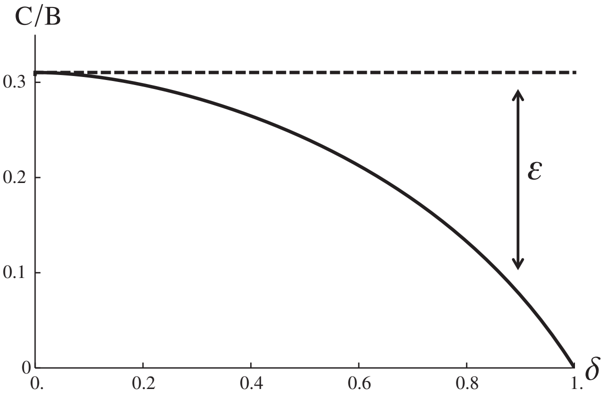

Fig. 1 Threshold cost-to-benefit ratio C/B under which selection favours altruism as a function of selection δ. The other parameters are set at $m = 0.1$, $N = 2$ and $C = 1$. The dashed line corresponds to the weak selection approximation (eqn 11), while the full line corresponds to the exact invasion condition (eqn 13). The difference between these two curves gives the error ε.

mutant lineage. The right side of eqn (11) is a measure of relatedness that is here demographically scaled, taking into account local competition (e.g. Queller, 1994; Lehmann & Rousset, 2010).

Strong selection

We now present the conditions for the invasion of altruism for arbitrary levels of selection δ. We first consider the case where there are only two individuals per patch ($N = 2$) as this turns out to be fully tractable. Eqn (2) then reduces to $\rho = e_1/d_1 + (2e_2b_1)/(d_1d_2) - 1$, and using eqns (1)–(9), invasion fitness reads:

$$\rho = \frac{(1-\delta C)\left(1+m+\delta 2(1-m)(B-C)+\delta^2(1-m)^2(B-C)^2\right)}{1+m+\delta(1-m)B} - 1. \tag{12}$$

In contrast to eqn (10), invasion fitness under strong selection (eqn 12) involves quadratic terms, and it is no longer possible to evaluate the invasion condition simply in terms of the cost-to-benefit ratio C/B. We therefore present results with the value $C = 1$, which is the maximum cost that can be impaired on an individual in the absence of benefits and corresponds to the most restrictive condition for altruism to invade. In the absence of benefits conferred by other patch members, an individual has zero fitness if selection is maximal ($\delta = 1$), that is the individual commits self-sacrifice with probability one. The conservative value $C = 1$ generates the greatest difference between the strong and weak selection regimes, but we note here that all our results below still hold when $C < 1$.

For an arbitrary intensity of selection, altruism spreads if:

$$\frac{1}{B} < \frac{(1-m)\left(1+2\delta(m-2+\delta-m\delta)+\sqrt{1+4\delta(1-\delta)}\right)}{2\left(3-m-(3-m)(1-m)\delta+(1-m)^2\delta^2\right)}, \tag{13}$$

which reduces to eqn (11) (with $C = 1$) when $\delta \to 0$. But what happens when selection intensity is not vanishingly small? Figs 1 and 2a illustrate that increasing δ reduces the scope for the invasion of altruism. When $\delta = 1$, the right side of eqn (13) is zero, and therefore, altruism can never spread when rare. This is not surprising as in this case, a single mutant individual has zero fecundity and so the mutant lineage goes extinct after the first round of reproduction. It is therefore impossible for relatedness to build up due to local genetic drift.

As shown in Figs 1 and 2a, inference made about the invasion of altruism based on the weak selection approximation is robust to reasonable changes in δ. We can gain quantitative insight into this robustness by considering the difference between the invasion condition under weak (eqn 11) and strong selection (eqn 13):

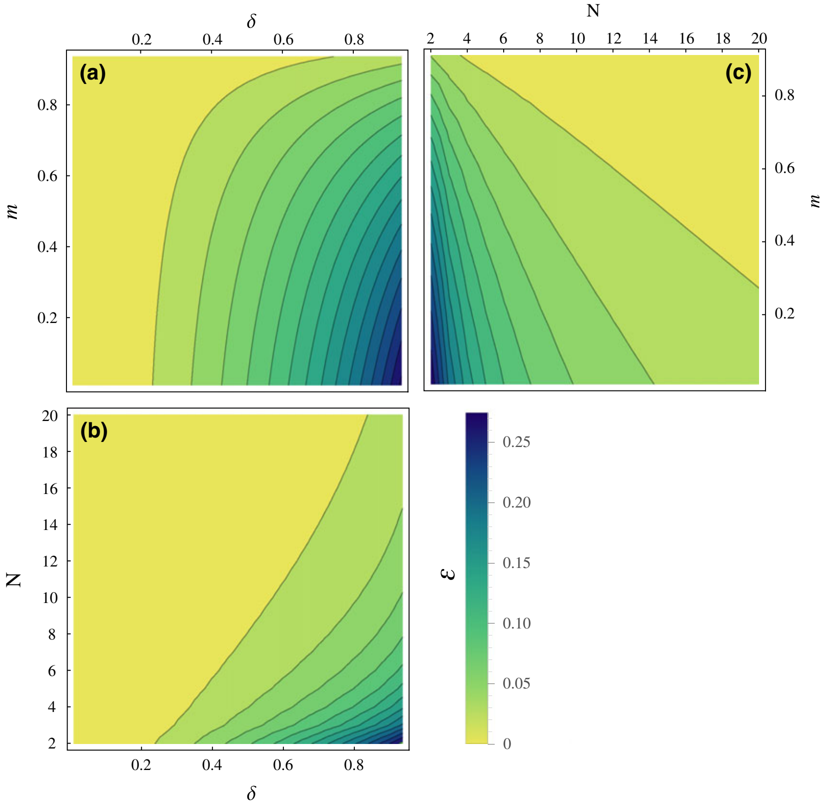

Fig. 2 Error ε between the approximated and true invasion condition. (a) Error as a function of selection intensity δ (x-axis) and migration m (y-axis) with $N = 2$ (eqn 14). (b) Error as a function of selection intensity δ (x-axis) and patch size N (y-axis) with $m = 0.1$ (using eqn 1 into eqn 5 along with eqns 8–9). (c) Error as a function of patch size N (x-axis) and migration m (y-axis) with $\delta = 0.9$. As indicated by the scale on the right of panel B, darker colours represent greater magnitude of error.

$$\epsilon = \frac{1}{2}(1-m)\left(\frac{2}{3-m} + \frac{1+2\delta(m-2+\delta-m\delta)-\sqrt{1+4\delta(1-\delta)}}{3-m-(3-m)(1-m)\delta+(1-m)^2\delta^2}\right), \tag{14}$$

which can be thought of as the error associated with using the weak selection approximation (Fig. 1). This error satisfies:

$$0 \leq \epsilon \leq 4\delta^2 \frac{(1-m)}{(3-m)^2}. \tag{15}$$

So, the error converges to zero at least as fast as δ^2 does ($O(\delta^2)$).

Furthermore, eqn (15) and Fig. 2a reveal the weak selection approximation is increasingly robust to increases in δ as migration increases. In particular, if migration is significantly larger than the intensity of selection ($m \gg \delta$), then the error is very small for all selection intensities δ (Fig. 2a). This can be understood by observing that when migration is large, demographic effects on allele frequency change balance with those of strong selection, and therefore, the overall effects of selection remain weak.

The condition that $m \gg \delta$ and its consequences on demography have been useful to develop approximations of evolutionary processes under limited dispersal. If in addition to $m \gg \delta$, selection intensity becomes vanishingly small ($\delta \to 0$), then the evolutionary dynamics can be approximated by the so-called quasi-equilibrium approximation (QE, Roze & Rousset, 2008), which assumes that genetic associations due to demographic factors go to equilibrium before those due to selective factors. Notably, the QE approximation is implicitly used in the application of the 'direct fitness method' (Taylor & Frank, 1996), and underpins most of the literature cited in the introduction. When the population is finite, the QE approximation can be refined to the diffusion approximation under limited dispersal (Roze & Rousset, 2003). The diffusion approximation is particularly useful to derive detailed information about the segregation process, like the fixation probability or the expected time to fixation, and it has been shown by simulations to be accurate when selection is of the order 10^{-2} (and $m > 10^{-2}$, Roze & Rousset, 2003). Here, we have seen that if one is interested only in the invasion condition of the mutant allele, then using the weak selection, approximation generates correct results when $m \gg \delta$, even if the selection intensity is very strong (Fig. 2a).

For arbitrary patch size N and strong selection, the invasion fitness of the mutant ρ can be expressed in terms of hypergeometric functions, which do not lead to a simple expression for invasion fitness. We therefore evaluate numerically the threshold value for $1/B$ under which altruism can invade (assuming that $C = 1$) and compute the difference between this threshold and the one found under weak selection (eqn 11), giving the error ε. Fig. 2b shows that for all values of N, the threshold value for $1/B$ decreases slowly relative to that found under weak selection as δ increases. In addition, we find that the error in using the weak selection approximation decreases markedly with increases in population size (Fig. 2b). Hence, the maximum error in using the weak selection approximation for all $N \geq 2$ is also given by eqn(1).

Altogether, we find that the invasion condition is less sensitive to increases in selection intensity (δ) when groups are large and migration is strong (Fig. 2a,b), although the effect of migration on the accuracy of the weak selection approximation vanishes as patch size increases (Fig. 2c). The effects of patch size and migration stem from the fact that in larger groups or when migration is strong, the total effect of local genetic drift decreases. As a result, the invasion fitness of an altruistic allele becomes less sensitive to variation in the selection intensity. However, it should be noted that the spread of the altruistic mutant still relies on genetic drift to initially build up relatedness, and therefore, the probability of its invasion decreases with patch size and migration.

Discussion

The weak selection approximation of population genetics (e.g. Nagylaki, 1992) has allowed for a significant progress in delineating the role of various environmental and demographic factors in the evolution of altruistic helping under limited dispersal (e.g. Lehmann & Rousset, 2010 for a review). Yet, the reliability of the conclusions based on weak selection approximations in the face of increased selection pressure had so far not been investigated in a systematic manner for structured populations. To gain insight into the robustness of the weak selection approximation, we derived the invasion fitness of an altruistic mutant in the infinite island model of dispersal. We modelled reproduction as a Moran process, unlike previous work with arbitrary selection strength, which had used a Wright–Fisher process (Eshel, 1972; Aoki, 1982; Motro, 1982; Wild & Fernandes, 2009). This allowed us to derive fully tractable equations for the invasion fitness (e.g. eqn 12). We then compared analytically the invasion condition found under weak selection (eqn 11) with that found under arbitrary strength of selection (eqn 13).

We find that when selection is so strong that a carrier of the altruistic mutant allele has zero fitness in the absence of any other altruist ($\delta = 1$, $C = 1$), an initial mutant allele can never spread as it is immediately driven to extinction after one episode of reproduction (eqn 13). If this result is hardly surprising, it highlights that for altruism to spread under limited dispersal, selection must be sufficiently weak for a local cluster of mutants to build up under the action of local genetic drift. In other words, relatedness must accumulate at the altruistic locus as a result of stochastic sampling within the patch where the initial mutant appeared. Only then can favourable mutant–mutant interactions occur and the mutant be picked up by positive selection.

The weak selection approximation under limited dispersal quantifies the role of genetic drift through the probability that homologous genes sampled in different individuals within the same patch are identical-by-descent in the absence of selection ($\delta = 0$, e.g. Rousset, 2004). This means that genetic drift is allowed to play its maximal role in building up relatedness between patch members. To evaluate the level of relatedness that favours altruism under stronger selection requires considering the effects of selection on the probabilities of identity-by-descent. Yet, simulations have shown that using neutral probabilities of identity-by-descent in invasion fitness provides a good approximation when the selection intensity is of the order $\delta = 0.01$ (Roze & Rousset, 2004), and even up to $\delta = 0.2$ (Lehmann *et al.*, 2007b; Rousset & Roze, 2007) in certain scenarios.

Our analysis gives formal support to the observation that the weak selection approximation is robust to moderate increases in selection intensity (eqn 15, Figs 1 and 2a,b) and sheds lights on the effects of patch size and migration on the validity of this approximation. When patches are very small, then the error associated with weak selection can be significant, unless migration is strong. In particular, the error becomes vanishingly small when migration is substantially larger than selection ($m \gg \delta$, Fig. 2a). As patch size increases, the error diminishes (Fig. 2b) and becomes largely independent of the migration rate (Fig. 2c). Therefore, unless the initial mutant produces no offspring at all ($\delta = 1$, $C = 1$), in which case local genetic drift cannot favour the emergence altruism, the spread of altruism is accurately predicted by the weak selection approximation under strong selection when either migration is substantially larger than selection ($m \gg \delta$) or patches are large ($N \gg 1$).

The model presented in this paper also illustrates that weak selection provides the most favourable conditions for the spread of altruism. This means that if the weak selection approximation predicts that altruism will not spread under certain conditions, for instance when kin competition exactly cancels out the kin selected benefits of altruism (Taylor, 1992a,b), it is very unlikely that increasing the intensity of selection will revert that result. This should also apply to cultural evolution where genetic drift is replaced by sampling effects occurring during social transmission (cultural drift, Cavalli-Sforza & Feldman, 1981). In Appendix C, we provide an example where the selection for cultural altruism vanishes under the weak selection approximation and show that this also holds under strong selection.

Furthermore, we expect that using the weak selection, approximation will also generate the conditions that are most favourable for the emergence of altruism under more complex demographic scenarios than the one presented here. Regardless of the underlying demographic process, whether it is Moran, Wright–Fisher or more realistic, the invasion of an altruistic mutant initially relies on genetic drift being able to build up relatedness in the face of negative selection, in order for a sufficient number of favourable mutant–mutant interactions to occur. As its name suggests, selection is at its weakest against genetic drift under the weak selection approximation. Therefore, like in our model, increasing selection pressure should generally disfavour the evolution of altruism, including in models with different sexes, various mating mechanisms or local demographic fluctuations with our parameter N now thought of as the local effective population size. Under isolation by distance, for example, simulations have shown that if the precise effect of selection strength is contingent on population structure, this structure becomes less important as selection strength increases and altruism is effectively counter selected (Szabó *et al.*, 2005; Segbreck *et al.*, 2011; Pinheiro *et al.*, 2012).

The model we have used is a deliberately simple one, capturing the broad features of selection on a trait resulting in fecundity and/or survival altruism. Yet, our results and previous work carried out for well-mixed populations can be used to infer on the robustness of the weak selection approximation for other altruistic traits, like dispersal, or more generally social traits under limited dispersal. It has been found that when only two alleles coding for a social behaviour segregate in a well-mixed population, and they mutate from one to another, selection intensity has no bearings on the conditions for one allele to be more frequent than the other (Antal *et al.*, 2009). This result and our model suggest that whatever the trait under scrutiny, if variability in the trait is due to the presence of only two alleles, then the weak selection approximation should still be efficient in predicting the direction of selection under limited dispersal when selection is strong – as long as migration is strong relative to selection or patch size is large.

However, it has also been shown that if there is a greater number of alleles present in a well-mixed population, then the ranking order of different alleles more often than not switches between the weak and strong selection modes (Wu *et al.*, 2013). This suggests that with more than two alleles segregating, the weak selection approximation may fail to predict the evolutionary outcome when selection is strong under limited dispersal. Nevertheless, as illustrated by the examples in Wu *et al.* (2013)'s study, we still expect that altruistic behaviour will be favoured under weak selection and disfavoured under strong selection in this case.

In summary, the weak selection approximation has been pivotal in understanding the selective pressure on altruistic traits, and more generally, on social behaviours.

We found that this approximation is robust to increases in selection intensity, especially when migration is greater than selection or when patches are large. In addition, we have highlighted that if it has any effect, increasing the selection intensity will tend to hinder the invasion of altruism. By simultaneously simplifying analysis and laying the best possible ground for the emergence of altruism, the weak selection approach to social evolution should not fail to detect whether altruism is selected under realistic environmental and demographic scenarios.

Acknowledgments

We would like to thank two anonymous reviewers for helpful comments. CM is funded by a grant to Laurent Keller, and LL by Swiss NSF grant PP00P3-123344.

References

Ajar, E. 2003. Analysis of disruptive selection in subdivided populations. *BMC Evol. Biol.* **3**: 22.

Akcay, E. & Van Cleve, J. 2012. Behavioral responses in structured populations pave the way to group optimality. *Am. Nat.* **179**: 257–269.

Alizon, S. & Taylor, P.D. 2008. Empty sites can promote altruistic behavior. *Evolution.* **62**: 1335–1344.

Anderson, S.P., de Palma, A. & Thisse, J.-F. 1992. *Discrete Choice Theory of Product Differentiation*. The MIT Press, Cambridge, MA.

Antal, T., Nowak, M.A. & Traulsen, A. 2009. Strategy abundance in games for arbitrary mutation rates. *J. Theor. Biol.* **257**: 340–344.

Aoki, K. 1982. A condition for group selection to prevail over counteracting individual selection. *Evolution.* **36**: 832–842.

Arbilly, M., Motro, U., Feldman, M.W. & Lotem, A. 2010. Co-evolution of learning complexity and social foraging strategies. *J. Theor. Biol.* **267**: 573–81.

van Baalen, M. & Rand, A. 1998. The unit of selection in viscous populations and the evolution of altruism. *J. Theor. Biol.* **193**: 631–648.

Bao, M. & Wild, G. 2012. Reproductive skew can provide a net advantage in both conditional and unconditional social interactions. *Theor. Popul. Biol.* **82**: 200–208.

Boyd, R. & Richerson, P.J. 1985. *Culture and the Evolutionary Process*. University of Chicago Press, Chicago.

Cavalli-Sforza, L. & Feldman, M.W. 1981. *Cultural Transmission and Evolution*. Princeton University Press, Princeton, NJ.

Chesson, P.L. 1984. Persistence of a Markovian population in a patchy environment. *Z. Wahrsh. Verwandte Gebeite* **66**: 97–107.

Eshel, I. 1972. On the neighbor effect and the evolution of altruistic traits. *Theor. Popul. Biol.* **11**: 258–277.

Ewens, W.J. 2004. *Mathematical Population Genetics*. Springer-Verlag, New York.

Frank, S.A. 1998. *Foundations of Social Evolution*. Princeton University Press, Princeton, NJ.

Fudenberg, D. & Levine, D.K. 1998. *Theory of Learning in Games*. MIT Press, Cambridge, MA.

Gandon, S. 1999. Kin competition, the cost of inbreeding and the evolution of dispersal. *J. Theor. Biol.* **200**: 345–364.

Gardner, A. 2010. Sex-biased dispersal of adults mediates the evolution of altruism among juveniles. *J. Theor. Biol.* **262**: 339–345.

Gardner, A. & West, S.A. 2006. Demography, altruism, and the benefits of budding. *J. Evol. Biol.* **19**: 1707–1716.

Grafen, A. 2007a. Detecting kin selection at work using inclusive fitness. *Proc. R. Soc. Lond. B Biol. Sci.* **274**: 713–719.

Grafen, A. 2007b. An inclusive fitness analysis of altruism on a cyclical network. *J. Evol. Biol.* **20**: 2278–2283.

Grimmett, G. & Stirzaker, D. 2001. *Probability and Random Processes*. Oxford University Press, Oxford.

Guillaume, F. & Perrin, N. 2006. Joint evolution of dispersal and inbreeding load. *Genetics.* **173**: 497–509.

Johnstone, R.A. & Cant, M.A. 2008. Sex differences in dispersal and the evolution of helping and harming. *Am. Nat.* **172**: 318–330.

Johnstone, R.A., Cant, M. & Field, J. 2012. Sex-biased dispersal, haplodiploidy and the evolution of helping in social insects. *Proc. R. Soc. B Biol. Sci.* **279**: 787–793.

Karlin, S. & Taylor, H.M. 1975. *A First Course in Stochastic Processes*. Academic Press, San Diego.

Kuijper, B. & Johnstone, R.A. 2012. How dispersal influences parent-offspring conflict over investment. *Behav. Ecol.* **23**: 898–906.

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

Lehmann, L. & Balloux, F. 2007. Natural selection on fecundity variance in subdivided populations: kin selection meets bet hedging. *Genetics.* **176**: 361–377.

Lehmann, L. & Perrin, N. 2002. Altruism, dispersal and phenotype-matching kin recognition. *Am. Nat.* **159**: 451–468.

Lehmann, L. & Rousset, F. 2010. How life-history and demography promote or inhibit the evolution of helping behaviors. *Philos. Trans. R. Soc. B* **365**: 2599–2617.

Lehmann, L., Perrin, N. & Rousset, F. 2006. Population demography and the evolution of helping behaviors. *Evolution.* **60**: 1137–1151.

Lehmann, L., Keller, L. & Sumpter, D. 2007a. The evolution of helping and harming on graphs: the return of the inclusive fitness effect. *J. Evol. Biol.* **20**: 2284–229.

Lehmann, L., Rousset, F., Roze, D. & Keller, L. 2007b. Strong reciprocity or strong ferocity? a population genetic view of the evolution of altruistic punishment. *Am. Nat.* **170**: 21–36.

Lehmann, L., Feldman, M.W. & Foster, K.R. 2008. Cultural transmission can inhibit the evolution of altruistic helping. *Am. Nat.* **172**: 12–24.

Leturque, H. & Rousset, F. 2002. Dispersal, kin competition, and the ideal free distribution in a spatially heterogeneous population. *Theor. Popul. Biol.* **62**: 169–180.

Lion, S. & van Baalen, M. 2007. Self-structuring in spatial evolutionary ecology. *Ecol. Lett.* **11**: 277–295.

Lion, S. & Gandon, S. 2010. Life history, habitat saturation and the evolution of fecundity and survival altruism. *Evolution.* **64**: 1594–606.

Luce, R.D. 1959. *Individual Choice Behavior*. Wiley, New York.

Massol, F., Calcagno, V. & Massol, J. 2009. The metapopulation fitness criterion: proof and perspectives. *Theor. Popul. Biol.* **75**: 183–200.

Metz, J. & Gyllenberg, M. 2001. How should we define fitness in structured metapopulation models? including an application to the calculation of evolutionary stable dispersal strategies. *Proc. R. Soc. Lond. B Biol. Sci.* **268**: 499–508.

Motro, U. 1982. Optimal rates of dispersal i. haploid populations. *Theor. Popul. Biol.* **21**: 394–411.

Nagylaki, T. 1992. *Introduction to population genetics*. Springer-Verlag, Heidelberg.

Ohtsuki, H. 2010. Evolutionary games in Wright's island model: kin selection meets evolutionary game theory. *Evolution.* **64**: 3344–53.

Ohtsuki, H. 2012. Does synergy rescue the evolution of cooperation? an analysis for homogeneous populations with non-overlapping generations. *J. Theor. Biol.* **307**: 20–28.

Ohtsuki, H., Hauert, C., Lieberman, E. & Nowak, M.A. 2006. A simple rule for the evolution of cooperation on graphs and social networks. *Nature.* **441**: 502–505.

Pinheiro, F.L., Santos, F.C. & Pacheco, J.M. 2012. How selection pressure changes the nature of social dilemmas in structured populations. *New J. Phys.* **14**: 073035.

Pen, I. 2000. Reproductive effort in viscous populations. *Evolution.* **54**: 293–297.

Queller, D.C. 1994. Genetic relatedness in viscous populations. *Evol. Ecol.* **8**: 70–73.

Rodrigues, A.M.M. & Gardner, A. 2012. Evolution of helping and harming in heterogeneous populations. *Evolution* **66**: 2065–2079.

Rogers, A.R. 1990. Group selection by selective emigration: the effects of migration and kin structure. *Am. Nat.* **135**: 398–413.

Rousset, F. 2004. *Genetic Structure and Selection in Subdivided Populations*. Princeton University Press, Princeton, NJ.

Rousset, F. & Roze, D. 2007. Constraints on the origin and maintenance of genetic kin recognition. *Evolution.* **61**: 2320–2330.

Roze, D. & Rousset, F. 2003. Selection and drift in subdivided populations: A straightforward method for deriving diffusion approximations and applications involving dominance, selfing and local extinctions. *Genetics.* **165**: 2153–2166.

Roze, D. & Rousset, F. 2004. The robustness of Hamilton's rule with inbreeding and dominance: Kin selection and fixation probabilities under partial sib mating. *Am. Nat.* **164**: 214–231.

Roze, D. & Rousset, F. 2008. Multilocus models in the infinite island model of population structure. *Theor. Popul. Biol.* **73**: 529–542.

Sandholm, W.H. 2011. *Population Games and Evolutionary Dynamics*. MIT Press, Cam bridge, MA.

Schlag, K.H. 1998. Why imitate, and if so, how? *J. Econ. Theory* **78**: 130–156.

Sozou, P.D. 2009. Individual and social discounting in a viscous population. *Proc. R. Soc. Lond. B Biol. Sci.* **276**: 2955–2962.

Szabó, G., Vukov, J. & Szolnoki, A. 2005. Phase diagrams for an evolutionary prisoner's dilemma game on two-dimensional lattices. *Phys. Rev. E.* **72**: 047107.

Taylor, P.D. 1992a. Altruism in viscous populations – an inclusive fitness model. *Evol. Ecol.* **6**: 352–356.

Taylor, P.D. 1992b. Inclusive fitness in a homogeneous environment. *Proc. R. Soc. Lond. B Biol. Sci.* **240**: 299–302.

Taylor, P.D. & Frank, S.A. 1996. How to make a kin selection model. *J. Theor. Biol.* **180**: 27–37.

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

Taylor, P.D. & Maciejewski, W. 2012. An inclusive fitness analysis of synergistic interactions in structured populations. *Proc. R. Soc. B* **279**: 4596–4603.

Van Dyken, J.D. 2010. The components of kin competition. *Evolution.* **64**: 2840–2854.

Van Dyken, J. & Wade, M.J. 2012. Origins of altruism diversity i: The diverse ecological roles of altruistic strategies and their evolutionary responses to local competition. *Evolution.* **66**: 2484–97.

Van Segbroeck, S., Santos, F.C., Lenaerts, T. & Pacheco, J.M. 2011. Selection pressure transforms the nature of social dilemmas in adaptive networks. *New J. Phys.* **13**: 013007.

Wild, G. & Fernandes A.D. 2009. Investing in the public good through conditional phenotypes of large effect. *J. Evol. Biol.* **22**: 927–941.

Wild, G. & Traulsen, A. 2007. The different limits of weak selection and the evolutionary dynamics of finite populations. *J. Theor. Biol.* **247**: 382–390.

Wright, S. 1931. Evolution in Mendelian populations. *Genetics.* **16**: 97–159.

Wu, B., Garcia, J., Hauert, C. & Traulsen, A. 2013. Extrapolating weak selection in evolutionary games. *PLoS Comput Biol* **9**: e1003381.

Yeh, A.Y.-C. & Gardner, A. 2012. A general ploidy model for the evolution of helping in viscous populations. *J. Theor. Biol.* **304**: 297–303.

Appendix A

Here, we show the equivalence between eqns (2) and (3). Using $w_i = \phi_i + e_i$, we have:

$$\begin{aligned}\rho &= \sum_{i=1}^{N}(w_i - 1)i\bar{t}_i \\ &= \sum_{i=1}^{N} e_i i\bar{t}_i + \sum_{i=1}^{N} \phi_i i\bar{t}_i - \sum_{i=1}^{N} i\bar{t}_i \qquad \text{(A-1)} \\ &= \sum_{i=1}^{N} e_i i\bar{t}_i - 1 + 1 + \sum_{i=1}^{N} \phi_i i\bar{t}_i - \sum_{i=1}^{N} i\bar{t}_i.\end{aligned}$$

Now, as $\sum_{i=1}^{N} \phi_i i\bar{t}_i$ is the expected number of successful offspring of the initial mutant that establish locally, $1 + \sum_{i=1}^{N} \phi_i i\bar{t}_i$ is the total size of the local mutant lineage (with 1 accounting for the initial mutant). But $\sum_{i=1}^{N} i\bar{t}_i$ is by definition the total size of the local mutant lineage, therefore $1 + \sum_{i=1}^{N} \phi_i i\bar{t}_i = \sum_{i=1}^{N} i\bar{t}_i$, and $\sum_{i=1}^{N}(w_i - 1)i\bar{t}_i = \sum_{i=1}^{N} e_i i\bar{t}_i - 1$, as required.

Appendix B

Here, we derive eqn (10). Substituting eqn (1) into eqn (5) and in turn into eqn (6) gives the necessary components to express individual fitness w_i (eqn 4) in terms of pay-off. Then, performing a first order Taylor series expansion around $\delta = 0$ on individual fitness, and substituting it into (8) gives for invasion fitness

$$\rho = \delta \sum_{i=1}^{N} \left(-C\frac{i}{N} + B\frac{i(i-1)}{N(N-1)} - (1-m)^2 (B-C)\frac{i^2}{N^2} \right) \bar{t}_i^{\circ} + O(\delta^2), \tag{B-1}$$

where the superscript $\circ$ in $\bar{t}_i^{\circ}$ is used to indicate that the sojourn time in state i is evaluated under neutrality ($\delta = 0$). This is sufficient because any effect of selection on $\bar{t}_i$ will only result in second order effects of selection on ρ. The neutral sojourn time $\bar{t}_i^{\circ}$ is found using eqn (8), but with neutral birth and death rates $b_i = (N - i)(1 - m)i/N^2$ and $d_i = i[N - i(1 - m)]/N^2$. With this, we find that:

$$\begin{aligned} &\sum_{i=1}^{N} \frac{mi}{N}\bar{t}_i^{\circ} = 1 \\ &\sum_{i=1}^{N} \frac{mi(i-1)}{N(N-1)}\bar{t}_i^{\circ} = R, \end{aligned} \tag{B-2}$$

where R is the probability two distinct individuals randomly sampled on the same patch carry an homologous gene identical-by-descent, and R satisfies the recursion:

$$R = (1-m)\left(\frac{1}{N} + \frac{N-1}{N}R\right). \tag{B-3}$$

Equation (B-2) gives the analogue of the moments obtained under the Wright–Fisher process (Ajar, 2003, eqns 34–36) for the Moran process. Using eqn (B-2), along with the identity $i^2/N^2 = i/N^2 + (N-1)i(i-1)/(N^2(N-1))$, gives for eqn (B-1):

$$\rho = \frac{\delta}{m}\left[-C + BR - (1-m)^2(B-C)\left(\frac{1}{N} + \frac{N-1}{N}R\right)\right], \tag{B-4}$$

which after substituting for R using eqn (B-3) produces eqn (10) of the main text. Eqn (B-4) takes the same form as the usual selection gradient computed with the quasi-equilibrium approach (e.g. Ajar 2003; Rousset, 2004).

Appendix C

Here, we analyse the spread of an altruistic mutant that is transmitted culturally by social learning (Cavalli-Sforza & Feldman, 1981; Boyd & Richerson, 1985). As before, the population is made up of an infinite number of patches, each of constant size N. The life cycle can be decomposed into two steps. In the first step, individuals interact socially with each other within patches and receive pay-offs from these interactions. There are two cultural strategies, or variants, in the population, a resident a and an altruistic mutant A. As before, an individual that adopts the mutant strategy sees its pay-off decrease by a cost C while its neighbours receive a benefit B. The resident, meanwhile, pays no cost. Therefore, the pay-offs received by a mutant and a resident in a patch with i mutants are also given by eqn (1).

In the second step, an individual from each patch updates its strategy by social learning. Namely, a randomly sampled individual within each patch revises its strategy according to its pay-off and the pay-offs of other individuals. Strategy revision is performed by social learning through pay-off-based imitation (Boyd & Richerson, 1985; Schlag, 1998). With probability $1 - m$, an individual chooses an individual at random, including itself, in its own patch and adopts the exemplar's variant as a function of exemplar's pay-off. With complementary probability m, the individual chooses an exemplar individual at random in another patch and again adopts its variant as a function of the exemplar's pay-off.

Choice and pay-off are related by a logit choice rule (e.g. Sandholm, 2011), whereby, conditional on revising its strategy locally, an individual in a patch with i mutants adopts the mutant variant with probability:

$$\frac{\frac{i}{N}\exp\left(\delta\pi_{\mathrm{A},i}\right)}{\frac{i}{N}\exp\left(\delta\pi_{\mathrm{A},i}\right) + \frac{N-i}{N}\exp\left(\delta\pi_{\mathrm{a},i}\right)}, \tag{C-1}$$

and with complementary probability, it adopts the resident variant. Hence, when an individual updates its action, it chooses an exemplar mutant with probability i/N and imitates him with a probability proportional to the weight $\exp(\delta\pi_{\mathrm{A},i})$. Similarly, it chooses an exemplar resident with probability $(N - i)/N$ and imitates him with a probability proportional to the weight $\exp(\delta\pi_{\mathrm{a},i})$. The logit choice rule is a standard way to model selective choice among different alternatives (e.g. Luce, 1959; Anderson *et al.*, 1992; Fudenberg & Levine, 1998; Arbilly *et al.*, 2010). The parameter δ in eqn (C-1) measures the sensitivity of choice to pay-off. When $\delta \to 0$, choice does not depend on pay-off and is proportional on the frequency of variants, while when δ increases, actions with higher pay-offs are more likely to be chosen. Hence, δ is a measure of the intensity of selection for cultural transmission.

Invasion fitness

To derive the invasion fitness of the mutant cultural variant, we note that the assumptions behind our development of ρ (eqn 3) do not restrict the variants to being genetically determined and that we can in fact use it for variants that are culturally transmitted. Under social learning, fitness w_i is the expected number of individuals that will learn the cultural variant from a focal mutant in a group with i mutants. It can again be decomposed into two terms, $w_i = \phi_i + e_i$, where ϕ_i is the expected number of individuals in the focal patch that adopt the variant from the focal mutant and e_i is the expected number of individuals from other patches that adopt the focal's variant. With our imitative logit choice rule, the components of fitness w_i are now

$$\begin{aligned} \phi_i &= \frac{N-1}{N} + \frac{1}{N}\frac{(1-m)\exp(\delta\pi_{\mathrm{A},i})}{\frac{i}{N}\exp(\delta\pi_{\mathrm{A},i}) + \frac{N-i}{N}\exp(\delta\pi_{\mathrm{a},i})} \\ e_i &= \frac{m}{N}\frac{\exp(\delta\pi_{\mathrm{A},i})}{\exp(\delta\pi_{\mathrm{a},0})}. \end{aligned} \tag{C-2}$$

To compute the invasion condition ρ (eqn 3), we also need the sojourn time $\bar{t}_i$ in state i (eqn 8). This depends on the birth and death probabilities associated with our imitative logit choice rule, which are:

$$\begin{aligned} b_i &= \frac{(N-i)}{N}\frac{(1-m)\frac{i}{N}\exp(\delta\pi_{\mathrm{A},i})}{\frac{i}{N}\exp(\delta\pi_{\mathrm{A},i}) + \frac{N-i}{N}\exp(\delta\pi_{\mathrm{a},i})} \\ d_i &= \frac{i}{N}\left[(1-m)\frac{\frac{N-i}{N}\exp(\delta\pi_{\mathrm{a},i})}{\frac{i}{N}\exp(\delta\pi_{\mathrm{A},i}) + \frac{N-i}{N}\exp(\delta\pi_{\mathrm{a},i})} + m\right]. \end{aligned} \tag{C-3}$$

Weak selection

To obtain the weak selection approximation of the invasion fitness of the mutant, eqns (1) and (C-2)–(C-3) are substituted into the fitness w_i (eqn 4), which is Taylor expanded to the first order around $\delta = 0$ and then substituted into eqn (3), giving

$$\rho = \delta\sum_{i=1}^{N}\left(-C\frac{i}{N} + B\frac{i(i-1)}{N(N-1)} - (1-m)(B-C)\frac{i^2}{N^2}\right)\bar{t}_i^{\circ} + O(\delta^2). \tag{C-4}$$

Using eqn (B-2), we find that the invasion fitness reads:

$$\rho = \frac{\delta}{m}\left[-C + BR - (1-m)(B-C)\left(\frac{1}{N} + \frac{N-1}{N}R\right)\right]. \tag{C-5}$$

Then, we note that in the absence of selection, the relatedness between two different individuals randomly sampled from the same patch is updated when an individual revises its strategy by copying a patch mate (with probability $1 - m$), whom it copies with probability $1/N$. So, eqn (B-3) also applies for the imitative logit choice rule. Substituting eqn (B-3) into eqn (C-5) eventually gives:

$$\rho = -\frac{\delta(1-R)C}{m}. \tag{C-6}$$

Hence, unless the trait results in a direct pay-off benefit to the focal ($C < 0$), the mutant cannot evolve under social learning when the revision protocol follows the imitative logit choice rule and there is only one individual updating its trait per unit time step. This result had been obtained in a previous study (Lehmann *et al.*, 2008, eqn 23–24) that studied the same model as here, but with the weights of the imitative protocol given directly by the pay-offs, rather than the exponential of the pay-offs (eqn C-1). In both cases, the updating process is the same when $\delta \to 0$, leading to the same invasion condition.

Strong selection

As with genetic transmission, we start by considering the case with $N = 2$ as this is fully tractable for arbitrary levels of selection. To calculate the invasion fitness in this case, it turns out to be more practical to use expression eqn (2). We substitute eqns (1) and (C-2)–(C-3), eqn (8) into eqn (2) with $N = 2$ to give:

$$\rho = -\frac{\left(1 - e^{-C\delta}\right)\left(e^{B\delta}\left(e^{C\delta} - m + 1\right) + m\right)}{e^{\delta(B+C)} + m}. \tag{C-7}$$

For $\delta > 0$, the numerator of eqn (C-7) is positive for any values of B when altruism is costly $C > 0$. Therefore, as with the weak selection approximation (eqn C-6), we find that regardless of whether effects on patch neighbours are positive (helping behaviour, $B > 0$) or negative (harming behaviour, $B < 0$), a mutant that result in a reduction of the material pay-off of a focal individual is always evolutionary unstable. Numerical exploration for arbitrary patch size N shows the same qualitative trend as predicted by eqn (C-7). Namely, a mutant with $C > 0$ and that either helps or harms patch neighbours is selected against.

Received 13 March 2014; revised 29 May 2014; accepted 7 July 2014