PROCEEDINGS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

An inclusive fitness analysis of synergistic interactions in structured populations

Peter Taylor and Wes Maciejewski

Proc. R. Soc. B 2012 **279**, doi: 10.1098/rspb.2012.1408 first published online 12 September 2012

PROCEEDINGS OF THE ROYAL SOCIETY B

CrossMark click for updates

Proc. R. Soc. B (2012) **279**, 4596–4603
doi:10.1098/rspb.2012.1408
Published online 12 September 2012

An inclusive fitness analysis of synergistic interactions in structured populations

Peter Taylor* and Wes Maciejewski

Department of Mathematics and Statistics, Queen's University, Kingston, Ontario, Canada K7L 3N6

We study the evolution of a pair of competing behavioural alleles in a structured population when there are non-additive or 'synergistic' fitness effects. Under a form of weak selection and with a simple symmetry condition between a pair of competing alleles, Tarnita *et al.* provide a surprisingly simple condition for one allele to dominate the other. Their condition can be obtained from an analysis of a corresponding simpler model in which fitness effects are additive. Their result uses an average measure of selective advantage where the average is taken over the long-term—that is, over all possible allele frequencies—and this precludes consideration of any frequency dependence the allelic fitness might exhibit. However, in a considerable body of work with non-additive fitness effects—for example, hawk–dove and prisoner's dilemma games—frequency dependence plays an essential role in the establishment of conditions for a stable allele-frequency equilibrium. Here, we present a frequency-dependent generalization of their result that provides an expression for allelic fitness at any given allele frequency p. We use an inclusive fitness approach and provide two examples for an infinite structured population. We illustrate our results with an analysis of the hawk–dove game.

Keywords: evolutionary game theory; non-additive games; relatedness; allele frequency; Price equation; frequency dependence

1. INTRODUCTION

An enormous body of significant work constructs analytical models for the genetical evolution of social behaviour. The key relationship here is the dependence of focal fitness on the behaviour (phenotypic value) of a number of interactants. These behaviours are typically correlated with individual genotypic values, and the resulting connection between fitness and genotype allows us to get hold of the manner in which selection changes the frequency of alleles coding for alternative behaviours. The central tool in this analysis is the covariance formula of Price [1]. It requires us to calculate the covariance between focal fitness and focal genotype, and the dependence of the former on the genotypic values of neighbouring individuals reduces the problem to one of calculating covariances between neighbouring genotypes (or between expressions involving neighbouring genotypes) and the focal genotype.

In building models of genetic change, we make explicit assumptions about how focal fitness depends on local genotypic values. The simplest models we work with are linear; that is, fitness effects among interactants are assumed to be additive—if Y and Z both interact with X, the effect on the fitness of X is the sum of the individual effects of Y and Z. If X, Y and Z have genotypic values x, y and z, then the Price equation requires calculation of the covariances $\mathrm{cov}(x, y)$ and $\mathrm{cov}(x, z)$. Much of the work on cooperation and altruism makes this assumption. Other models use nonlinear functions and considerable attention has been paid to quadratic expressions [2]. These arise naturally in haploid models in which genotypic values are either 0 or 1. In this case, there are four possibilities for the genotypic pair (y, z), and a quadratic function is sufficient to describe all possible joint genotypic effects on the fitness of X of the interactions of Y and Z with X. In this case, the Price equation may involve covariances $\mathrm{cov}(x, xy)$ and $\mathrm{cov}(x, yz)$, and these can be more difficult to calculate.

In fact, it has been pointed out many times that, at the end of the day, all selection cares about is the multivariable *linear* regression of fitness on genotypic values, as selection is only able to act on additive genetic effects. This idea was perhaps first given prominence in Falconer's [3] classic book on quantitative genetics in which the breeding value of an individual was defined as 'the sum of the average of the effects of the two alleles present'. That is for two genes *within* individuals, but the same conditions apply to genes *between* individuals. More recently, this central idea has been amplified by Queller [4,5], Frank [6] and Gardner *et al.* [7]. To rephrase the idea, suppose that the fitness W of X depends in any manner we wish on the genotypic values of Y and Z. Then, use the linear regression of W on y and z to replace the function $W(y, z)$ by the linear function $W^*(y, z)$. Then, for calculating the genetic effects of selection (say on allele frequency), we can replace W by the linear function W^*. Having said that, we still might be faced with a formidable calculation in a particular population model, as the regression coefficients can be difficult to calculate.

Recently, Tarnita *et al.* [8] have shown that in a structured population with a simple symmetry condition between a pair of competing alleles, quadratic ('synergistic') effects can be handled with standard linear methods. This is an interesting result, because it not only simplifies our calculations, but also provides an interesting and unexpectedly simple mathematical form for allelic fitness. However, their calculation of fitness uses a long-term average measure of allele frequency, preventing them from analysing frequency-dependent behaviour, which is a critical component of the analysis in non-additive games such

* Author for correspondence (peter.taylor@queensu.ca).

Electronic supplementary material is available at http://dx.doi.org/10.1098/rspb.2012.1408 or via http://rspb.royalsocietypublishing.org.

Received 20 June 2012
Accepted 21 August 2012

as hawk–dove and prisoner's dilemma. Our purpose here is to generalize their result to obtain a frequency-dependent condition for allelic fitness. We use an inclusive fitness approach in both a finite and an infinite structured population model, and illustrate our results with an analysis of the hawk–dove game played in Wright's [9] island model with structured demes.

2. THE INCLUSIVE FITNESS MODEL

(a) *Demographics*

We work with a structured population represented as a graph, a set of nodes each occupied by a single asexual haploid breeder, together with edges between certain pairs of nodes. Nodes joined by an edge are called *adjacent*, and the set of nodes adjacent to a given node is called its *neighbourhood*. Each edge carries two different pieces of information: first, the probability that the two breeders will engage in a 'primary' interaction (e.g. play a game [10]), and second, the probability that an offspring from one node will displace the breeder at the other. For simplicity, we assume that offspring dispersal between nodes i and j is symmetric. We work with two different population update processes: a non-overlapping generations model (Wright–Fisher) and a continuous-time process (Moran) with either birth–death or death–birth updating [11,12].

(b) *A two-allele model*

We suppose that there are two alleles A and B assorting at a fixed locus and let individual genotypic value be 1 for A and 0 for B. We let the genotypic value x of an individual X be the frequency of A in its genotype. In order to counter the effects of genetic drift towards homozygosity, we find it convenient to make different assumptions in a finite and infinite population. In the former, we assume that genetic mutation acts from B to A at rate μ and from A to B at rate ν, and in the latter, we use long-range migration and send offspring to a 'distant' node at which it shares no common ancestor with the local breeders. Ohtsuki [13] formalizes this with 'a node at infinity'. Thus, in our infinite population model, we ignore genetic mutation, for which the effects are typically orders of magnitude smaller than those of migration. We let the *allele frequency* p be the average frequency of A, where in an infinite population, this average is taken over *space* (over the whole population), and in a finite population, it is taken over *time* (over all states of the population as the frequency of A wanders between 0 and 1). Then, in a neutral (no selection) finite population, the allele frequency will be $p = \mu/(\mu + \nu)$, and in an infinite population model, it will be the allele frequency of the long-range immigrants.

(c) *Primary and secondary fitness effects*

We assume that individuals engage in pairwise interactions (for example, in the playing of a two-person matrix game [10]) with behaviours or strategies determined by the alleles A and B. The primary fitness effect $w(x, y)$ on an individual X with genotype x whose partner Y has genotype y, has four possible values (for x, $y = 0, 1$), which we denote as a, b, c and d so that

$$\begin{aligned} w(x,\ y) &= axy + bx(1-y) + c(1-x)y + d(1-x)(1-y) \\ &= d + (b-d)x + (c-d)y + (a+d-b-c)xy, \end{aligned} \tag{2.1}$$

where a, b, c and d are assumed to be small. This case has been much studied in evolutionary game theory [10,14,15] in which the interactants play a game with pay-off matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

However, in a structured population, these primary interactions will typically have 'secondary' fitness effects [16,17]—we use the terminology of West & Gardner [18], which must also be accounted for, and the overall fitness effect on a focal individual X will combine the primary effects with the secondary effects from primary interactions in the neighbourhood. All of these will have the form of equation (2.1), with various genotypes in the role of x and y, and coefficients a, b, c and d (or a fixed multiple of these). For example, if Y interacts with Z, obtaining a fecundity increase of $w(y, z)$, and the resulting increased competition for breeding spots increases the mortality of X with probability r, the interaction has a secondary effect on X of $-rw(y, z)$ and this will be added to the fitness of X. In addition, in a heterogeneous population (with different types of nodes—for example, graphs that are not regular, or nodes with different offspring dispersal patterns), different terms will have to be weighted by appropriate reproductive values (RVs). When all such effects have been accounted for, the overall fitness effect on X will be a linear combination of terms of the form (2.1), and thus will have the form

$$W_{\mathrm{X}} = f_1 a + f_2 b - f_3 c - f_4 d, \tag{2.2}$$

where the f_i are quadratic polynomials in breeder genotypic values. Note that it is convenient for our purpose to put a negative sign in front of the last two terms (which represent pay-offs to players with allele B).

The primary fitness effect (2.1) is called 'additive' when $a + d = b + c$ and the synergistic xy term disappears. In this case, the secondary fitness effects will also be additive, and hence this will be the case for the overall fitness effect W_{X} in (2.2). The word 'additive' comes from the observation that when $a + d = b + c$, it is possible to assign a fitness effect to each of the two alleles separately, so that the effect of the alleles acting together will be the sum of the separate effects. This relationship is more transparent with the use of the matrix form found in equation (3.2).

(d) *Allele frequency change*

A general objective in evolutionary modelling is to get an appropriate measure of the selective advantage of a particular allele. Here, we will take this to be the *selective* rate of change in the frequency of A—that is, we ignore effects of mutation, migration, random sampling, and so on. Our basis for the calculation will be the classic formula of Price [1], but we will work with that using an inclusive fitness approach. Roughly speaking, the inclusive fitness effect W_{IF} of an allele A is (proportional to) the rate of increase in the population-wide frequency of A. But it is important to be precise about how the measurement is made.

First note that the rate of change of allele frequency will depend on the configuration of the alleles A and B, and we make an assumption that this is at long-term neutral equilibrium. This refers to a steady probabilistic state attained by the alleles at *neutrality*; that is, with allele A having the same behaviour as B ($a = b = c = d$).

We imagine that there is a switch governing the effect of A. We begin with the switch off, and let the population attain equilibrium, and then we turn it on, so that A has an altered behaviour with a small fitness effect, and measure the rate of allele frequency change. However, we make this measurement in a different way in a finite and infinite population. We will discuss both cases here, but we provide only the infinite population equations, leaving the rather more technical (but completely analogous) finite population equations for the electronic supplementary material, §D.

(e) *An infinite population*

Here, the long-range migration (and the infinite size) will guarantee that at equilibrium every possible 'local' configuration of alleles will be represented in the population at the correct frequency. With the population at that equilibrium, we turn the A switch on, and use Price's [1] formula to measure the selective rate of change of the population-wide allele frequency:

$$\frac{\mathrm{d}\bar{x}}{\mathrm{d}t} = \mathrm{cov}(x, W_{\mathrm{X}}). \tag{2.3}$$

A more general version of Price's equation has a second term to account for non-selective forces of change (such as mutation causing offspring genotype to differ from parental genotype), but here we are measuring only the changes due to selection. Equation (2.3) uses the notation belonging to a continuous-time model such as a Moran process. In discrete time models, in which the derivative would be replaced by $\Delta\bar{x}$, the change over a single time step or generation, we must divide the right side by average fitness, though we can avoid that by normalizing so that mean fitness is one.

(f) *A finite population*

In this case, even at neutrality, random sampling will cause the population allele frequency $\bar{x}$ to drift and oscillate between fixation and non-fixation states. Selection, of course, can act only in the unfixed states and the covariance in Price's formula (2.3) will depend on the state. The convention we adopt [19] is to let the overall allele frequency change be the average of the state-dependent changes given by equation (2.3), where different population states are weighted by their long-term neutral frequency. The resulting generalization of Price's equation is formulated in the electronic supplementary material.

(g) *Inclusive fitness*

The inclusive fitness effect of the allele A, as originally defined by Hamilton [20], is formulated as follows. We take a focal A individual in the population at neutral equilibrium, and turn the A switch on. Then, we record the changes in personal fitness experienced by all individuals whose fitness is affected by the focal behaviour (primarily or secondarily), and the inclusive fitness effect of A is the sum of these, each such change weighted by the relatedness of the focal individual to the affected individual:

$$W_{\mathrm{IF}} = \sum\nolimits_i R_i \Delta w_i. \tag{2.4}$$

Typically, one of the summands is the focal individual itself as its behaviour is expected to affect its own fitness, and in this case, the relatedness to itself is, by definition, $R = 1$. We remark that in a heterogeneous population, those fitnesses w_i must incorporate RV. For example, in the Moran model, fitness is the fecundity rate minus the mortality rate, and while the mortality of X is always weighted by the RV of X, in calculating the fecundity of X, each offspring must be weighted by the RV belonging to its destination node.

The Price equation appeared in 1970, though the covariance form had already appeared in a paper by Robertson [21], but it took some years after that for the mathematical relationship between Hamilton's inclusive fitness and the Price equation to be properly formulated. What the Price equation gives us is what Hamilton [22] originally identified as 'neighbour-modulated' fitness, and is usually now more simply referred to as personal fitness. Unlike inclusive fitness, which takes a focal A *actor* and tabulates the different effects of its behaviour on all others, personal fitness takes a focal A *recipient* and tabulates the different effects of the A behaviour in the population on its fitness. A considerable body of work [2,19,23–25,26] (and references in these) has established that under quite general conditions, these 'inverse' approaches give us the same result and that the inclusive fitness effect can be written as

$$W_{\mathrm{IF}} = \frac{\mathrm{cov}(x, W_{\mathrm{X}})}{\mathrm{cov}(x, x)}. \tag{2.5}$$

A comparison of (2.5) with (2.3) allows us to use W_{IF} as our measure of the selective advantage of A.

From equation (2.2), the inclusive fitness effect (2.5) can be written as

$$W_{\mathrm{IF}} = \frac{\mathrm{cov}(x,f_1)}{\mathrm{cov}(x,x)}a + \frac{\mathrm{cov}(x,f_2)}{\mathrm{cov}(x,x)}b - \frac{\mathrm{cov}(x,f_3)}{\mathrm{cov}(x,x)}c - \frac{\mathrm{cov}(x,f_4)}{\mathrm{cov}(x,x)}d. \tag{2.6}$$

Using the fact established earlier that the f_i are quadratic polynomials in breeder genotypic values, the quotients of covariances will be linear expressions in terms such as

$$R_{\mathrm{X}\to\mathrm{Y}} = \frac{\mathrm{cov}(x,y)}{\mathrm{cov}(x,x)} \quad \text{and} \quad R_{\mathrm{X}\to\mathrm{YZ}} = \frac{\mathrm{cov}(x,yz)}{\mathrm{cov}(x,x)}. \tag{2.7}$$

The first of these is the standard coefficient of relatedness of X to Y, and in a structured population, this can be calculated by recursive techniques, which typically often use the notions of genetic 'identity by descent' (IBD) or 'identity in state' (IIS). This coefficient is known to be independent of p [2,19,23,26]. The second of these is a 'generalized' relatedness coefficient, which can also be calculated by recursive techniques using the notions of IBD or IIS, but the calculations are more complex (see [27] and [13]) and the coefficients generally depend on the population allele frequency p. However, we will use the fact that $R_{\mathrm{X}\to\mathrm{YZ}}$ is linear in p. The argument for this is somewhat technical and is presented in the electronic supplementary material, §A. The finite population analogues of equations (2.5)–(2.7) are presented in the electronic supplementary material, §D.

3. A FREQUENCY-DEPENDENT VERSION OF THE RESULT OF TARNITA *ET AL.*

Our purpose here is to provide a frequency-dependent version of a result of Tarnita *et al.* [8] for the selective advantage of an allele A. Consider a haploid structured population with two alleles A and B playing the matrix game $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Tarnita *et al.* [8] show that in a finite population with a symmetric mutation rate and weak selection (a, b, c and d small), there exists a parameter σ dependent on the population structure but independent of the pay-offs a, b, c and d for which the condition that allele A be selectively favoured over B can be written as

$$\sigma a + b > c + \sigma d. \tag{3.1}$$

For our generalization of this result, we work with a standard model for altruism [2] in which an 'altruist' X gives benefit B to its partner Y at cost C, and gets a synergistic bonus D if Y is also an altruist. The pay-off matrix in this case is

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} B - C + D & -C \\ B & 0 \end{bmatrix} \tag{3.2}$$

and for the remainder of the paper this is the notation we use. We note that our version of the matrix is actually general, as we can subtract any constant from all entries of the matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ without affecting the analysis (as this gives the same fitness bonus to all players). With this terminology, equation (3.1) can be written as

$$(\sigma - 1)B - (\sigma + 1)C + \sigma D > 0. \tag{3.3}$$

Writing this in the abstract form $\beta B - \gamma C + \delta D > 0$, this tells us that $\delta = (\beta + \gamma)/2$, and hence equation (3.1) can be written in the form

$$\beta B - \gamma C + \left(\frac{\beta + \gamma}{2}\right) D > 0. \tag{3.4}$$

This is a striking result, as it tells us that synergistic interactions can be handled with additive models. Indeed, the coefficients are independent of the pay-offs, and hence β and γ can be determined from an analysis of the additive game $\begin{bmatrix} B - C & -C \\ B & 0 \end{bmatrix}$ and will therefore only involve the standard relatedness coefficients $R_{X \to Y}$. The result is also at first surprising, as it is known that with synergistic interactions, the conditions for allele frequency increase generally depend on allele frequency p [2], but there is no frequency dependence in β or γ and hence there is none in equation (3.4). This confusion is resolved with the realization that Tarnita *et al.* [8] work with the standard finite population measure, described earlier, which takes p to be the long-term average allele frequency, and hence their assumption of a symmetric mutation rate will set p to equal $\frac{1}{2}$, and their condition cannot possibly be p-dependent.

Our main result is the following. Suppose that an infinite or a finite structured haploid asexual population has average allele frequency p, and has two alleles A and B at a single locus playing the evolutionary game $\begin{bmatrix} B - C + D & -C \\ B & 0 \end{bmatrix}$ with weak selective effects (small pay-offs). Then the inclusive fitness effect of A can be written in the form

$$W_{IF} = \beta B - \gamma C + \left[\frac{\beta + \gamma}{2} + \left(p - \frac{1}{2}\right)\alpha\right] D, \tag{3.5}$$

where α, β and γ are independent of both the matrix pay-offs and the allele frequency p. As mentioned earlier, if $D = 0$, there are no synergistic effects and W_{IF} involves only the linear relatedness coefficients $R_{X \to Y}$, and the coefficients β and γ are linear combinations of these coefficients. The coefficient α typically requires the calculation of the more complex higher-order coefficients $R_{X \to YZ}$. The analysis leading to equation (3.5) is found in the electronic supplementary material, §B.

There are a number of special cases of equation (3.5) that are worth highlighting.

(a) *Symmetric allele frequency*

If the allele frequency p equals $\frac{1}{2}$, the α term in equation (3.5) vanishes and we obtain equation (3.4). In an infinite population, this occurs when long-range immigrants have allele frequency $\frac{1}{2}$, and in a finite population, it occurs with symmetric mutation as assumed in [8]. In this case, even with synergistic pay-offs ($D \neq 0$), the inclusive fitness effect involves only the standard relatedness coefficients $R_{X \to Y}$.

(b) *Small rates of long-range migration or mutation*

It turns out that α has the same order as the rate of long-range migration (infinite population) or the mutation rate (finite population), so that if these are negligible, the α term can be neglected, and equation (3.4) provides a good approximation to equation (3.5). The argument for this is found in the electronic supplementary material, §C. The significance of this for finite populations needs to be emphasized. When mutation is rare, the simpler equation (3.4) can be used, even without the assumption of symmetric mutation.

(c) *Transitive population structures*

The population is called transitive [28] if, given two nodes i and j, there is a bijection of the node set that preserves the edges and maps i to j. Roughly speaking, the population 'looks the same' from every node. Note that in preserving the edges, the information they carry (interaction and dispersal probabilities) must also be preserved. Transitivity allows us to get a good hold on the relationship between the primary and the secondary fitness effects, giving us special forms for the inclusive fitness effect in the additive ($D = 0$) case [12,16,17,28,29].

Consider a transitive population with either a Wright–Fisher demography (non-overlapping generations) or a Moran process with B–D updating (continuous reproduction). If the population is infinite, it turns out that the inclusive fitness benefits of a primary fecundity gift to a relative are exactly counterbalanced by the resulting competitive effects of the offspring produced, and, as a result, W_{IF} is independent of B, and from equation (3.5), the inclusive fitness effect will have the general form

$$W_{IF} \sim -C + \left[\frac{1}{2} + \left(p - \frac{1}{2}\right)\alpha\right] D, \tag{3.6}$$

where we use '$\sim$' to denote 'up to a multiplicative constant'.

If the population is finite of size n, the inclusive fitness effect of an additive ($D = 0$) primary interaction has the form

$$W_{\mathrm{IF}} \sim -\frac{B}{n-1} - C, \tag{3.7}$$

[12,17,28,29]. Combining this with equation (3.5) gives us the general form

$$W_{\mathrm{IF}} \sim -\frac{B}{n-1} - C + \left[\frac{n-2}{2(n-1)} + \left(p - \frac{1}{2}\right)\alpha\right]D, \tag{3.8}$$

and, again, the α-term can be ignored if mutation is rare.

(d) *No local secondary effects*

Suppose that all secondary effects of a primary interaction are experienced by individuals chosen at random from the population. Such individuals, by definition, have zero relatedness with the primary interactants, and in this case $w(x, y)$ in equation (2.1) provides the entire local fitness effect of the interaction, and the inclusive fitness effect can be written as

$$W_{\mathrm{IF}} = \frac{\mathrm{cov}(x, w)}{\mathrm{cov}(x, x)} = -C + R_{\mathrm{X}\to\mathrm{Y}}B + R_{\mathrm{X}\to\mathrm{XY}}D. \tag{3.9}$$

A standard result is that $R_{\mathrm{X}\to\mathrm{XY}} = p + (1-p)R_{\mathrm{X}\to\mathrm{Y}}$. Its derivation has appeared in a number of places [2,7,30] and is provided at the end of the electronic supplementary material, §A. With this, we have

$$\begin{aligned} W_{\mathrm{IF}} &= -C + R_{\mathrm{X}\to\mathrm{Y}}B + (p + qR_{\mathrm{X}\to\mathrm{Y}})D \\ &= -C + R_{\mathrm{X}\to\mathrm{Y}}B + \left(\frac{1 + R_{\mathrm{X}\to\mathrm{Y}}}{2} + \left(p - \frac{1}{2}\right)(1 - R_{\mathrm{X}\to\mathrm{Y}})\right)D. \end{aligned} \tag{3.10}$$

Equation (3.10) still displays frequency dependence, but involves only the standard coefficients of relatedness. It is also found in eqn (3.3) of a paper by Gardner *et al.* [7].

4. TWO EXAMPLES OF AN INFINITE ISLAND MODEL

Our main focus here has been on the infinite population case, and, without providing the calculations (which are found elsewhere), we provide two examples of equation (3.5) in an infinite structured population. The two examples have a number of common features. Both are set in an infinite island model with identical demes [9]. In both examples, individuals are haploid and asexual, and we use a Moran process with a 'birth–death' protocol [11]; thus, generations are overlapping and continuous, and the time between successive offspring births for each individual is exponentially distributed. Offspring always replace an existing breeder in their native deme with probability h, and in a distant deme with probability $1 - h$. Primary interactants play the game with matrix $\begin{bmatrix} B - C + D & -C \\ B & 0 \end{bmatrix}$ and pay-offs provide a small fecundity increment; that is, they affect the reproductive rate. In each example, we will specify (i) the primary partners for each player and (ii) the offspring dispersal probabilities. Finally, both examples have a transitive structure, and therefore we expect to obtain the form of equation (3.6).

(a) *Example 1. Infinite island model with random mixing demes of size n*

Breeders choose partners for the matrix game at random from among their $n - 1$ deme-mates. Offspring remain on their home deme with probability h, and in that case, they replace a random breeder on the deme including the parent. Under these assumptions, we have:

$$W_{\mathrm{IF}} \sim -C + \left(\frac{1}{2} + \left(p - \frac{1}{2}\right)\left(\frac{n(1-h)}{n + 2h - nh}\right)\right)D. \tag{4.1}$$

This example has been studied by Ohtsuki [13], but with a Wright–Fisher process (non-overlapping generations).

Note the factor of $(1 - h)$ in the $(p - \frac{1}{2})$ term. This is expected from our discussion above of small long-range migration. If we let h approach 1, this rate approaches zero and equation (4.1) is approximated by

$$W_{\mathrm{IF}} \sim -C + \frac{1}{2}D. \tag{4.2}$$

Of course, in the limit, as h approaches 1, we might expect each deme to behave as a randomly mixed *finite* population of size n, and thus we might expect equation (4.2) to hold approximately for such a population with a small mutation rate; but it turns out that this is not the case, a result that emphasizes a significant difference between migration and mutation. In a transitive finite population of size n, under a Moran process with birth–death updating and small mutation rate, the inclusive fitness effect is given by equation (3.8), and because the α term is negligible, this is

$$W_{\mathrm{IF}} \sim -\frac{B}{n-1} - C + \frac{n-2}{2(n-1)}D, \tag{4.3}$$

and is different from equation (4.2)—though, as we might expect, they converge as n gets large.

(b) *Example 2. Island model with structured demes of size 4*

The population consists of an infinite number of demes of size 4, each consisting of two dyadic patches. Primary interactions are between patchmates, and thus each breeder has only one partner. Offspring who stay on their home deme (probability h) displace the parent's patchmate (but not the parent itself) with probability $\frac{1}{2}$ and displace each breeder in the opposite patch with probability $\frac{1}{4}$ (figure 1). Under these assumptions, we have

$$\begin{aligned} W_{\mathrm{IF}} \sim &-2(2+h)C + \Bigg((2+h) \\ &+ \left(p - \frac{1}{2}\right)\frac{(1-h)(12 + 2h + h^2)}{3 - h}\Bigg)D. \end{aligned} \tag{4.4}$$

Again, there are three things to note. First, equation (4.4) has the form of equation (3.5); second, the population structure is homogeneous, and thus it even has the form of equation (3.6); and third, the $(p - \frac{1}{2})$ term is of order $1 - h$ and can be ignored when the migration rate is small. We do not include here the calculations behind equation (4.4); it is provided for illustrative purposes only.

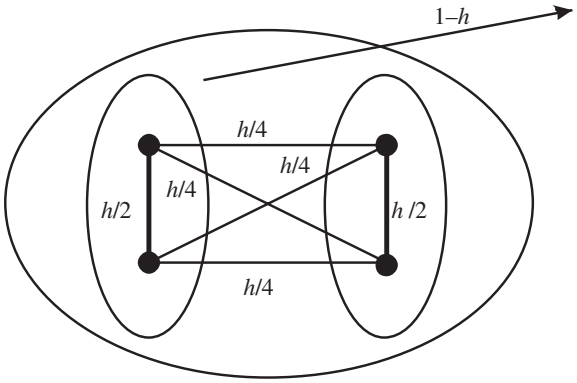

Figure 1. Structured deme with $n = 4$ breeders.

5. APPLICATION TO THE HAWK–DOVE GAME

In evolutionary game theory, the hawk–dove game is one of the earliest examples [10,31,32]. It is a good choice for this paper, as it is familiar and, for a range of parameters, has a stable polymorphic equilibrium.

The matrix for the hawk–dove game (note that dove is row 1, and hawk is row 2) is

$$\begin{bmatrix} \frac{v}{2} & 0 \\ v & \frac{v-k}{2} \end{bmatrix} \sim \begin{bmatrix} k & k-v \\ k+v & 0 \end{bmatrix}. \tag{5.1}$$

The matrix on the left is the standard one in use. Here, v is the prize and k is the amount by which a fight over the prize reduces its value. Doves will not fight, and when two doves encounter the prize, they split it. Two hawks will fight over it and then split the remains. In a hawk–dove encounter, the hawk simply takes the prize at no cost. The matrix on the right puts the original matrix in the form $\begin{bmatrix} B-C+D & -C \\ B & 0 \end{bmatrix}$ by subtracting $(v-k)/2$ and, for convenience, multiplying by 2. We have $C = v - k$, $B = v + k$ and $D = -k$.

Note that if all offspring were to migrate a long distance, we would essentially have a random mixing population and the game pay-offs would provide total fitness (no secondary effects). In this case, it is clear from the matrix in (5.1) that hawk behaviour would invade a pure dove population and, for $k > v$, dove behaviour would invade a pure hawk population. As a result, for $k > v$, we expect a stable intermediate equilibrium. To calculate this equilibrium, we let A play dove and B play hawk, and then W_A will exceed W_B when $pk + (1-p)(k-v) > p(k+v)$, and this simplifies to give

$$p < 1 - \frac{v}{k}. \tag{5.2}$$

As predicted, if there are no secondary effects, a stable intermediate equilibrium requires $k > v$ and is given by $p^* = 1 - v/k$ [10,32].

Now we set the game into the structured populations of examples 1 and 2. In example 1, we set $W_{IF} > 0$ in equation (4.1) with $C = v - k$ and $D = -k$. When this is rearranged, we get the condition that W_I be positive and that the dove frequency p increase to be

$$p < \left(\frac{1}{2} - \frac{v}{k}\right)\left(\frac{n+2h-nh}{n(1-h)}\right) + \frac{1}{2}. \tag{5.3}$$

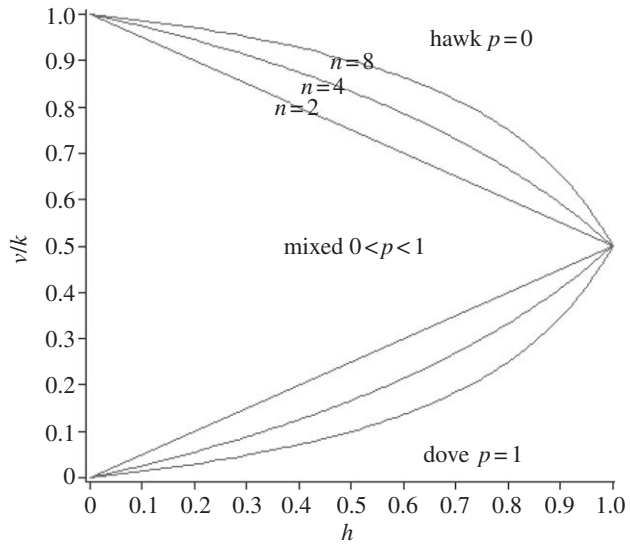

Figure 2. The hawk–dove game in Wright's island model with unstructured demes of size n. The diagram depicts the nature of the stable equilibrium for three values of n in terms of the prize/penalty ratio v/k and the probability h that offspring stay on their native deme.

In condition (5.3), when the r.h.s. is greater than 1, $p = 1$ is a stable equilibrium, and when the r.h.s. is less than 0, $p = 0$ is a stable equilibrium. These conditions are illustrated in figure 2. When the r.h.s. of condition (5.3) is between 0 and 1, it will give us p^*, the allele frequency at a stable polymorphism.

Now look at the two-patch island model of example 2. We set $W_{IF} > 0$ in condition (4.4) with $C = v - k$ and $D = -k$, and rearrange to get that A increases in frequency when

$$p < \left(\frac{1}{2} - \frac{v}{k}\right)\left(\frac{2(2+h)(3-h)}{(1-h)(12+2h+h^2)}\right) + \frac{1}{2}. \tag{5.4}$$

Again, when the r.h.s. of condition (5.4) exceeds 1, p increases to 1, and when it is less than 0, p decreases to 0, as illustrated in figure 3. When the r.h.s. of condition (5.4) is between 0 and 1, it will give us p^*, the allele frequency at a stable polymorphism. Note that this expression for p^* is linear in v/k and this allows us to obtain the value of p^* at any point in figure 3 with a linear interpolation between the $p^* = 0$ and $p^* = 1$ boundaries. The same remark applies, by the way, in example 1 (condition (5.3)).

6. DISCUSSION

A genetical model of behaviour is additive when the fitness of a focal individual depends linearly on the genotypic values of a number of neighbouring individuals. Linear models were typically used in the theoretical development of kin selection, partly because the linearity of the equations allows for simple calculations.

It is worth mentioning the special case in which the dependence of focal fitness on genotype is mediated by behaviour (phenotype) and genetic effects on behaviour are small. In that case, even if the dependence of focal fitness on phenotype is complex, fitness will still depend additively on genotype to first order in the behavioural

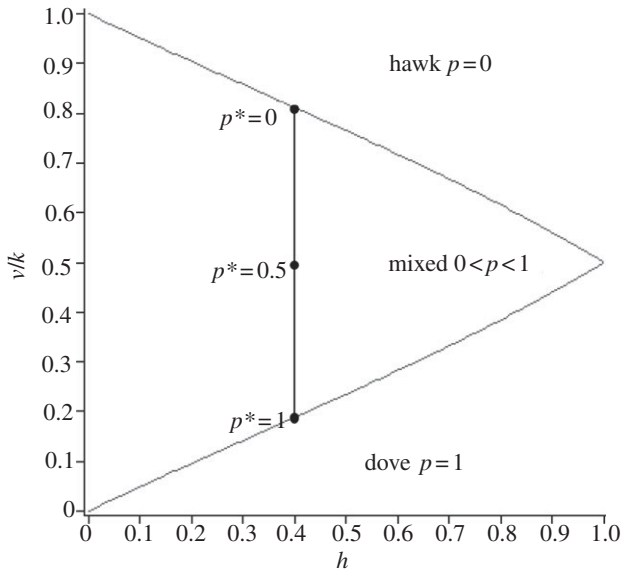

Figure 3. The hawk–dove game in an infinite island model with demes of size 4 structured as two dyadic patches. The diagram depicts the nature of the stable equilibrium in terms of the prize/penalty ratio v/k and the probability h that offspring stay on their native deme. The vertical line illustrates that when there is a stable mixed equilibrium p^*, its value in terms of the ratio v/k can be obtained with a linear interpolation between the $p^* = 0$ and the $p^* = 1$ boundaries.

effects. Indeed, if breeder X with genotype x and phenotype $h_X = h_0 + \varepsilon x$ interacts with breeder Y, the fitness effect will have the form

$$w = \varepsilon\left(\frac{\partial w}{\partial h_X}x + \frac{\partial w}{\partial h_Y}y\right) + o(\varepsilon^2)$$

and to first order in ε this is linear in the genotypic values [33].

However, there are good reasons for extending our analytical reach to interactions with nonlinear fitness effects and the most active such area is found in the study of non-additive evolutionary games. This theory had its start in the early days of hawk–dove and prisoner's dilemma. More recently, in the field of 'evolutionary game theory', games are typically played on graphs, and the results emerge from the interplay between the structure of the game matrix and the structure of the graph.

A significant feature of non-additive fitness effects is the frequency dependence of the results [2,30]. When focal fitness depends linearly on genotypic value, then, even though individual fitness depends on the allele frequency among interactants, the population-wide success of an allele (i.e. its rate of frequency change) will be independent of population-wide allele frequency. To get a dependence on allele frequency, we need non-additive genotypic fitness effects, and this is well illustrated by the dependence of the α term on p in equation (3.5).

The early work on these games [2,32] was set in unstructured ('open') population models and took account only of primary interactions, and as we have seen, the analysis in that case involves only linear methods even though the conditions we obtain (equation (3.10)) are frequency-dependent.

We have suggested that the calculations involved in the analysis of non-additive fitness effects are more complex than those for additive effects. It might be better to say that they are more difficult to execute. For example, non-additive games belong to the former case and require the calculation of covariances of the form $\mathrm{cov}(x, yz)$, whereas additive games only feature $\mathrm{cov}(x, y)$. In fact, the same basic recursive approach is used in both cases. For example, if we use a standard pedigree analysis to calculate $\mathrm{cov}(x, y)$, we use a recursive argument to obtain the probability that breeders X and Y are 'IBD' (i.e. are derived from a common ancestor). If they are, they are both A with probability p and are otherwise both B; if they are not, they are independent and therefore have covariance zero. In the parallel calculation of $\mathrm{cov}(x, yz)$, we have three interacting individuals, and there are several cases (all three IBD, X and Y IBD but not Z, Y and Z IBD but not X, etc.). A recursive argument is still used, but it requires the solution of a system of equations.

Thinking in terms of the hawk–dove game, the results of Tarnita *et al.* [8] come at first as a surprise in that they tell us that the success of either strategy is frequency-independent. Of course when we look carefully at what they have shown, we see that the measure they adopt of the selective advantage of an allele is its long-term average rate of increase in frequency, and it thus represents an average over all frequencies. What is in fact interesting and unexpected in these results is that the *average* rate of increase of either allele has a particularly simple and elegant mathematical form (seen in equation (3.4)), which, in an inclusive fitness analysis, can be calculated in terms of standard linear relatedness coefficients. It is also interesting that the finite population result of Tarnita *et al.* [8] extends readily to asymmetric mutation rates provided these rates are small.

Our work here uses an inclusive fitness approach to extend this analysis to population models with a fixed but arbitrary equilibrium allele frequency p. The interesting conclusion we obtain is that our result (equation (3.5)) continues to display the Tarnita *et al.* [8] form (their eqn (2.7), our equation (3.4)), but includes a frequency-dependent term, and of course obtains exactly their form for the symmetric allele frequency $p = \frac{1}{2}$.

This work was supported by a grant from the Natural Sciences and Engineering Council of Canada. The final version of the manuscript has profited greatly from comments of both Andy Gardner and Corina Tarnita.

REFERENCES

1 Price, G. R. 1970 Selection and covariance. *Nature* **227**, 520–521. (doi:10.1038/227520a0)
2 Queller, D. C. 1985 Kinship, reciprocity and synergism in the evolution of social behaviour: a synthetic model. *Nature* **318**, 366–367. (doi:10.1038/318366a0)
3 Falconer, D. S. 1960 *Introduction to quantitative genetics*, 2nd edn. London, UK: Longman.
4 Queller, D. C. 1992 Quantitative genetics, inclusive fitness and group selection. *Am. Nat.* **139**, 540–558. (doi:10.1086/285343)
5 Queller, D. C. 1992 A general model for kin selection. *Evolution* **46**, 376–380. (doi:10.2307/2409858)
6 Frank, S. A. 1998 *Foundations of social evolution*. Princeton, NJ: Princeton University Press.
7 Gardner, A., West, S.A. & Wild, G. 2011 The genetical theory of kin selection. *J. Evol. Biol.* **24**, 1020–1043. (doi:10.1111/j.1420-9101.2011.02236.x)
8 Tarnita, C. E., Ohtsuki, H., Antal, T., Fu, F. & Nowak, M. A. 2009 Strategy selection in structured

populations. *J. Theor. Biol.* **259**, 570–581. (doi:10.1016/j.jtbi.2009.03.035)
9 Wright, S. 1931 Evolution in Mendelian populations. *Genetics* **16**, 97–159.
10 Maynard Smith, J. 1982 *Evolution and the theory of games*. Cambridge, UK: Cambridge University Press.
11 Ohtsuki, H. & Nowak, M. A. 2006 Evolutionary games on cycles. *Proc. R. Soc. B* **273**, 2249–2256. (doi:10.1098/rspb.2006.3576)
12 Taylor, P. D. 2010 Birth–death symmetry in the evolution of a social trait. *J. Evol. Biol.* **23**, 2569–2578. (doi:10.1111/j.1420-9101.2010.02122.x)
13 Ohtsuki, H. 2010 Evolutionary games in Wright's island model: kin selection meets evolutionary game theory. *Evolution* **64**, 3344–3353. (doi:10.1111/j.1558-5646.2010.01117.x)
14 Nowak, M. A. & May, R. M. 1992 Evolutionary games and spatial chaos. *Nature* **359**, 826–829. (doi:10.1038/359826a0)
15 Nowak, M. A., Sasaki, A., Taylor, C. & Fudenberg, D. 2004 Emergence of cooperation and evolutionary stability in finite populations. *Nature* **428**, 646–650. (doi:10.1038/nature02414)
16 Taylor, P. D. 1992 Altruism in viscous populations—an inclusive fitness model. *Evol. Ecol.* **6**, 352–356. (doi:10.1007/BF02270971)
17 Grafen, A. & Archetti, M. 2008 Natural selection of altruism in inelastic homogeneous viscous populations. *J. Theor. Biol.* **252**, 694–710. (doi:10.1016/j.jtbi.2008.01.021)
18 West, S. & Gardner, A. 2010 Altruism, spite and greenbeards. *Science* **327**, 1341–1344. (doi:10.1126/science.1178332)
19 Rousset, F. & Billiard, S. 2000 A theoretical basis for measures of kin selection in subdivided populations: finite populations and localized dispersal. *J. Evol. Biol.* **13**, 814–825. (doi:10.1046/j.1420-9101.2000.00219.x)
20 Hamilton, W. D. 1964 The genetical evolution of social behaviour I. *J. Theor. Biol.* **7**, 1–16. (doi:10.1016/0022-5193(64)90038-4)
21 Robertson, A. 1966 A mathematical model of the culling process in dairy cattle. *Anim. Prod.* **8**, 95–108. (doi:10.1017/S0003356100037752)
22 Hamilton, W. D. 1963 The evolution of altruistic behaviour. *Am. Nat.* **97**, 354–356. (doi:10.1086/497114)
23 Michod, R. E. & Hamilton, W. D. 1980 Coefficients of relatedness in sociobiology. *Nature* **288**, 694–697. (doi:10.1038/288694a0)
24 Charlesworth, B. 1980 Models of kin selection. In *Evolution of social behaviour: hypotheses and empirical tests* (ed. H. Markl), pp. 11–26. Weinheim, Germany: Chemie.
25 Taylor, P. D. & Frank, S. 1996 How to make a kin selection argument. *J. Theoret. Biol.* **180**, 27–37. (doi:10.1006/jtbi.1996.0075)
26 Taylor, P. D., Day, T. & Wild, G. 2007 From inclusive fitness to fixation probability in homogeneous structured populations. *J. Theor. Biol.* **249**, 101–110. (doi:10.1016/j.jtbi.2007.07.006)
27 Antal, T., Ohtsuki, H., Wakeley, J., Taylor, P. & Nowak, M. 2009 Evolution of cooperation by phenotypic similarity. *Proc. Natl Acad. Sci. USA* **106**, 8597–8600. (doi:10.1073/pnas.0902528106)
28 Taylor, P. D., Day, T. & Wild, G. 2007 Evolution of cooperation in a finite homogeneous graph. *Nature* **447**, 469–472. (doi:10.1038/nature05784)
29 Taylor, P. D., Lillicrap, T. & Cownden, D. 2010 Inclusive fitness analysis on mathematical groups. *Evolution* **65**, 849–859.
30 Lehmann, L. & Keller, L. 2006 Synergy, partner choice and frequency dependence: their integration into inclusive fitness theory and their interpretation in terms of direct and indirect fitness effects. *J. Evol. Biol.* **19**, 1426–1436. (doi:10.1111/j.1420-9101.2006.01200.x)
31 Maynard Smith, J. & Price, G. 1973 The logic of animal conflict. *Nature* **246**, 15–18. (doi:10.1038/246015a0)
32 Grafen, A. 1979 The hawk dove game played between relatives. *Anim. Behav.* **27**, 905–907. (doi:10.1016/0003-3472(79)90028-9)
33 Grafen, A. 1985 Hamilton's rule OK. *Nature* **318**, 310–311. (doi:10.1038/318310a0)

Electronic Supplementary Material for

An inclusive fitness analysis of synergistic interactions in structured populations

Peter Taylor
Wes Maciejewski
Queen's University

Proc. Roy. Soc B, 2012.

ESM-A The linearity of $R_{\mathrm{X}\to\mathrm{YZ}}$

We begin in an infinite population and show that the relatedness coefficient $R_{\mathrm{X}\to\mathrm{YZ}} = \dfrac{\mathrm{cov}(x, yz)}{\mathrm{cov}(x, x)}$ is linear in the overall population allele frequency p. There are a number of ways to do this but here we make use of the concept of *identity by descent* (IBD). We say that two genes are IBD if they have a common ancestor where we make the assumption that a gene that migrates a long distance will never meet a common ancestor. We begin with the numerator of $R_{\mathrm{X}\to\mathrm{YZ}}$. Write

$$\mathrm{cov}(x, yz) = \mathrm{E}(xyz) - \mathrm{E}(x)\mathrm{E}(yz). \qquad \text{(A1)}$$

The first term E(xyz) is simply the probability that X, Y and Z are all A and this will be

- p if all three individuals are IBD at the locus in question,
- p^2 if only two of the three are IBD, and
- p^3 if none of the three are IBD

Thus E(xyz) is a linear combination of p, p^2 and p^3 so is a polynomial of degree at most 3 (and at least 1). By a similar argument, E(xy) is a polynomial of degree 1 or 2, and since E(x) = p, cov(x, yz) is a polynomial of degree 2 or 3. Note now that if p = 0 or 1, cov(x, yz) will equal zero, and thus cov(x, yz) must be divisible by $p(1–p)$. It follows that cov(x, yz) has the form $p(1–p)(s+rp)$ for parameters s and r that are independent of p. Finally, cov(x, x) = var(x) = $p(1–p)$ in a haploid population, so that the quotient $R_{\mathrm{X}\to\mathrm{YZ}}$ must have the form $s+rp$.

An interesting simple example of this is found in an infinite population, when one of Y and Z is actually X. We have:

$$\begin{aligned}\mathrm{cov}(x,xy) &= \mathrm{E}(x^2y) - \mathrm{E}(x)\mathrm{E}(xy) = \mathrm{E}(xy) - \mathrm{E}(x)\mathrm{E}(xy)\\ &= (1-p)\mathrm{E}(xy) = (1-p)[\mathrm{E}(x)\mathrm{E}(y) + \mathrm{cov}(x, y)]\\ &= (1-p)[p^2 + p(1-p)R_{\mathrm{X}\to\mathrm{Y}}] = p(1-p)[p + (1-p)R_{\mathrm{X}\to\mathrm{Y}}]\end{aligned}$$

Hence: $R_{\mathrm{X}\to\mathrm{XY}} = \dfrac{\mathrm{cov}(x, xy)}{\mathrm{cov}(x, x)} = p + (1-p)R_{\mathrm{X}\to\mathrm{Y}}$.

The argument in a finite population is parallel but technically more complex and a comment on this is found at the end of ESM-D.

ESM-B. A variant of the argument of Tarnita et al (2009).

Here we use an inclusive-fitness framework to import the elegant argument of Tarnita et al (2009) into an infinite population with long-range migration at a particular allele frequency p.

Tarnita et al (2009) begin by showing that the condition for allele A to be favoured over B has the form

$$k_1 a + k_2 b > k_3 c + k_4 d \qquad \text{(B1)}$$

where the k_i are independent of the payoffs but do depend on the population structure, the update rule (how old breeders are replaced by new), the mutation rate and the population size. Our analogue of eq. (B1) comes by writing eq. (6) for the inclusive-fitness effect of A in the form:

$$W_{IF\text{-}A} = k_1(p)a + k_2(p)b - k_3(p)c - k_4(p)d \qquad \text{(B2)}$$

Unlike Tarnita et al (2009), we need to explicitly display the dependence of the k_i on the equilibrium allele frequency p, but we emphasize that they are independent of the payoffs, as the covariances in eq. (6) are calculated in the neutral population (Taylor et. al 2007a). The same is true of other aspects of the Tarnita et al (2009) argument that depend on the A-B symmetry, and essentially for the same reason, that the components of the inclusive fitness effect are covariances and they are calculated in the neutral population in which A and B have the same effect.

Following the Tarnita et al (2009) argument, we observe that the inclusive-fitness effect of B must be the analogue of (B2) with the payoffs suitably permuted and p replaced by $q = 1-p$:

$$\begin{aligned} W_{IF\text{-}B} &= k_1(q)d + k_2(q)c - k_3(q)b - k_4(q)a \\ &= -k_4(q)a - k_3(q)b + k_2(q)c + k_1(q)d \end{aligned} \qquad \text{(B3)}$$

Since the inclusive-fitness effect of an allele is proportional to its initial (when the switch is turned on) rate of increase in frequency (Taylor et. al 2007a, eq. 3.6):

$W_{IF\text{-}B} = -\lambda W_{IF\text{-}A}$

for some $\lambda > 0$, and hence, from (B2) and (B3):

$k_4(q) = \lambda k_1(p)$, $k_3(q) = \lambda k_2(p)$, $k_2(q) = \lambda k_3(p)$ and $k_1(q) = \lambda k_4(p)$.

Thus

$k_4(q) = \lambda k_1(p) = \lambda^2 k_4(q)$.

It follows that $\lambda^2 = 1$, and since $\lambda > 0$, $\lambda = 1$. Thus

$k_4(q) = k_1(p)$ and $k_3(q) = k_2(p)$

and eq. (B2) can be written:

$$W_{IF\text{-}A} = k_1(p)a + k_2(p)b - k_2(q)c - k_1(q)d. \qquad \text{(B4)}$$

Now using the payoff matrix: $\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} B-C+D & -C \\ B & 0 \end{bmatrix}$ (B4) becomes:

$$W_{IF\text{-}A} = [k_1(p) - k_2(q)]B - [k_1(p) + k_2(p)]C + k_1(p)D \quad \text{(B5)}$$

Using the fact that in the absence of synergy (D=0), W_{IF} involves only the standard relatedness coefficients $R_{X\to Y}$ (Queller 1985), we write eq. (B5) as

$$W_{IF\text{-}A} = \beta B - \gamma C + k_1(p)D \quad \text{(B6)}$$

where β and γ are independent of p. A comparison of eqs. (B5) and (B6) allows us to write:

$$k_1(p) = \frac{\beta+\gamma}{2} + \frac{k_2(q)-k_2(p)}{2}. \quad \text{(B7)}$$

The last term on the right is linear in p (ESM-A) and is zero when p = ½, and so it must have the form $(p - ½)\alpha$ where α is independent of p and this gives us the form of eq. (12).

ESM-C. The p – ½ term in eq. (12) is first order in the long-term migration rate.

To make this argument, we refer to a couple of equations in ESM-B. Note that the off-diagonal entries b and c in the payoff matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ can contribute to an individual's fitness only when both alleles are present in the focal neighbourhood. When long-range migration is rare, local neighbourhoods will tend towards homozygosity, and must await a migration event for heterozygosity to be restored. It follows from eq. (B2) that the coefficients $k_2(p)$ and $k_3(p)$ of b and c will be (at least) first order in the long-range migration rate. It then follows from eq. (B7) that this will be the case for the difference between $k_1(p)$ and $(\beta+\gamma)/2$, and hence from eq. (B5) it will also hold for α.

We remark that this argument can be just as easily made for a finite population. In this case, for the entries b and c to appear in the calculation both alleles must be present in the population and as a mutation is necessary to free the population from fixation, the relative amount of time the population will be unfixed, and hence the coefficients b and c, will be first order in the mutation rate.

ESM-D. Allele-frequency change in a finite population.

By the *state* of the population, we will mean the specification of the A-B configuration among the nodes, with isomorphic configurations generally identified. A finite population does not have an equilibrium state in the same way as an infinite population and, certainly when the population is small, random sampling of genes will cause the population allele frequency $\bar{x}$ to drift and fixation will inevitably occur. At this point we need mutation to unfix the population and restart what becomes a continual drift-fixation-mutation cycle. Selection, of course, can only act in the unfixed states and Price's (1970) equation (3) will then give the selective component of allele-frequency change. But of course what it gives us will depend very much on the state of the population.

Our measure of overall allele-frequency change will be an average of the change given by eq. (3) where each population state is weighted by its long-term average neutral frequency of occurrence. (Rousset and Billiard 2000, Taylor et. al. 2007a). We write this as:

$$\mathrm{E}\left[\frac{\mathrm{dE}(x)}{\mathrm{d}t}\right] = \mathrm{E}\left[\mathrm{cov}(x, W_{\mathrm{X}})\right] \qquad \text{(D1)}$$

Here we use round brackets to signal a within-state calculation, and square brackets for a calculation over all states with their long-term frequencies (Taylor et al (2007a). Thus the covariance on the right is calculated in each population state and the expectation is taken over all states with their long-term frequency. The inclusive fitness effect is still given as $\mathrm{W}_{\mathrm{IF}} = \sum_i R_i \Delta w_i$ (eq. 4) but the relatedness and the fitness effects become long-term averages over all states. This leads to the following formula:

$$\mathrm{W}_{\mathrm{IF}} = \frac{\mathrm{E}\left[\mathrm{cov}(x, W_{\mathrm{X}})\right]}{\mathrm{E}\left[\mathrm{cov}(x, x)\right]} \qquad \text{(D2)}$$

which is the finite-population analogue of eq. (5) and the corresponding coefficients of relatedness are the analogue of eq. (7):

$$R_{\mathrm{X}\to\mathrm{Y}} = \frac{\mathrm{E}\left[\mathrm{cov}(x, y)\right]}{\mathrm{E}\left[\mathrm{cov}(x, x)\right]}, \qquad R_{\mathrm{X}\to\mathrm{YZ}} = \frac{\mathrm{E}\left[\mathrm{cov}(x, yz)\right]}{\mathrm{E}\left[\mathrm{cov}(x, x)\right]}. \qquad \text{(D3)}$$

To work with $\mathrm{E}\left[\mathrm{cov}(x,\bullet)\right]$ we use the covariance decomposition theorem over all states (Ross 1998): $\mathrm{cov}[x,\bullet] = \mathrm{E}\left[\mathrm{cov}(x,\bullet)\right] + \mathrm{cov}[\mathrm{E}(x), \mathrm{E}(\bullet)]$, as it is $\mathrm{cov}[\bullet,\bullet]$ to which the IBD argument of ESM-A can be applied. See Taylor et al (2007a) for technical details.

Ross, S.M. 1998. A First Course in Probability, Fifth ed. Prentice-Hall Inc, Englewood Cliffs, NJ.