(This is a sample cover image for this issue. The actual cover is not yet available at this time.)

Journal of Theoretical Biology 325 (2013) 76–82

Contents lists available at SciVerse ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Inclusive and personal fitness in synergistic evolutionary games on graphs

Peter Taylor*

Department of Math & Stats, Queen's University, Kingston, ON, Canada K7L 3N6

HIGHLIGHTS

- Analysis of non-additive evolutionary games on a graph.
- Personal- and inclusive, fitness formulation of the equation for allele-frequency change.
- Interpretation of the generalized relatedness coefficients.
- Example of calculation in a finite cycle graph.

ARTICLE INFO

Article history:
Received 13 September 2012
Received in revised form
2 February 2013
Accepted 6 February 2013
Available online 13 February 2013

Keywords:
Non-additive games
Relatedness
Allele frequency
Price equation
Transitivity

ABSTRACT

I study the evolution of a pair of competing behavioural alleles in a finite graph-structured population when there are non-additive or "synergistic" fitness effects. I begin with the Price equation and extend it to both a personal-fitness and an inclusive-fitness formulation. I thereby obtain an extension of "Hamilton's Rule" to synergistic effects and I calculate and interpret the generalized relatedness coefficients. I present an example of the analysis in a cycle graph with 4 nodes.

1. Introduction

An enormous body of significant work constructs analytical models for the genetical evolution of social behavior. The key relationship here is the dependence of focal fitness on the behaviour (phenotype) of a number of interactants. These phenotypes are typically correlated with individual genotypes which may in turn be correlated among interactants. When this is fed into the covariance formula of Price (1970) for allele-frequency change, we get an expression each term of which can be "factored" as a product of a fitness effect and a relatedness between interactants. There are many particular variations of this formulation, but they all go back to Hamilton's path-breaking inclusive-fitness analysis.

In these models, the relatedness coefficients emerge from the dependence of focal fitness W_0 on the genotypic values x_i of various primary and secondary interactants. When this dependence can be assumed to be linear, these coefficients have the form $\mathrm{cov}(x_0, x_i)$ and they deliver the classic coefficients of relatedness (Michod and Hamilton, 1980), but in non-linear models, the covariances, for example $\mathrm{cov}(x_0, x_ix_j)$, are more difficult both to calculate and to interpret (Queller, 1985; Lehmann and Keller 2006; Tarnita et al., 2009; Ohtsuki, 2010).

Here, I study a two-player evolutionary game (Maynard Smith, 1982; Queller, 1985; Nowak and May, 1992; Nowak et al., 2004) in a finite population structured as a graph (Lieberman et al., 2005) in which there are synergistic fitness effects (Queller, 1985) leading to a quadratic dependence of fitness on neighbouring genotypes. I study carefully the relationship between the personal-fitness and the inclusive-fitness formulations, and I show that these provide equivalent interpretations of the components of the Price equation in terms of fitness effects and coefficients of consanguinity (Michod and Hamilton, 1980). I illustrate the method with a cyclic graph of size 4.

2. Population structure

The first thing to say is that I am going to try to be systematic in my use of subscripts. The analysis will be switching from a recipient-centred point of view (personal fitness) to an actor-centred point of view (inclusive fitness). To help the reader keep track of this, I will choose a subscript notation which will distinguish actor and recipient.

* Tel.: +1 613 533 2434; fax: +1 613 542 9883.
E-mail address: Peter.taylor@queensu.ca

0022-5193/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jtbi.2013.02.002

2.1. Demographics

I work with a finite structured population represented as a graph, a set of nodes, indexed by i and j, etc., each occupied by a single asexual haploid breeder, together with edges between each pair of nodes (Lieberman et al., 2005). Each edge (i, j) carries two different pieces of information. The first is the probability e_{ij} that the individual on node i, as actor, will engage in a "primary" interaction (play the game) with the individual on node j, as recipient. The second is the probability d_{ij} that an offspring from node i will displace the individual at node j. I assume that interactions are symmetric (that is, $e_{ij}=e_{ji}$) but I do not require offspring dispersal to be symmetric. I allow d_{ii} to be non-zero.

To keep the analysis simple I assume that the graph is homogeneous or *transitive* (Taylor et al., 2007b). What that means is that given any pair i_0, j_0 of nodes there is a bijection of the node set mapping i_0 to j_0 which preserves the d_{ij} and the e_{ij} (such bijections are called *isomorphisms*). With this assumption, the graph "looks the same" from every node and that allows us to work with a single "focal" node to which I will generally assign the index 0. The extension of these results to non-transitive graphs is discussed later.

I work with a continuous-time population process (Moran) with a birth–death (BD) updating (Ohtsuki and Nowak, 2006; Taylor, 2010).

2.2. A two-allele model

I suppose that there are two alleles A and B assorting at a fixed locus and let the genotypic value x_i of the breeder on node i be the frequency of A in its genotype. I suppose that mutation occurs at birth with a small probability μ and both A and B mutate to an A form with probability p and to a B form with probability $1-p$. Thus the effective (phenotypic) rate of mutation is $\mu(1-p)$ from A to B and μp from B to A, but the genotypic rate μ is higher than either of these. The reason for modeling genetic mutation in this manner is that it gives us the following critical property (Taylor et al., 2007a): if the breeders at two nodes are IBD (identical by descent), they are both A with probability p, and if they are not IBD, they are each independently A with probability p. Here, individuals are IBD if they have a common ancestor with no intervening mutation. When making IBD arguments, we tend to assume automatically that these properties hold, but it is important to note that they will not unless we use the mutation rate μ (instead of $\mu(1-p)$ or μp) as the IBD "breaker." Note that under this process, the neutral (no selection) equilibrium allele frequency (the frequency of A) will be p.

The existence of synergistic effects will introduce quadratic genotype expressions x_ix_j and I will find it useful to express these with a double subscript $x_ix_j=x_{ij}$. This notation has some conceptual as well as technical value in that in the inclusive fitness framework, synergistic effects are created by a pair of actors i and j and it makes sense to consider them as a single generalized actor ij.

2.3. Primary and secondary fitness effects

I assume that in each time-step individuals engage in pairwise interactions, playing the game with payoff matrix

$$\begin{bmatrix} b-c+d & -c \\ b & 0 \end{bmatrix} \tag{1}$$

(Maynard Smith, 1982; Queller, 1985; Nowak and May, 1992; Nowak et al., 2004) where the first row gives the payoffs to an A-player (with an A or B partner) and the second row belongs similarly to B. This matrix form provides the following interpretation of an actor-recipient interaction. Each A-individual, as actor, gives benefit b to its partner at cost c, but if the partner is also A, it gives a synergistic bonus d. I assume that the payoffs represent small increments in fecundity, small enough that we can ignore second-order effects in these payoffs. The payoffs are added to the baseline fecundity of 1 for the next reproductive bout.

It must be noted that in a structured population, these primary interactions will typically have "secondary" fitness effects (Taylor, 1992; Grafen and Archetti, 2008—I use the terminology of West and Gardner (2010)) which must also be accounted for, and the overall fitness effect on an individual will combine the primary effects with the secondary effects from primary interactions in the neighbourhood. Indeed since breeder i interacts with j with probability e_{ij}, the primary fitness effect of these interactions on any breeder j is:

$$w_j=-cx_j+\sum_{i\neq j}e_{ij}(bx_i+dx_{ij}) \tag{2}$$

and the overall fitness effect on the breeder at node k is

$$\begin{aligned} W_k &= w_k-\sum_j d_{jk}w_j \\ &= -cx_k+\sum_{i\neq k}e_{ik}(bx_i+dx_{ik})-\sum_i\left[-cx_id_{ik}+\sum_{j\neq i}e_{ij}(bx_i+dx_{ij})d_{jk}\right] \end{aligned} \tag{3}$$

I remark, that my actor-recipient interpretation of the ordering of the subscripts will help us in the interpretation of the terms in Eq. (3). Here, i is an actor and k is a recipient. The first two terms describe the primary effects of the game i plays with k. The final summation over i describes the secondary effects of i's actions due to the dispersal of offspring to k. In the first term in the square brackets, i acts on itself and d_{ik} carries the mortality effect to k. In the second term i acts on j and d_{jk} carries the effect to k.

3. Price's formula for allele-frequency change

We begin with Price's (1970) covariance formula for allele-frequency change over a single time step (the time between single reproductive events). The classic version of Price's theorem is that the selective change in average allele frequency $\bar{x}$ is equal to the covariance over the population of individual fitness W with individual genotype x: $\Delta\bar{x}=\mathrm{cov}(W,x)$. Note that I am concerned here with the *selective* change so that I ignore the effects of mutation. Price's original formulation added a separate term to account for such changes.

Here I will specialize this formula in two important ways. The first of these starts with the observation that the change in $\bar{x}$ will depend on the population-wide configuration of the alleles A and B, that is on the *state* of the population. If the population is large, we tend not to worry about this as the "average" state will change very little over time. But in a small population, the selective change $\Delta\bar{x}$ can be quite different in different states and what we do is take a long-term average of $\Delta\bar{x}$, effectively an average of its value over all states, each state weighted by its frequency of occurrence (Rousset and Billiard, 2000). This gives us the formulation:

$$\Delta\bar{x}=\sum\nolimits_S\pi_S\mathrm{cov}(W,x) \tag{4}$$

Here, π_S is the long-term frequency of state S, and the covariance is taken over the population in each state.

The problem with this formulation is that it is difficult to work with directly unless the population is very small, as the covariance is hard to calculate in any particular state. What Hamilton's (1964) seminal analysis provided was a formulation which essentially interchanged the within- and between-population processes. A recipient-centred formulation of his approach would

work with the equation:

$$\Delta\bar{x} = \frac{1}{N}\sum\nolimits_k \mathrm{cov}\,[W_k, x_k] \tag{5}$$

In Eq. (5), we take a fixed node k in a fixed state and calculate its fitness W_k, and then we calculate the "longterm" covariance between W_k and x_k over all states. I use square brackets for the covariance to signal that this is not a covariance over a population, but rather correlates changes in genotype and fitness at a fixed node over a long interval of time. That is, the covariance is taken over all states S with each state weighted by its frequency π_S. Finally we take the average of this over all nodes k. It turns out that there are powerful recursive methods using identity by descent to calculate the covariance in Eq. (5) (see the example below) and that is what makes the analysis tractable. The transition from Eq. (4) to Eq. (5) is essentially an interchange of summations and the steps of the analysis are shown in Appendix 1.

The second specialization takes off from the observation that the frequency with which different population states occurs depends on the selective regime. To eliminate this problem, we calculate the covariance in the Price equation using the *neutral* equilibrium state frequencies, that is, we use the equilibrium attained when A has the same behaviour as B, that is, we set $b=c=d=0$, so that although the game is played, the payoffs are zero and there is no selective difference between the alleles. It turns out that the error introduced by using the neutral distribution will be second order in the selective effects b, c, and d, so that we will still get the correct first-order conditions. (Taylor et al., 2007a, Tarnita and Taylor, in preparation).

Finally we observe that in a homogeneous population, $\mathrm{cov}[W_k, x_k]$ is independent of the node k (all nodes will exhibit the same long-term behavior), and we can dispense with the average over k, and simply work with a "focal" node. This gives us our final form:

$$\Delta\bar{x} = \mathrm{cov}[W_0, x_0] \tag{6}$$

where the covariance is taken over all population states. Eq. (6) is the form of the Price equation I will work with. Just to summarize: we take a focal node 0 and calculate the covariance between its fitness and genotype over all population states, using the neutral values of the equilibrium state frequency π_S.

4. Personal fitness

Henceforth, I will take W to be the selective fitness *effect* rather than the fitness itself. These differ by the constant 1 so will have the same covariance with x. When we put Eq. (3) into the covariance Eq. (5), the terms x_i, and x_{ik} and x_{ij} in Eq. (3) becomes covariances: $\mathrm{cov}[x_i, x_k]$, $\mathrm{cov}[x_{ik}, x_k]$ and $\mathrm{cov}[x_{ij}, x_k]$. The personal-fitness formulation is obtained from the Price equation by converting these genotypic covariances into coefficients of relatedness defined as follows:

$$\text{actor } i \text{ to recipient } j: \quad R_{i-j} = \frac{\mathrm{cov}[x_i, x_j]}{\mathrm{var}[x]} \tag{7}$$

$$\text{actor } ik \text{ to recipient } k: R_{ik-k} = \frac{\mathrm{cov}[x_{ik}, x_k]}{\mathrm{var}[x]} \tag{8}$$

$$\text{actor } ij \text{ to recipient } k: R_{ij-k} = \frac{\mathrm{cov}[x_{ij}, x_k]}{\mathrm{var}[x]} \tag{9}$$

For example, in Eq. (9), we think of i and j working as joint actors to produce the synergistic effect d. These coefficients provide precise measures of the genetic similarity between two (or among three) individuals and will be discussed more fully later. Using these in the expansion of Eq. (6) (in which k has been replaced by 0) gives us what is called the personal-fitness effect W_{PF}:

$$W_{\mathrm{PF}} = -cR_{0-0} + \sum_{i\neq 0} e_{i0}(bR_{i-0} + dR_{i0-0}) - \sum_i \left[-cR_{i-0}d_{i0} + \sum_{j\neq i} e_{ij}(bR_{i-0} + dR_{ij-0})d_{j0}\right] \tag{10}$$

Essentially, x_i has been replaced by the relatedness of i to the focal 0 and x_{ij} has been replaced by the relatedness of ij to the focal 0. Eqs. (6)–(9) tell us that W_{PF} in Eq. (10) is the one-step change in allele frequency divided by the genetic variance:

$$W_{\mathrm{PF}} = \frac{\Delta\bar{x}}{\mathrm{var}(x)} \tag{11}$$

To summarize, Price's formula takes a focal node and follows the genotype and fitness of the individual at that node over time, and tells us that the allele A will increase in frequency under the action of selection if there is a positive correlation between the genotype and the fitness. Personal fitness takes this and drills down to the level of the social partners. It says that for the allele to increase in frequency, it must be more highly correlated with the genes of those social partners whose behaviour has a positive effect on fitness than with the genes of those social partners whose behaviour does not. And it is the relatedness coefficients that measure this correlation.

5. Inclusive fitness

Personal fitness formulates the fitness of a focal recipient as it is affected by the behaviour of a number of actors. Inclusive fitness turns the tables around and takes the focal individual to be the actor focusing on how its behaviour affects the fitness of a number of recipients. In this process it applies a weight to these recipients which corresponds to how closely they are related to the focal actor. These weights are essentially the relatedness coefficients. The conceptual advantage of this formulation is that the focal individual is placed in the driver's seat in the sense that we can imagine it choosing its behaviour to maximize an average measure of community fitness where the average is constructed using relatedness as weights (Grafen, 2006; West and Gardner, 2013). This average is called the inclusive fitness of the focal individual.

I now show how this IF formulation relates to the PF formula of Eq. (10). Start with Eq. (3) and imagine all the summations expanded, so that all terms are written out. To get the PF formulation we classify the terms according to the value of the recipient index k and group all those with the same k. We then use the transitivity of the graph to argue that the groups are really all the same (i.e., the $k=1$ group has the same set of terms as the $k=2$ group, etc.) and then we take only one of those groups and set the index k to 0. To get the IF formulation we take the same expanded form of Eq. (3), but now we classify the terms according to the value of the actor index i and group all those with the same i. We again use the transitivity of the graph to argue that the groups are all the same and we take only one of those groups and set the index i to 0. When we replace genotypic values with relatedness coefficients we get the formulation:

$$W_{\mathrm{IF}} = -cR_{0-0} + \sum_{k\neq 0} e_{0k}(bR_{0-k} + dR_{0k-k}) - \sum_k \left[-cR_{0-k}d_{0k} + \sum_{j\neq 0} e_{0j}(bR_{0-k} + dR_{0j-k})d_{jk}\right] \tag{12}$$

It should be clear from this analysis that Eqs. (10) and (12) have the same set of terms but just differently organized, so that $W_{\mathrm{IF}} = W_{\mathrm{PF}}$.

The quantity W_{IF} is called the *inclusive-fitness effect* of the action of allele A, and as I have suggested above, it has a powerful agent-based interpretation. We can imagine a focal individual who bears a silent allele A asking what the fitness effect might be of activating that allele, the effect, not only on its own fitness, but on the fitness of all others affected by the focal behaviour *who might also bear the allele A*. That is the job of the relatedness coefficients, to measure the probability that an affected individual will indeed carry A, conditional on the focal having A. Of course what we are really measuring here is the effect of the focal behaviour on the allele A itself, more precisely *the effect of the action of the allele on its own frequency*. Indeed, Price's formula tells us that that is exactly what both Eqs. (10) and (12) are measuring.

6. The interpretation of relatedness

The relatedness coefficients can be calculated in terms of the coefficients of consanguinity G (Michod and Hamilton, 1980) defined as follows. Let G_{ij} be the probability that the individuals on nodes i and j are IBD and let G_{ijk} be the probability that the individuals on nodes i, j and k are all IBD. [Recall that individuals are IBD (identical by descent) if they have a common ancestor with no intervening mutation.] Then:

$$R_{i-j} = G_{ij} \tag{13a}$$

$$R_{ij-i} = G_{ij} + (1-G_{ij})p \tag{13b}$$

$$R_{ij-k} = G_{ijk} + (G_{ik} - G_{ijk})p + (G_{jk} - G_{ijk})p \tag{13c}$$

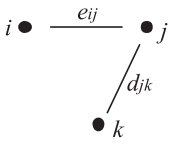
Eq. (13a) is straightforward. If a pair of nodes i and j are IBD then they have the same allele and that is A with probability p, B with probability $1-p$. If they are not IBD, they are each independently A with probability p. Thus $\mathrm{cov}[x_i,x_j] = G_{ij}\mathrm{var}[x] + (1-G_{ij})(0) = G_{ij}\mathrm{var}[x]$. A similar but more complex argument for Eq. (13c) is found in Appendix 2, and Eq. (13b) is a special case (set $k=i$) of Eq. (13c).

The G-coefficients can be calculated with a standard recursive argument (Ohtsuki, 2010) and an example is provided in Appendix 3. My objective here is to interpret these IBD expressions and see that they square with our intuition of what relatedness ought to be measuring. Before doing this, I must point out that a standard way to normalize relatedness is to set the relatedness of an actor to an average node in the population to be zero (Michod and Hamilton, 1980; Rousset and Billiard, 2000; Taylor et al., 2007a; Taylor 2008), but I have not done this in Eq. (13). When mean fitness is zero (or simply constant) this does not affect final result and it is more transparent for us to work with the unnormalized form. I remark however that Eq. (13a) does provide the other standard normalization, that an individual has relatedness 1 to itself.

I now turn to the interpretation of the R-expressions and I begin by noting that Eq. (10) and (12) provide a different story for PF and IF. With PF we have a focal recipient and the job of W_{PF} is simply to tell us how its fitness as an A-individual compares with its fitness as a B-individual. More precisely W_{PF} measures the difference between the two. The IF story is a bit more complicated but also a bit more interesting. Here we have a focal actor and the job of W_{IF} is to measure the effect of its A-behaviour on the overall frequency of A. That is the approach I use for the analysis in Fig. 1.

7. Example: Calculation in a 4-cycle

The population consists of four nodes arranged on the vertices of a square, numbered (around the square) as 0, 1, 2 and 3 (Fig. 2).



case	configuration		probability	$k=$
	IBD	allelic		
1	$i \equiv j \equiv k$		G_{ijk}	A
2	$i \equiv k$	j=A	$(G_{i0} - G_{ij0})p$	A
3	$j \equiv k$	j=A	$(G_{j0} - G_{ij0})p$	A
4a	$i \equiv j$	k=A	$(G_{j0} - G_{ij0})p$	A
4b	$i \equiv j$	k=B	$(G_{j0} - G_{ij0})(1\text{-}p)$	B
5a	none	j=A, k=A	$g_0 p^2$	A
5b	none	j=A, k=B	$g_0 p(1-p)$	B

Fig. 1. Analysis of the term $e_{ij}dR_{ij-k}d_{jk}$. My objective here is to provide an intuitive interpretation for the form of the relatedness coefficient R_{ij-k} found in Eq. (13c). I use an inclusive fitness interpretation. A parallel interpretation using personal fitness is also available and is in fact a bit simpler. I begin with an actor i bearing allele A. This affects the behaviour of i and the objective of inclusive fitness is to measure the effect of this on the frequency of A. Now this effect will be realized through many pathways, and I take as an example only one of these––the effect on the mortality of k through a synergistic fecundity gift to j. First of all, the probability that this pathway is available is $e_{ij}d_{jk}$. I assume this is the case and set this equal to 1. Second, j will receive the gift d only if it carries A so we need to include the probability of that. Thirdly, to assess the effect of this on the frequency of A *through* k we need to know whether k carries A or B. In the table, I list the different IBD and allelic configurations in which j will receive d and I record the probability of each. Here g_0 is the probability there are *no* IBD relationships among the three (Appendix 2). The effect through k on the frequency of A will be obtained as the difference between the cases in which k carries A and B. The notation $i \equiv j$ indicates that the breeders on nodes i and j are IBD, and the notation j=A says that the breeder on node i has allele A. Entries in the IBD column specify *all* the relationships among the three, so that, for example, in case 2, j is *not* IBD to i and k. The calculation of the probabilities relies on the result that if the breeders at two nodes are IBD, they are both A with probability p, and if they are not IBD, they are each independently A with probability p. Note that g_0 in 5a and 5b is the probability of no IBC relationship among the three nodes. Cases 1, 2 and 3 all provide an effect on the frequency of A. Case 4 has two subcases, an effect 4a on A and an effect 4b on B. The probabilities are in the ratio $p/(1-p)$ and this is the same as the ratio of the frequency of A to B. That tells us that the frequency effects cancel each other. The same holds for cases 5a and 5b. Thus, the net effect on the frequency of A is found in the top three cases and these are exactly the three IBD components of the relatedness coefficient R_{ij-k} in Eq. (13c).

Individuals that share a common side are called neighbours. In each time step each individual plays the $\begin{bmatrix} b-c+d & -c \\ b & 0 \end{bmatrix}$ game with *each* of its two neighbours. Payoffs provide small increases in fecundity. I use a BD updating rule so that changes in fecundity produce mortality effects. Offspring disperse with equal probability to the two neighbouring nodes.

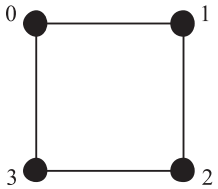


Fig. 2. A 4-cycle.

7.1. Personal fitness

I begin with a PF approach. I take node 0 as the focal recipient. The fecundity increment of individual j is

$$w_j = (b + dx_j)(x_{j+1} + x_{j-1}) - 2cx_j \tag{14}$$

where we treat our indices "mod 4" so that $3+1=0$, etc. These are the effects of the primary interactions; the secondary effects are the changes in mortality from fecundity changes of the two neighbours. Focal fitness is then:

$$W_0 = \text{fecundity} - \text{mortality} = w_0 - \frac{w_1 + w_3}{2} \tag{15}$$

which simplifies to:

$$W_0 = b(x_1 + x_3 - x_0 - x_2) - c(2x_0 - x_1 - x_3)$$

$$+d\left(\frac{x_{01}+x_{03}-x_{12}-x_{23}}{2}\right) \tag{16}$$

To convert this to a PF formulation, we introduce the following relatedness notation:

R_1 to neighbour e.g., R_{0-1}
R_2 across diagonal e.g., R_{0-2}
R_3 neighbouring actors interacting within e.g., R_{01-1}
R_4 neighbouring actors interacting outside e.g., R_{01-2}

Then Eq. (16) simplifies as:

$$W_{PF}=b(2R_1-R_0-R_2)-c(2R_0-2R_1)+d(R_3-R_4) \tag{17}$$

7.2. Inclusive fitness

I take node 0 as the focal actor. This approach allows us to write a version of Eq. (17) almost by inspection. We have a focal fecundity gift of b to each neighbour (R_1) but this also produces a mortality effect felt equally at nodes 0 and 2 (R_0 and R_2). The focal fecundity cost is $2c$ (R_0) and this produces a mortality effect felt equally at nodes 1 and 3 (R_1). Finally the focal gift of d to each neighbour is joint with that neighbour (R_3) and the mortality effect is felt equally at nodes 0 and 2 (R_3 and R_4). This gives us:

$$W_{IF}=b(2R_1-R_0-R_2)-c(2R_0-2R_1)+d\left(2R_3-2\frac{R_3+R_4}{2}\right) \tag{18}$$

and this is the same as Eq. (17).

It remains to calculate the relatedness coefficients and Eq. (13) give us these in terms of the coefficients of consanguinity G. To specialize G to the current example, I introduce:

$$\begin{array}{ll} G_1 & \text{CC to neighbour e.g., } G_{01} \\ G_2 & \text{CC across diagonal e.g., } G_{02} \\ G_3 & \text{CC among any three nodes e.g., } G_{012} \end{array} \tag{19}$$

Then Eq. (13) gives us:

$$\begin{aligned} R_1&=G_1 \\ R_2&=G_2 \\ R_3&=G_1+(1-G_1)p \\ R_4&=G_3+(G_1-G_3)p+(G_2-G_3)p \end{aligned} \tag{20}$$

The G-coefficients are calculated recursively in terms of the most recent replacement of a breeder by an offspring. In all cases, a replacement offspring is a mutant with probability μ and is then not IBD to any breeder. Take nodes 0 and 1 and suppose 1 is most recently replaced. The offspring came either from 0 (giving us $G_0=1$) or from 2 (giving us G_2). That gives us the first equation below. Now take nodes 0 and 2, and suppose 2 is most recently replaced. The offspring came either from 1or 3 (giving us in both cases G_1). That gives us the second equation below. Finally take nodes 0, 1 and 2. If 2 is the most recently replaced, the offspring came either from 1 (giving us G_1) or from 3 (giving us G_3). A similar argument applies for node 0. If 1 is the most recently replaced, the offspring came either from 0 or 2 (giving us in both cases G_2). That gives us the third equation below. We have:

$$\begin{aligned} G_1&=(1-\mu)\frac{G_2+1}{2} \\ G_2&=(1-\mu)G_1 \\ G_3&=(1-\mu)\left[\frac{2}{3}\frac{G_1+G_3}{2}+\frac{1}{3}G_2\right] \end{aligned} \tag{21}$$

To first order in μ, these solve to give

$$\begin{aligned} G_1&=1-3\mu \\ G_2&=1-4\mu \\ G_3&=1-5\mu \end{aligned}$$

and when these are put into Eq. (18), we get

$$W_{IF}=(-2b-6c+2d)\mu. \tag{22}$$

Note that this has the form provided by Tarnita et al. (2009) and Taylor and Maciejewski (2012):

$$W_{IF}=\left(\beta b-\gamma c+\frac{\beta+\gamma}{2}d\right)\mu$$

and in particular (to first order in μ) is independent of the long-term allele frequency p.

8. Discussion

I have reformulated the Price formula (Eq. (4)) for allele-frequency change with both a personal-fitness and an inclusive-fitness approach. Historically this methodology emerged from Hamilton's original (1964) inclusive-fitness approach. The idea was already out there (Wright, Fisher, Haldane, e.g., Haldane 1955) that the evolutionary success of a genetic trait which was "social" (affecting the fitness of others) would depend on the fitness of a number of "related" interactants, but Hamilton (1964) provided the first formal account of how this might be calculated. Hamilton's approach had the additional stunning property that all the various fitness effects which needed to go into the calculation could be credited, in a carefully weighted manner (the weights being essentially the R's), to a single focal "actor" who in principle could use this "inclusive-fitness" quantity to decide whether a potential action would be evolutionarily favoured. In that sense, we could view the actor as a maximizing agent (Grafen 2006, 2009).

Inclusive fitness and personal fitness are in many ways "mirror-image" accounting schemes for calculating the genetic fitness of an item of social behaviour (West and Gardner, 2013). In the inclusive-fitness approach there is a single focal actor and the calculation produces a weighted sum of the effects of its behaviour on the fitness of all related recipients. In the personal-fitness approach there is a single focal recipient and the calculation produces a weighted sum of the effects on its fitness of the behaviour of all related actors. A number of studies have shown that, for additive interactions, these two approaches yield equivalent calculations; indeed they are simply different ways of classifying the many different pathways through which behaviour affects fitness. One of my objectives in this work is to check that the argument continues to work for non-additive interactions such as those found in the game matrix (1).

Synergistic effects, in the first instance, challenge our notion of the actor, and require us to allow a pair of individuals i and j to act together, and I even regard the pair as "the actor ij." However, this also requires more complex relatedness coefficients and a second objective of mine is to provide an interesting formula (Eq. (13c)) for the triple relatedness coefficient R_{ij-k} in terms of coefficients of consanguinity G, and to give an intuitive interpretation of the terms of this formula. Along with this, I provide an example (Eq. (21)) of the recursive calculation of the triple coefficients of consanguinity G_{ijk}. All this work derives in an essential way from the insights of Michod and Hamilton (1980), Queller (1985), Tarnita et al. (2009) and Ohtsuki (2010).

I have restricted attention here to transitive graphs Taylor et al. (2007b), graphs for which the population structure looks the

same from every node, but the arguments generalize readily to non-transitive graphs. In this case we have a class-structure in the graph, the classes being the orbits under the group of all graph isomorphisms (bijections of the node set which preserve structure, in this case the e_{ij} and the d_{jk}). Here different classes can have different reproductive value (*RV*) and we must incorporate *RV* in our definition of fitness *W* (Taylor, 2008; Tarnita and Taylor, in preparation). It is also the case that, in Eq. (6), one must choose a focal individual from each relevant class (Taylor and Frank, 1996—here the term "direct fitness" was introduced, but "personal fitness" is now a common and much better term). That is, in the PF approach we need a focal individual selected from each recipient class and in the IF approach we need a focal individual selected from each actor class.

I remark that in classic IF studies, the class structure typically emerges from different roles in the social structure, e.g., male–female, parent–offspring, and in this case behaviour is typically class specific, e.g. parental care, so that we have a single actor class (the parent), but more than one recipient class (parent and offspring). In this case, in a PF approach requires more than one focal recipient but an IF approach needs only one focal actor. Of course, in these models we are often interested in two interacting traits (parental provisioning and offspring begging) but these are two traits, not one, and in an IF approach each are analyzed with a single focal individual, though the W_{IF}(parent) and the W_{IF}(offspring) will each depend on the level of both traits.

I have ended with an illustration of the calculation in 4-cycle population. In preparing this example I found myself discovering that my technical approach to relatedness in a finite population over the past decade was more complicated than it needed to be. To explain this point more fully, I begin with Price's formula (Eq. (4)). In its original 1970 appearance, the covariance was thought of (and explicitly presented!) as calculated over all individuals (nodes) in the population. That works fine for a large population in which all possible local configurations of alleles are represented somewhere, but in small populations this would not be the case at any one time and the covariance might better be regarded as taken over the "long-term" population. Well, let's be careful: selection happens at each particular time-step and we must use the covariance over the population at a fixed time to give us the allele-frequency change in the next step (so the 1970 formula certainly works for both finite and infinite population models). The problem is that in a small population, this covariance will change over time and the reasonable way to accommodate that (and this is the significant contribution of Rousset and Billiard, 2000) is to take a long-term average of the covariance and I write that as E[cov(*W*, *x*)]. To help us keep our thinking clear, I adopt the notation of using round brackets for expectation/covariance over a population and square brackets for expectation/covariance over the long-term. The problem is that it is difficult to explicitly calculate this average.

What we *can* do, in a population of constant size, is effectively interchange the order of the brackets and this is explained in Appendix 1. Note that the new formulation we get from this, cov[W_0, x_0], does correspond to the standard recursive approach we use to calculate relatedness coefficients. Indeed, if W_0 depends on x_i, that requires us to calculate cov[x_i, x_0] and the argument that this is $G_{ij} \times \mathrm{var}[x]$ is essentially a calculation at a long-term equilibrium.

One final remark: there is an important sense in which the personal and inclusive fitness approaches do *not* require additive interactions (e.g., Hamilton, 1970; Queller, 1992; Gardner et al., 2011). The idea is to write fitness W_k (Eq. (3)) as a linear regression in the genotypes x_i and then let the regression coefficients serve as the new fitness effects $\hat{b}_{ik}$ in place of the payoffs *b*, *c*, and *d*. Thus, $W_k = \Sigma_i \hat{b}_{ik} x_i + \varepsilon_k$ with cov(W_k, ε_k)=0. However, we still have the task of calculating the $\hat{b}_{ik}$ and the resulting inclusive-fitness expression does not have the direct physical interpretation in terms of the primary payoffs and probabilities of interaction and dispersal that we gain with our personal- and inclusive-fitness interpretations.

Acknowledgements

The development of these ideas was aided through interactions with colleagues at a meeting on "Inclusive fitness: theory and practice" sponsored by the Kavli Royal Society International Centre. Comments of Stu West and a number of conversations with Corina Tarnita and Wes Maciejewski on closely related work also assisted my thinking. Comments of two reviewers of an earlier draft had an enormous effect on the shape of this paper. This work was supported by a grant from the Natural Sciences and Engineering Council of Canada.

Appendix 1. Price's covariance formula for allele-frequency change

The first thing to note about the formula $\Delta\bar{z} = \mathrm{cov}(W,x)$ is that the covariance will depend on the population state—the configuration of A and B among nodes. Traditional applications of the formula have a large population in mind, large enough so that at any moment all local configurations of the alleles are appropriately represented, and $\Delta\bar{z}$ would not change much from generation to generation. But for a small population, this will not be the case and to get a reasonable measure of allele frequency change, we want to take a long-term average, i.e., an average over all possible population states, each weighted by frequency of occurrence (Rousset and Billiard, 2000). We write this as

$$\mathrm{E}[\Delta\bar{z}] = \mathrm{E}[\mathrm{cov}(W,x)] \tag{1.1}$$

Here we have adopted the bracket notation of Taylor et al. (2007a), whereby round brackets signal expectation or covariance over the population (that is, over the node set) and square brackets signal expectation or covariance over all population states (that is, over the long term).

In order to calculate this, we execute a sequence of steps whose effect is to interchange the order of brackets, as follows.

$$\mathrm{E}[\mathrm{cov}(W,x)] \tag{1.1}$$

$$= \mathrm{cov}[W,x] \tag{1.2}$$

$$= \mathrm{E}(\mathrm{cov}[W_i,x_i]) \tag{1.3}$$

$$= \mathrm{cov}[W_0,x_0] \tag{1.4}$$

The covariance in (1.2) is between the fitness and the genotype of a random node in a random state. The equivalence of (1.1) and (1.2) follows from the covariance decomposition theorem (Ross 1998) using the population state as the class:

$$\mathrm{cov}[W,x] = \mathrm{E}[\mathrm{cov}(W,x)] + \mathrm{cov}[\mathrm{E}(W),\mathrm{E}(x)] \tag{1.5}$$

The first term on the left is the average within-state covariance and the second is the covariance over all states of average fitness with average genotype. This second term is zero as in a population of constant size, average population fitness E(*W*) is constant. And that gives us the equivalence.

In (1.3), the variables are subscripted to emphasize the interpretation of the covariance. Here we take a fixed node *i* and calculate the covariance between its fitness and its genotype as time changes, that is over all population states. And then we average the results over all nodes. The equation follows again from the covariance decomposition theorem, but using the nodes

as the classes:

$$\mathrm{cov}[W,x] = \mathrm{E}(\mathrm{cov}[W_i,x_i]) + \mathrm{cov}(\mathrm{E}[W_i],\mathrm{E}[x_i]) \quad 1.6$$

The first term on the left is the average within-node covariance and the second is the covariance over all nodes of long-term average fitness with long-term average genotype. This second term is zero as the long-term average $\mathrm{E}[x_i]$ must equal p for all i, since the covariance calculations use the neutral distribution. And that gives us (1.3). Finally, we use the transitivity of the graph to note that all nodes exhibit the same long-term behaviour, so that the covariance in Eq. (1.3) is the same for all i and the expectation is obtained from any fixed focal node.

Appendix 2. The triple IBD calculation

Suppose that we have three distinct nodes: 1, 2 and 3. Use the notation $i \equiv j$ to indicate that the breeders on nodes i and j are IBD. For convenience, I introduce the coherent notation:

$g_{123} = G_{123} = \mathrm{Prob}(1 \equiv 2 \equiv 3)$.
$g_{ij} = G_{ij} - G_{ijk} = \mathrm{Prob}(i \equiv j$ but neither are IBD to the third node $k)$.
$g_0 = \mathrm{Prob}$(no two of the three nodes are IBD).

Note that

$$g_{123} + g_{12} + g_{13} + g_{23} + g_0 = 1$$

as these are a complete set of the five disjoint conditions on the IBD configuration for the three nodes. If we take these five conditions to define the five possible IBD classes, then the covariance decomposition theorem gives us

$$\mathrm{cov}[x_1,x_{23}] = [g_{123}\mathrm{cov}_{123} + g_{12}\mathrm{cov}_{12} + g_{13}\mathrm{cov}_{13} + g_{23}\mathrm{cov}_{23} + g_0\mathrm{cov}_0] + \mathrm{cov}(\mathrm{E}[x_1],\ \mathrm{E}[x_{23}]).$$

where each of the subscripted covariances on the right is the long-term covariance of x_1 with x_{23} conditional on the corresponding IBD condition. Here the expression in the square brackets is the *average within class* covariance, and the expression at the right is the *between class* covariance where $\mathrm{E}[x]$ denotes the long-term average taken over each class. But this last term is zero as $\mathrm{E}[x_1] = p$ in each of the five classes.

Now I calculate the conditional covariances. Note first that $\mathrm{cov}_{23} = \mathrm{cov}_0 = 0$ as in these classes, x_1 and x_2x_3 are independent. For cov_{12}, x_3 is independent of x_1 and x_2 and hence:

$$\mathrm{cov}_{12}[x_1,x_2x_3] = \mathrm{E}[x_3]\mathrm{cov}_{12}[x_1,x_2] = p\ \mathrm{cov}[x,x] = p\mathrm{var}[x]$$

The same argument shows that $\mathrm{cov}_{23} = p\mathrm{var}[x]$. Finally $\mathrm{cov}_{123}[x_1, x_2\ x_3] = \mathrm{cov}[x, xx] = \mathrm{var}[x]$.

Putting all this together:

$$\mathrm{cov}[x_1,x_2x_3] = g_{123}\mathrm{cov}_{123} + g_{12}\mathrm{cov}_{12} + g_{13}\mathrm{cov}_{13} = [G_{123} + (G_{12} - G_{123})p + (G_{13} - G_{123})p]\mathrm{var}[x]$$

and this gives us Eq. (13c). Eq. (13b) is obtained by specializing Eq. (13c) to the case $k = i$.

Reference

Gardner, A., West, S.A., Wild, G., 2011. The genetical theory of kin selection. J. Evol. Biol. 24, 1020–1043.
Grafen, A., 2006. Optimization of inclusive fitness. J. Theor. Biol. 238, 541–563.
Grafen, A., 2009. Formalizing Darwinism and inclusive fitness theory. Philos. Trans. R. Soc. London, Ser. B 364, 3135–3141.
Grafen, A, Archetti, M., 2008. Natural selection of altruism in inelastic homogeneous viscous populations. J. Theor. Biol. 252, 694–710.
Haldane, J.B.S., 1955. Population genetics. New Biol. 18, 34–51.
Hamilton, W.D., 1964. The genetical evolution of social behaviour I. J. Theor. Biol. 7, 1–16.
Hamilton, W.D., 1970. Selfish and spiteful behaviour in an evolutionary model. Nature 228, 1218–1220.
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.
Lieberman, E., Hauert, C., Nowak, M.A., 2005. Evolutionary dynamics on graphs. Nature 433, 312–316.
Maynard Smith, J., 1982. Evolution and the Theory of Games. Cambridge UniversityPress, Cambridge.
Michod, R.E., Hamilton, W.D., 1980. Coefficients of relatedness in sociobiology. Nature 288, 694–697.
Nowak, M.A., May, R.M., 1992. Evolutionary games and spatial chaos. Nature 359, 826–829.
Nowak, M.A., Sasaki, A., Taylor, C., Fudenberg, D., 2004. Emergence of cooperation and evolutionary stability in finite populations. Nature 428, 646–650.
Ohtsuki, H., 2010. Evolutionary games in Wright's island model: kin selection meets evolutionary game theory. Evolution 64, 3344–3353.
Ohtsuki, H., Nowak, M.A., 2006. Evolutionary games on cycles. Philos. Trans. R. Soc. London, Ser. B 273, 2249–2256, http://dx.doi.org/10.1098/rspb.2006.3576.
Price, G.R., 1970. Selection and covariance. Nature 227, 520–521.
Queller, D.C, 1985. Kinship, reciprocity and synergism in the evolution of social behaviour: a synthetic model. Nature 318, 366–367.
Queller, D.C, 1992. A general model for kin selection. Evolution 46, 376–380.
Ross, S.M., 1998. In: Fifth (Ed.), A First Course in Probability. Prentice-Hall Inc, Englewood Cliffs, NJ.
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.
Tarnita, C.E., Ohtsuki, H., Antal, T., Fu, F., Nowak, M.A., 2009. Strategy selection in structured populations. J. Theor. Biol. 259, 570–581, http://dx.doi.org/10.1016/j.jtbi.2009.03.035.
Tarnita, C.E., Taylor, P.D. (in prep).
Taylor, P.D., 1992. Altruism in viscous populations––an inclusive fitness model. Evol. Ecol. 6, 352–356.
Taylor, P.D., 2008. Decompositions of Price's formula in an inhomogeneous population structure. J. Evol. Biol 22, 201–213, http://dx.doi.org/10.1111/j.1420-9101.2008.01640.x.
Taylor, P.D., 2010. Birth–death symmetry in the evolution of a social trait. J. Evol. Biol. 23, 2569–2578.
Taylor, P.D., Frank, S., 1996. How to make a kin selection argument. J. Theor. Biol. 180, 27–37.
Taylor, P.D., Day, T., Wild, G., 2007a. From inclusive fitness to fixation probability in homogeneous structured populations. J. Theor. Biol. 249, 101–110, http://dx.doi.org/10.1016/j.jtbi.2007.07.006.
Taylor, P.D., Day, T., Wild, G., 2007b. Evolution of cooperation in a finite homogeneous graph. Nature 447, 469–472.
Taylor, P.D., Maciejewski, W. 2012. An inclusive fitness analysis of synergistic interactions in structured populations. Philos. Trans. R. Soc. London, Ser. B 279, 4596–4603, http://dx.doi.org/10.1098/rspb.2012.1408.
West, S., Gardner, A., 2013. Adaptation and social evolution. Current Biology, in press.
West, S., Gardner, A., 2010. Altruism, spite and greenbeards. Science 327, 1341–1344.