doi:10.1111/j.1420-9101.2006.01196.x

Direct fitness or inclusive fitness: how shall we model kin selection?

P. D. TAYLOR, G. WILD & A. GARDNER

Department of Mathematics and Statistics, Queen's University, Kingston, ON, Canada

Keywords:

Class structure;
Hamilton's rule;
Price equation;
relatedness;
reproductive value.

Abstract

Two standard mathematical formulations of kin-selection models can be found. Inclusive fitness is an actor-centred approach, which calculates the fitness effect on a number of recipients of the behaviour of a single actor. Direct fitness is a recipient-centred approach, which calculates the fitness effect on the recipient of the behaviour of a number of actors. Inclusive fitness offers us a powerful heuristic, of choosing behaviour to maximize fitness, but direct fitness can be mathematically easier to work with and has recently emerged as the preferred approach of theoreticians. In this paper, we explore the fundamental connection between these two approaches in both homogeneous and class-structured populations, and we show that under simple assumptions (mainly fair meiosis and weak selection) they provide equivalent formulations, which correspond to the predictions of Price's equation for allele frequency change. We use a couple of examples to highlight differences in their conception and formulation, and we briefly discuss a two-species example in which we have a class of 'actor' that is never a 'recipient', which the standard direct fitness method can handle but the usual inclusive fitness cannot.

Introduction

Kin selection is a deterministic evolutionary force that acts when individual fitness is affected by traits expressed by relatives. The objective of evolutionary modelling is to obtain conditions under which a kin-selected trait will change in frequency or in character. A standard approach is to assume that the trait in question is genetically determined and to find conditions for the change in frequency of the underlying alleles. The classic covariance formula of Price (1970) provides a simple direct equation for allele frequency change, but when fitness is affected by the behaviour of others who might also carry the alleles in question, the calculations required by this approach can be complicated.

Inclusive fitness, a powerful heuristic method introduced by Hamilton (1964), was designed to handle these complications. Hamilton introduced his idea with the observation 'if we were to follow the usual approach of the progress due to natural selection in a generation, we should attempt to give formulas for the neighbour-modulated fitness…' and he suggests that the calculations for this might be 'rather unwieldy'. This prompted him to introduce an 'alternative approach', which restricts attention to the fitness effects of a single focal actor, and this is what we call inclusive fitness. The decades following Hamilton's paper saw a wonderful development of his idea, simplifying it, extending it, and relating it to various exact genetic formulations. But the neighbour-modulated approach, now usually called direct fitness (Taylor & Frank, 1996; Frank, 1998), has also been developed and the main point of this article is that these two approaches are computationally equivalent, although in different models one approach or the other often seems more immediate or natural. In fact, in recent years, direct fitness has emerged as the preferred method of modelling kin selection (e.g. Gandon, 1999; Perrin & Mazalov, 2000; Day, 2001; Leturque & Rousset, 2002, 2003; Wild & Taylor, 2005; Pen, 2006). However, it is important to say that inclusive fitness, in tracking the various fitness effects of a single individual's behaviour, mirrors the way evolutionary biologists think, and likely for that reason, it remains the preferred mode of analysis by most biologists.

Correspondence: Peter D. Taylor, Department of Mathematics and Statistics, Queen's University, Kingston, ON, Canada K7L 3N6.
Tel.: +1 613 533 2434; fax: +1 613 533 2964;
e-mail: taylorp@post.queensu.ca

Here, we describe and compare these two heuristic approaches with particular attention to their formulation in a class-structured population (Taylor, 1990; Taylor & Frank, 1996; Rousset, 2004). Under a wide range of assumptions the two approaches come up with the same conditions for the evolutionary increase or decrease in allele frequency, but there are conceptual and technical differences in their formulation and our purpose here is to make these clear. Basically, we will argue that each approach arises from the other by a simple re-indexing process. Suggestions have been made that direct and inclusive fitness are essentially different (Frank, 1997, 1998), but it seems to us that this arises from (interesting) attempts to extend the methods beyond their normal range of applicability, e.g. to fitness interactions between species. We will present an example to try to clarify this situation.

In our experience, the difficulties that often arise in the formulation of inclusive and direct fitness arguments, and in their comparison, have to do with the wide range of notations and assumptions found in the literature, and it seems to us that there is a real need for a simple general scheme and a self-contained exposition of it. A key ingredient to a clear understanding and easy implementation of any method is a good notational scheme and the literature has not consistently provided one (not even in our own papers). We have tried to remedy that here and to employ a common scheme to facilitate comparison between the two approaches. We will apply our formulation to three models of altruistic behaviour, a model of altruistic interaction between offspring, a class-structured model in which offspring provide a benefit to parents, and a host–parasite model with interaction between species.

We consider a population of individuals. This might be homogeneous consisting of individuals all of the same type or heterogeneous with individuals of different classes (male, female, or adult, juvenile or large, small, etc.) or even different species. Our objective will be to track the changing frequency of an allele found in a number of members of the population. This allele will affect the behaviour of a number of individuals who carry it and this in turn will affect the fitness of this individual and of other interactants and this will cause a change in the frequency of the allele. This is a complex situation – many individuals in different situations affecting the fitness of different configurations of other individuals. We will handle this by classifying the interactions and then adding the effects of the different types of interactions that affect a random individual. This is essentially a differential approximation, justified by an assumption of small effects (Grafen, 1985a).

We use the term 'interaction' to refer to a basic 'cause and effect' unit. Each interaction will involve a single actor and a single recipient and it is crucial to be aware of the distinction between these. An actor is characterized by its phenotype; a recipient is characterized by its genotype. Some individuals in the population will be both actors and recipients; as an actor their phenotype is what counts; as a recipient their genotype is what counts. Other individuals might be only recipients and will have only genotypes; still others might be only actors and have only phenotypes. Each type of interaction will have an effect on allele frequency and the overall population-wide change will be obtained by adding up these effects.

Table 1 A summary of the main notation used in the text.

Symbol	Explanation
c_j	$u_j v_j$, the reproductive value of class j
dw_k	Effect of deviant behaviour on the fitness of kth recipient
dw_{jk}	Effect of deviant behaviour on the fitness of the kth recipient belonging to class j
G	Genotypic value of the focal individual
G_j	Genotypic value of a focal individual belonging to class j
G_k	Genotypic value of the kth recipient belonging to the social group of the focal individual
$\bar{G}$	Population-wide genotypic value, also the frequency of the target allele
$\bar{G}_j$	Average genotypic value among individuals belonging to class j
n_j	The number of class j recipients in the social group of the focal actor (inclusive fitness) or equivalently the number of class j actors in the social group of the focal recipient (direct fitness)
P	Phenotypic value of the focal individual
P_k	Phenotypic value of the kth actor belonging to the social group of the focal individual
$\hat{P}$	Normal phenotypic value
R_k	$\mathrm{cov}(G,P_k)/\mathrm{cov}(G,P)$, the relatedness of the kth actor to the focal recipient in his social group (direct fitness arguments), or $\mathrm{cov}(G_k,P)/\mathrm{cov}(G,P)$, the relatedness of the focal actor to the kth recipient in his social group (inclusive fitness arguments). Note that both formulations of relatedness are equivalent
R_{jk}	$\mathrm{cov}(G_{jk},P)/\mathrm{cov}(G,P)$, the relatedness of the focal actor to the kth recipient of class j (inclusive fitness arguments). This is equivalent to the coefficient, R_{kj} used in direct fitness arguments
R_{kj}	$\mathrm{cov}(G_j,P_k)/\mathrm{cov}(G,P)$, the relatedness of the kth actor to the class j focal recipient (direct fitness arguments). This is equivalent to the coefficient, R_{jk} used in inclusive fitness arguments.
u_j	The steady-state frequency of class j
v_j	The reproductive value of an individual belonging to class j
W	Fitness of the focal individual
W_j	Fitness of a focal individual belonging to class j
$\bar{W}$	Population-wide average fitness
$\hat{W}$	Normal fitness
$\hat{W}_j$	Fitness of a normal class j individual
W_{dir}	Direct fitness effect
W_{inc}	Inclusive fitness effect
w_i	The i-fitness of a random individual defined as her contribution to class i in the next generation
w_{ij}	$= E[w_i \mid$ individual belongs to class $j]$, the average i-fitness of a random individual in class j
$\mathbf{w}$	$= [w_{ij}]$, a square matrix that maps the frequency distribution of classes from one time step to the next

An example of an individual which is a recipient but not an actor is found in a class-structured population in which adult females care for juveniles. The adults are both actor (in providing care) and recipient (in paying the cost of care), but the juveniles are only recipients (in receiving the care). An example of an individual which is an actor but not a recipient is found in a two species interaction in which a parasite inhabits a host. If we are interested only in a parasite trait, the parasite is both actor and recipient, but the host who responds to the parasite is only an actor. For this last example, there will usually be alternative or more general modelling approaches which would work better. Our notation is found in Table 1.

Homogeneous population

In a homogeneous population, interactions are symmetric and each individual is both actor and recipient. The fundamental result describing allele frequency change is Price's (1970) covariance formula. If we let the genotypic value G of an individual be the frequency of the target allele in its genotype and let its fitness be W, then Price's formula tells us that the population-wide allele frequency $\bar{G}$ will increase exactly when G and W are positively correlated. Precisely:

$$\Delta\bar{G} = \frac{1}{\bar{W}}\mathrm{cov}(G, W), \qquad (1)$$

where $\bar{W}$ is population-wide average fitness. Some simplifying assumptions are needed for this, notably that meiosis is 'fair', that on average there is no change in allele frequency in transmission from parent to offspring.

In building a model, we use behaviour to mediate the relationship between genotype and fitness. The genotypic value G of a focal individual will be correlated with the phenotypic values P_i of a number of actors and the P_i will in turn affect the fitness W of the focal individual. For example, P might denote the probability of performing an altruistic act. Often we handle the effect on fitness by treating fitness as a function of these phenotypic values: $W = W(P_0,P_1,P_2,\ldots)$. To further simplify matters we assume that this function can be treated as linear. Of course, this will seldom be the case but it will be a good approximation in a 'small-effects' argument, in which phenotypic values deviate slightly from normal behaviour $\hat{P}$, and we use a first-order differential approximation $W = \hat{W} + \mathrm{d}W$ (Grafen, 1985a). Here the differential fitness (due to the deviant behaviour) is

$$\mathrm{d}W = \sum_k \frac{\partial W}{\partial P_k}|_{\hat{P}}(P_k - \hat{P}), \qquad (2)$$

where the derivatives are evaluated in the uniform $\hat{P}$ population. Then eqn (1) becomes:

$$\Delta\bar{G} = \frac{1}{\bar{W}}\sum_k \frac{\partial W}{\partial P_k}|_{\hat{P}}\mathrm{cov}(G, P_k). \qquad (3)$$

At this point it is worth mentioning Queller's (1992) paper which shows that if fitness W is regressed directly on genotypic value (of an individual and neighbours), we get a simple general form of inclusive or direct fitness without the many assumptions needed to make phenotypic arguments work. This elegant result is mainly of theoretical interest, as we typically get hold of the partial regressions by using phenotype as an intermediate variable. (Queller's, 1992 discussion of the role of phenotype was to some extent foreshadowed by Cheverud, 1984.)

Direct fitness

Equation (3) arises directly from Price's formula and for that reason is called a direct fitness analysis of the fitness interaction (Taylor & Frank, 1996), also referred to as 'neighbour-modulated' fitness (Hamilton, 1964). More precisely, we define the direct fitness increment to be:

$$W_{\mathrm{dir}} = \sum_k \frac{\partial W}{\partial P_k}R_k, \quad \text{where } R_k = \frac{\mathrm{cov}(G, P_k)}{\mathrm{cov}(G, P)} \qquad (4)$$

with everything evaluated at $P = P_k = \hat{P}$. Here P is the phenotypic value of the focal individual and R_k is the relatedness of the kth actor (or the type k actor depending on the setup) to the focal recipient. Provided $\mathrm{cov}(G,P) > 0$ (which can always be arranged), W_{dir} will have the same sign as the change $\Delta\bar{G}$ in average allele frequency.

Inclusive fitness

The direct fitness approach fastens attention on a random individual recipient and adds up the effects on its fitness of the behaviour of all actors. An alternative formulation, inclusive fitness, introduced by Hamilton (1964), again takes a random individual but adds up the effects of its behaviour on the fitness of all recipients. This is essentially a re-allocation of the direct fitness effects, each effect being credited to the actor rather than to the recipient. But with this different formulation comes a different conceptualization. Instead of a fitness function $W(P_1,P_2,P_3,\ldots)$ which expresses the fitness of an individual in terms of the behaviour of others, the inclusive fitness version starts with a random actor, with phenotypic value P, and tabulates the effects of its behaviour on the fitness of a number of recipients, with genotypic values G_k. This actually already requires the 'allocation of effects' found in eqn (2) and for this reason, inclusive fitness can be more difficult to formulate; that is, inclusive fitness 'begins' with the differential effects of behaviour on fitness. Let $\mathrm{d}w_k$ be the effect of a deviant act on the fitness of the kth recipient. Then the inclusive fitness effect is:

$$W_{\mathrm{inc}} = \sum\nolimits_k \mathrm{d}w_k R_k, \quad \text{where } R_k = \frac{\mathrm{cov}(G_k, P)}{\mathrm{cov}(G, P)}, \qquad (5)$$

where R_k is the relatedness of the focal actor to the kth recipient (or to the type k recipient depending on the setup).

Notice that although R_k appears to be differently defined in eqns (4) and (5) the difference between formulae is purely technical and derives from the re-indexing. If we take these effects to be the fitness derivatives from eqn (2): $dw_k = \partial W/\partial P_k$, then eqns (4) and (5) are equivalent. However, the formulae are thought of in different ways. The terms of the direct fitness effect (eqn 4) are effects of the behaviour of others on the fitness of the focal individual; the terms of the inclusive fitness effect (5) are the effects of the behaviour of the focal individual on the fitness of others.

An example of cooperation between patchmates

As an example we take a patch with n asexual haploid breeders and consider a cooperative interaction between random offspring. An actor will have three categories of fitness effects: on itself (the cost c), on its partner (the benefit b) and on others who experience increased competitive effects due to the extra products of the cooperation. Note that the list of 'recipients' of an item of behaviour must include all those whose fitness is affected by the behaviour, not only the 'direct' effects, in this case the cost and benefit, but also 'indirect' effects of altered competition for limited resources, first identified by Hamilton (1971) and attributed to what he called 'population viscosity' and incorporated into inclusive fitness arguments by Wilson *et al.* (1992) and Taylor (1992a,b). Continuing with the argument, we take $dw_0 = -c$ (the cost to the actor), $dw_1 = b$ (the benefit to the partner) and $dw_2 = -(b - c)$ (the loss of fitness by competitors). The inclusive fitness effect is then:

$$W_{\text{inc}} = -cR_0 + bR_1 - (b - c)R_2, \tag{6}$$

where $R_0 = 1$ is the relatedness of the actor to itself, R_1 is the relatedness between offspring native to the same patch and R_2 is the relatedness of the actor to the competitively displaced individuals.

To identify these displaced individuals in the last term of eqn (6), we need further details of the population structure. As an example, suppose that the population has Wright's (1943) island structure and the islands are breeding patches. Suppose that after the fitness interaction there is random dispersal of offspring at rate d to an infinitely distant patch, followed of course by the arrival of an equal number of immigrants. Finally, we have competition for the n breeding spots. In that case, of the $b - c$ additional individuals created, a proportion d will disperse and will compete with those who are unrelated to the actor. The remaining $1 - d$ will stay on the patch and will compete with the post-dispersal inhabitants. Now a proportion d of these will be immigrants and will also be unrelated to the actor, and the remaining $1 - d$ will be native to the patch. That is two factors of $1 - d$ and the relatedness of the actor to the competitively displaced offspring will be $R_2 = (1 - d)^2 R_1$, where R_1 as above, is relatedness between offspring native to the same patch. These verbal arguments can in fact be tricky and the direct fitness calculation will give us a chance to check our reasoning. This analysis allows us to write the inclusive fitness effect as:

$$W_{\text{inc}} = -cR_0 + bR_1 - (b - c)[(1 - d)^2 R_1]. \tag{7}$$

Notice that our original description of the model used an inclusive fitness format – specifying the effects of the behaviour. This is typically the most natural way to present a model, and it also guides our intuition and our reasoning. This accounts in large part for the popularity of the inclusive fitness paradigm.

The direct fitness approach requires us to write the fitness of an actor as a function of the behaviour of all potential interactants. Suppose that phenotypic value P is the probability of cooperating, and let P_0 and P_1 be the phenotypic values of the actor and of a random offspring native to the same patch. Then if there are N individuals on a patch, the number after dispersal will be

$$[N + NP_1(b - c)](1 - d) + Nd = N[1 + P_1(b - c)(1 - d)].$$

If neutral fitness is normalized at $\hat{W} = 1$, then normalized fitness of an actor can be written as:

$$\begin{aligned} W(P_0, P_1) &= d(1 - cP_0 + BP_1) + (1 - d)\frac{1 - cP_0 + bP_1}{1 + P_1(b - c)(1 - d)} \\ &= 1 - cP_0 + bP_1 - (b - c)(1 - d)^2 P_1 \end{aligned} \tag{8}$$

The first expression partitions the fitness according to whether the actor disperses (probability d) or does not disperse (probability $1 - d$), and the second expression provides a first-order approximation which assumes that c and b are small. This approximation gives us a linear expression for W, and the direct fitness effect is clearly

$$W_{\text{dir}} = -cR_0 + bR_1 - (b - c)(1 - d)^2 R_1 \tag{9}$$

and we see that this is the same as the inclusive fitness effect.

However, note that the terms of eqns (7) and (9), are interpreted in a different way. The terms of the inclusive fitness (eqn 7) are the fitness effects on all individuals (weighted by relatedness) of a single interaction. The terms of the direct fitness (eqn 9) are the fitness effects on a single individual of all interactions (again weighted by relatedness). This leads to a different analysis of the two last terms, the competitive effects. In eqn (7) the dispersal effect $(1 - d)^2$, is counted as a component of the relatedness term (although there are other ways to do the analysis) whereas in eqn (9) this is counted as a component of the fitness effect (see also Gardner & West, 2006).

A class-structured population

Reproductive value

Now we suppose that there is a class structure in the population. Two important examples are found in gender classes and age classes. For an example which combines these, take the four classes to be adult males, adult females, juvenile males and juvenile females. In a neutral population we let u_j be the frequency of class j and v_j be the reproductive value of a random class j individual (Taylor, 1990). We let c_j be the class j reproductive value, and we normalize reproductive values so that

$$c_j = u_j v_j \quad \text{and} \quad \sum_j c_j = 1. \tag{10}$$

With this normalization, c_j can be interpreted as the probability that the ancestor (today) of an allele selected at random in a distant future generation resides in class j. To interpret $v_j = c_j/u_j$ we begin by choosing a random class j individual today and ask for the probability that the ancestor (today) of an allele selected at random in a distant future generation belongs to that particular individual. Then the v_j are proportional to these probabilities.

There is the question of how to measure and normalize fitness in such a population. In one time unit an individual might contribute to several classes, for example, through fecundity and survival, or through male and female offspring. We will let w_i be the 'i-fitness' of a random individual defined as her contribution to class i in the next generation. Now let w_{ij} be the average i-fitness of a class j individual in the neutral population. Then the matrix $\mathbf{w} = [w_{ij}]$ is the (neutral) fitness matrix of the population. In an age-structured population, this is just the Leslie matrix. Then the dominant right eigenvector of $\mathbf{w}$ is the class frequency vector $[u_j]$ and the dominant left eigenvector is the individual reproductive value vector $[v_i]$ (Taylor, 1990, 1996). The dominant eigenvalue λ of $\mathbf{w}$ is the growth rate of the population when it has attained stable proportions $[u_j]$, and we assume that the neutral population stays constant in size and take $\lambda = 1$. Then the eigenvector equations are:

$$\begin{aligned} u_i &= \sum_j w_{ij} u_j \\ v_j &= \sum_i v_i w_{ij} \end{aligned}. \tag{11}$$

From eqn (11) it makes sense to define the 'fitness' W of a random individual (in any class) to be the sum of i-fitness components, each weighed by the individual class reproductive value (Taylor, 1990, eqn 13):

$$W = \sum v_i w_i. \tag{12}$$

With this definition, eqn (11) shows that in the neutral population average class j fitness is v_j. We point out that different treatments of direct fitness normalize fitness in different ways, for example average class j neutral fitness is set equal to v_j (as here) in Taylor (1990) but set equal to 1 in Taylor & Frank (1996). These different normalizations will lead to different multiplicative constants in the equations and this can cause confusion. We are convinced that the most natural definition of fitness in this general context is eqn (12) and hence average class j fitness is normalized at v_j. However, for completeness we will provide the version of our final direct fitness formula for fitness normalized to 1 (see eqn 18, below).

We suppose that the actors belong to a single class. That seems a reasonable assumption, as actors belonging to different classes are apt to be doing different things or at least to find themselves in different circumstances, and can often be treated separately (Pen & Taylor, 2005; Wild & Taylor, 2005). However, the recipients will generally live in several classes and our formulation will group them by class.

Price's formula

In a class-structured population, at first it is not clear what allele frequency ought to mean. We can measure the average frequencies $\bar{G}_j$ within each class, but how are they to be averaged to get an overall measure? We ask in what way do we want a change in the various $\bar{G}_j$ to count as an increase in 'overall' frequency, and the answer to this is that what we want to measure is the increase in descendent genes in the long-term future of the population and we achieve this with a class reproductive value weighting (Fisher 1930, Price and Smith 1972). More precisely, multiplying the $\bar{G}_j$ by class frequencies u_j converts frequency to numbers of genes, and then multiplying by individual reproductive values v_j converts this to asymptotic numbers. The product of these gives class reproductive values c_j (eqn 10) and these are the weights we choose. We define

$$\bar{G} = \sum_j c_j \bar{G}_j. \tag{13}$$

An interesting observation is that in a neutral population, $\bar{G}$ will not change no matter how the mutant allele is distributed among the classes (Taylor, 1990, eqn 9 cites an observation of Uyenoyama for this).

With this definition of allele frequency, Taylor (1990, eqn 16 with $r = 1$) shows that Price's formula for the change in $\bar{G}$ over one time unit has the form:

$$\Delta\bar{G} = \sum_j u_j \mathrm{cov}(G_j, W_j), \tag{14}$$

where G_j and W_j are the genotype and fitness of a random class j individual (fitness normalized so that neutral fitness is v_j) and u_j is the class frequency. If you are wondering what happened to the denominator of average fitness (found in eqn 1) it is essentially found in the u_j

multiplier. Use the fact that average class j fitness is v_j together with eqn (10) that $\sum_j u_j v_j = 1$.

Now we treat W_j as a function of the phenotypic values P_k of a number of actors, and use the first-order Taylor expansion to get $W_j = \hat{W}_j + \mathrm{d}W_j$, where:

$$\mathrm{d}W_j = \sum_k \frac{\partial W_j}{\partial P_k}|_{\hat{P}}(P_k - \hat{P}). \tag{15}$$

Substitute this into eqn (14):

$$\Delta\bar{G} = \sum_j u_j \sum_k \frac{\partial W_j}{\partial P_k}|_{\hat{P}}\mathrm{cov}(G_j, P_k) \quad (W_j \text{ normalized to } v_j). \tag{16}$$

Direct fitness

As before, we begin by selecting a random focal recipient. Now in this case recipients can belong to different classes, and it is not so clear what it means to choose one 'at random'. It turns out that the way to do this is to select a random class j recipient with probability the class frequency u_j (Taylor & Frank, 1996). Note that this depends on our decision to normalize class j fitness to v_j, but see below for an alternative normalization. The direct fitness increment is then:

$$W_{\mathrm{dir}} = \sum_j u_j \sum_k \frac{\partial W_j}{\partial P_k}|_{\hat{P}} R_{kj}, \quad \text{where } R_{kj} = \frac{\mathrm{cov}(G_j, P_k)}{\mathrm{cov}(G, P)}, \tag{17}$$

where R_{kj} is the relatedness of the kth actor to the focal class j recipient and where we use G to denote the actor's genotype. If j is the actor class and we choose k so that the kth actor is the class j recipient, then $G = G_j$ and $P = P_k$ and $R_{kj} = 1$. A comparison of eqns (16) and (17) shows that direct fitness has the same sign as the change in average allele frequency $\bar{G}$.

It is worth pointing out that in some models (e.g. a sex-ratio model in Wild & Taylor, 2005) it is more natural to normalize average class fitness to be 1. In this case, we need to include the v_j along with the u_j weights, and that gives us c_j weights (eqn 10). For example, those are the weights found in Taylor & Frank (1996) but not in Taylor (1990). With this weighting, the direct fitness increment eqn (17) becomes

$$W_{\mathrm{dir}} = \sum_j c_j \sum_k \frac{\partial W_j}{\partial P_k}|_{\hat{P}} R_{kj} \quad (W_j \text{ normalized to } 1). \tag{18}$$

Inclusive fitness

We choose a random focal actor with phenotypic value P and let $\mathrm{d}w_{jk}$ be the 'effect' of a deviant act on the fitness of the kth recipient in class j which has genotypic value G_{jk}. Then the inclusive fitness effect is:

$$W_{\mathrm{inc}} = \sum_j \sum_k \mathrm{d}w_{jk} R_{jk}, \tag{19}$$

where

$$R_{jk} = \frac{\mathrm{cov}(G_{jk}, P)}{\mathrm{cov}(G, P)} \tag{20}$$

is the relatedness of the actor to the kth recipient in class j and where we use G to denote the actor's genotype (if j is the actor class it will be one of the G_{jk} – if the actor is recipient k in class j, then $G = G_{jk}$ and $R_{jk} = 1$). Again, a re-indexing gives an equivalence between eqns (18) and (19) and thus the direct and inclusive fitness formulations lead to the same result. This is not quite obvious as eqn (18) has the factor u_j and eqn (19) does not. It has to be noted that if an actor affects the fitness of n_j recipients in class j, then a random class j recipient will be affected by $(u_0/u_j)n_j$ actors, where u_0 is the frequency of the actor class.

A class-structured example

We illustrate these equations with a model that displays the difficulties that might arise in a class-structured population. Suppose we have an asexual haploid population with two classes, juveniles (class 1) and breeding adults (class 2) with Leslie matrix

$$\mathbf{w} = \begin{bmatrix} 0 & 10 \\ 1/20 & 1/2 \end{bmatrix}. \tag{21}$$

The columns give the expected output after 1 year. A juvenile (column 1) has probability 1/20 of becoming a breeding adult next year; otherwise it dies. A breeding adult (column 2) recruits 10 offspring into the juvenile class next year and in addition has probability 1/2 of retaining its breeding status another year. The matrix $\mathbf{w}$ has dominant eigenvalue 1 with right eigenvector $\mathbf{u} = \left[\begin{smallmatrix}10\\1\end{smallmatrix}\right]$ and left eigenvector $\mathbf{v} = [1,\ 20]/30$. Thus, at the annual census, there are 10 juveniles to every adult and an adult has 20 times the reproductive value of a juvenile. The class reproductive values are:

$$\begin{aligned} c_1 &= u_1 v_1 = 1/3 \\ c_2 &= u_2 v_2 = 2/3 \end{aligned} \tag{22}$$

and these are already normalized to have sum 1. In the neutral population the average class j fitnesses are

$$\begin{aligned} \bar{W}_1 &= \sum_i v_i w_{i1} = \left(\frac{20}{30}\right)\left(\frac{1}{20}\right) = \frac{1}{30} = v_1, \\ \bar{W}_2 &= \sum_i v_i w_{i2} = \left(\frac{1}{30}\right)(10) + \left(\frac{20}{30}\right)\left(\frac{1}{2}\right) = \frac{2}{3} = v_2 \end{aligned} \tag{23}$$

and as expected, these are the individual reproductive values.

Now suppose the mutant allele causes a juvenile to behave cooperatively by giving benefit b to a neighbouring adult at cost c (e.g. a helping behaviour). The problem is to find conditions on b and c for this behaviour to be adaptive and to increase in frequency. Again we take phenotypic value P to represent the probability of cooperating, with $\hat{P}$ as the resident value.

The first thing to ask is whether we are to interpret these fitness increments additively or multiplicatively. We are of course free to do either, but given the way we have normalized fitness (as reproductive value) the most natural interpretation is multiplicative. Thus, an adult with resident fitness v_2 who received a benefit of $b = 0.15$ would get a 15% increase in fitness, giving her a new fitness of

$$1.15v_2 = v_2(1 + b) = v_2 + v_2 b. \tag{24}$$

Her additive increment is not b but $v_2 b$.

Inclusive fitness

Take a random juvenile actor. There are two recipients, one in each class, the actor herself with fitness effect $\mathrm{d}w_1 = -v_1 c$ and her adult neighbour with fitness effect $\mathrm{d}w_2 = v_2 b$. Then eqn (19) gives

$$\mathrm{W}_{\mathrm{inc}} = -v_1 c R_1 + v_2 b R_2 = -v_1 c + v_2 b R, \tag{25}$$

where R is the relatedness of a juvenile to her adult neighbour. This is Hamilton's rule in this setting.

Direct fitness

In this case, we choose a random recipient in each class, a juvenile and an adult with fitnesses W_1 and W_2. The juvenile will be affected only by herself (as actor)

$$W_1(P_1) = v_1(1 - cP_1) \tag{26}$$

and the adult will be affected, on average, by 10 juvenile actors:

$$W_2(P_1, P_2, \ldots, P_{10}) = v_2(1 + b(P_1 + P_2 + \cdots + P_{10})). \tag{27}$$

Then eqn (17) gives

$$\begin{aligned} W_{\mathrm{dir}} &= u_1 \frac{\mathrm{d}W_1}{\mathrm{d}P_1} R_1 + u_2 \sum_k \frac{\partial W}{\partial P_k} R_{k2} \\ &= 10(-v_1 c) R_1 + 1(10 v_2 b) R_2 \\ &= 10(-v_1 c + v_2 b R), \end{aligned} \tag{28}$$

where we have replaced the R_{k2} by their average value R_2 and then called this R, the relatedness between a juvenile actor and her adult neighbour. This is equivalent to the inclusive fitness eqn (25).

Interactions between species

In a homogeneous population all individuals can play the role of both actor and recipient. In a class-structured population there will typically be individuals who are recipients but not actors. In this case, inclusive fitness, which is actor centred, and direct fitness, which is recipient centred, provide equivalent modelling approaches. Are there examples of individuals who are actors but not recipients? If so, a direct fitness approach should work well, but inclusive fitness is problematic. The reason is simple – inclusive fitness works with one focal actor and it requires this actor to behave in a deviant manner. But this deviation needs to come from an altered genotype (as the whole point is to track allele frequency change) and if the actor is not a recipient it would not have a genotype. Direct fitness solves this problem by having at least two types of actor, one of which has a genotype and can also play the role of recipient. This observation seems to have been first made in Queller (1992).

Here, we discuss a simple example of an interaction between species, similar to an example of Frank (1997), which illustrates the issue. Suppose a parasite inhabits a host and the host carries a locus, which determines the level of 'cooperative' behaviour towards the parasite. Cooperation exacts a fitness cost for the host but elicits a response from the parasite (e.g. reduced virulence), which enhances the fitness of the host. There are different ways through which this response might work, perhaps a plastic reaction, perhaps, if the parasite has a relatively short generation time, a genetic change. We are interested in tracking the cooperative behaviour of the host.

Let W be the host fitness, let G and P be the host genotype and phenotype, and let P_1 be the parasite phenotype. Our assumptions above are that P_1 depends on P (i.e. the response of the parasite to the host) and W depends on both P and P_1.

To formulate a direct fitness model in our notation, Frank (1997) would likely write $W = W(P,P_1)$ and use a version of eqn (3) or eqn (4):

$$W_{\mathrm{dir}} = \frac{\partial W}{\partial P} \mathrm{cov}(G, P) + \frac{\partial W}{\partial P_1} \mathrm{cov}(G, P_1). \tag{29}$$

With a normalization the first of these becomes a relatedness coefficient, but the second does not, and this moves us, quite reasonably, to a slightly extended concept of direct fitness. Perhaps one could try to extend inclusive fitness to handle the situation in which actors belong to different classes, but our guess is that in most cases alternative modelling approaches would be more reasonable. In the example above, our preference would be to treat the effect of the parasite on host fitness as one of two pathways by which the host affects her own fitness through her behaviour. Formally, we might use another notation for the above two-variable host fitness, say $\tilde{W}(P, P_1)$, treat P_1 as a function of P, and write host fitness W as a function only of P:

$$W(P) = \tilde{W}(P, P_1(P)). \tag{30}$$

Then inclusive fitness is found in eqn (1) with only a single summand:

$$W_{\mathrm{inc}} = \frac{\mathrm{d}W}{\mathrm{d}P} R = \left(\frac{\partial \tilde{W}}{\partial P} + \frac{\partial \tilde{W}}{\partial P_1} \frac{\mathrm{d}P_1}{\mathrm{d}P} \right) R, \tag{31}$$

where R is the relatedness of the actor to herself, which is of course 1. This seems the simplest, most direct and most natural way to handle the argument.

Conclusions

We have examined two classic formulations of kin-selection, direct and inclusive fitness. Direct fitness is a reformulation of Price's covariance formula, and therefore, under suitable assumptions, principally fair meiosis and small genetic effects, it provides a true measure of allele frequency change. Inclusive fitness is essentially a reorganization of the direct fitness calculation and therefore also measures allele frequency change. In particular, direct and inclusive fitness always give the same answer.

The two formulations use slightly different forms of the relatedness coefficient, $\mathrm{cov}(G,P_k)/\mathrm{cov}(G,P)$ in direct fitness (eqn 4) and $\mathrm{cov}(G_k,P)/\mathrm{cov}(G,P)$ in inclusive fitness (eqn 6), and in their generalizations in (eqn 17) and (eqn 20). The first was introduced by Orlove & Wood (1978), the second by Michod & Hamilton (1980) and both forms have been compared and their merits discussed by many others, notably Seger (1981), Grafen (1985b) and Queller (1985, 1992). There are two things to mention here. First, these two forms emerge naturally from the two methodologies, direct fitness which is centred on a focal recipient (i.e. eqn 4) and inclusive fitness which is centred on a focal actor (i.e. eqn 6). The many discussions in the literature of the relationship between these two forms often fail to draw attention to this critical connection. Secondly, the two forms are in fact equivalent (Queller, 1992). This is clear if the two approaches are regarded as sums over fitness interactions, in each case the same sum, but organized differently, with exactly the same relatedness weight applied to each fitness effect. To emphasize this we write both expressions as $\mathrm{cov}(G_R,P_A)/\mathrm{cov}(G,P)$. The numerator is the covariance between recipient genotype and actor phenotype, and the denominator is the covariance between genotype and phenotype in the actor class.

We have assumed small selective effects (weak selection) and it is important to note where this is used. In fact, the only place we need this is to obtain the (approximate) linearity of individual fitness as a function of the phenotypes of others (eqn 2), but if fitness happens to be linear in phenotype, we do not require weak selection at this point. If W is linear in the P_k, then all our equations are valid, even if fitness effects are large. The other place where inclusive and direct fitness arguments typically use weak selection is in the calculation of relatedness. The covariances in eqns (4) and (6) are typically calculated by assuming phenotype is linearly related to genotype and then using the notion of identity by descent, and this last argument is valid only when the alleles are neutral. When selection is weak, these 'neutral' covariances provide a good approximation. Work is ongoing on a more complete examination of the effects of strong selection on inclusive and direct fitness.

The popularity of direct fitness in the theoretical literature in recent years reflects the fact that it is often mathematically more natural to formulate. But the inclusive fitness paradigm continues to have a powerful presence, no doubt because it positions the modeller as an agent choosing behaviour to maximize fitness. This 'individual as maximizing agent' analogy (Grafen, 1999) allows us to put ourselves in the position of an individual organism and ask: how can I maximize my inclusive fitness? Not only does this constitute a powerful theoretical construct, but it is also a natural question for us humans to ask, as in our day-to-day lives our behavioural decisions are typically optimal, albeit with regard to complex payoff functions.

Acknowledgements

PDT is supported by a grant from the Natural Sciences and Engineering Research Council of Canada. The authors thank Amy Hurford for her comments on an early draft, Jeff Fletcher for feedback and a number of specific comments and two referees for extensive and valuable comments.

References

Cheverud, J.M. 1984. Evolution by kin selection: a quantitative genetic model illustrated by maternal performance in mice. *Evolution* **38**: 766–777.

Day, T. 2001. Population structure inhibits evolutionary diversification under resource competition. *Genetica* **112–113**: 71–86.

Frank, S.A. 1997. Multivariate analysis of correlated selection and kin selection, with an ESS maximization method. *J. Theor. Biol.* **189**: 307–316.

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

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

Gardner, A. & West, S.A. 2006. Demography, altruism, and the benefits of budding. *J. Evol. Biol.* (in press).

Grafen, A. 1985a. Hamilton's rule ok. *Nature* **318**: 310–311.

Grafen, A. 1985b. A geometric view of relatedness. *Oxford Surv. Evol. Biol.* **2**: 28–89.

Grafen, A. 1999. Formal Darwinism, the individual-as-maximizing agent analogy and bet-hedging. *Proc. R. Soc. Lond. B* **266**: 799–803.

Fisher, R. A. 1930. The genetical theory of natural selection. Clarendon Press, Oxford.

Hamilton, W.D. 1964. The genetical evolution of social behaviour, I and II. *J. Theor. Biol.* **7**: 1–52.

Hamilton, W.D. 1971. Selection of selfish and altruistic behaviour in some extreme models. In: *Man and Beast: Comparative Social Behavior* (J. F. Eisenberg & W. S. Dillon, eds), pp. 59–91. Smithsonian Institute Press, Washington, DC.

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

Leturque, H. & Rousset, F. 2003. Joint evolution of sex ratio and dispersal: conditions for higher dispersal rates from good habitats. *Evol. Ecol.* **17**: 67–84.

Michod, R.E. & Hamilton, W.D. 1980. Coefficients of relatedness in sociobiology. *Nature* **288**, 694–697.
Orlove, M.J. & Wood, C.L. 1978. Coefficients of relationship and coefficients of relatedness in kin selection: a covariance form for the rho formula. *J. Theor. Biol.* **73**: 679–686.
Pen, I. 2006. When boys want to be girls: effects of mating system and dispersal on parent–offspring sex ratio conflict. *Evol. Ecol. Res.* **8**: 103–113.
Pen, I. & Taylor, P.D. 2005. Modeling information exchange in worker-queen conflict over sex allocation. *Proc. R. Soc. Lond. B* **272**: 2403–2408.
Perrin, N. & Mazalov, V. 2000. Local competition, inbreeding, and the evolution of sex-biased dispersal. *Am. Nat.* **155**: 116–127.
Price, G.R. 1970. Selection and covariance. *Nature* **227**: 520–521.
Price, G.R. & Smith, C.A.B. 1972. Fisher's Malthusian parameter and reproductive value. *Ann. Hum. Genet.* **36**: 1–7.
Queller, D.C. 1985. Kinship, reciprocity and synergism in the evolution of social behaviour. *Nature* **318**, 366–367.
Queller, D.C. 1992. A general model for kin selection. *Evolution* **46**: 376–380.
Rousset, F. 2004. *Genetic Structure and Selection in Subdivided Populations*. Princeton University Press, Princeton, NJ.
Seger, J. 1981. Kinship and covariance. *J. Theor. Biol.* **91**: 191–213.
Taylor, P.D. 1990. Allele frequency change in a class-structured population. *Am. Nat.* **135**: 95–106.
Taylor, P.D. 1992a. Altruism in viscous populations—an inclusive fitness model. *Evol. Ecol.* **6**, 352–356.
Taylor P.D. 1992b. Inclusive fitness in a homogeneous environment. *Proc. R. Soc. Lond. B* **249**: 299–302.
Taylor, P.D. 1996. Inclusive fitness arguments in genetic models of behaviour. *J. Math. Biol.* **34**: 654–674.
Taylor, P.D. & Frank, S.A. 1996. How to make a kin selection model. *J. Theor. Biol.* **180**: 27–37.
Wild, G. & Taylor, P.D. 2005. A kin-selection approach to the resolution of sex-ratio conflict between mates. *J. Theor. Biol.* **236**: 126–136.
Wilson, D.S., Pollock, G.B. & Dugatkin, L.A. 1992. Can altruism evolve in purely viscous populations? *Evol. Ecol.* **6**: 331–341.
Wright, S. 1943. Isolation by distance. *Genetics* **28**: 114–138.

Received 3 March 2006; revised 19 June 2006; accepted 21 June 2006