

THE CENTRAL CONCEPTS OF INCLUSIVE FITNESS

3000 word article in the Oxford University Press Encyclopedia of Evolution, January 2002.

Peter D. Taylor
 Dept. of Mathematics and Statistics
 Queen's University
 Kingston ON Canada K7L 3N6
 taylorp@post.queensu.ca
 telephone: 613 533 2434

Troy Day
 Dept. of Zoology
 University of Toronto
 Toronto ON Canada M5S 3G5
 dayt@zoo.utoronto.ca
 telephone: 416-946-5563

Why do the plants in a certain populations grow to a particular height (on average) and not to another? What determines the relative numbers of male and female offspring in a population? How should an individual allocate resources between growth and fecundity, or between foraging and defence? What proportion of an individual's offspring should disperse and attempt to breed far from home? These are examples of the questions we consider here, and the answers we seek are *evolutionary*—that “deviant” individuals with alternative traits have lower fitness than “normal” individuals who exhibit the established trait. For example, we might try to find evidence that a plant which was shorter than average would receive less sunlight or disperse less pollen, whereas a plant which was taller than average would consume too much energy in growth or would take too long to attain maturity and would have less reproductive success.

But there's an extra complication that enters here. It's possible that by growing a bit taller the individual would get *more* offspring, but overall the trait (of increased height) would be at a disadvantage. The reason is that the neighbours of the taller individual would have slightly less fitness and, because of limited dispersal, they might be related to the taller individual and thus share the “tallness” genes. As a result, this would contribute a negative fitness force to such genes. In short, if what's good for the individual is bad for the group, and if there is some within-group genetic similarity, the evolutionarily stable (ESS) level of the trait will be a compromise between the two. We can look at it another way too—at the ESS a shorter individual could have has less personal fitness, if its relatives had more. In this context, tallness would be called “selfish” and shortness would be called “altruistic” and the ESS represents a balance.

The central problem in modelling the evolution of social characters, those whose behaviour affects the fitness of neighbours, is to find a good way to account for the effect of these interactions. Here we will describe a remarkable approach to this problem known as the inclusive fitness method. This method gives us an expression which can tell us whether or not a deviant behaviour will increase in frequency. The analysis applies to a wide range of characters, but for concreteness we will work here with the example of plant height.

As a simple illustration, suppose that the height deviation is caused by a mutation at a certain locus which alters its owner's height, and thereby changes the fitness of the owner and of exactly one neighbour by the amounts b_0 and b respectively. To calculate the resulting change in the

frequency of the mutant allele in the next generation we must weight these two fitness changes by the “effectiveness” of the owner and the neighbour in propagating the mutant allele, that is by its frequency in their gametes. Suppose this is p_0 and p respectively. Then it would seem that the allele should increase in frequency exactly when $p_0 b_0 + pb$ is positive. This is the simple idea behind the inclusive fitness method, and with some elaboration, it works.

If we take the above alteration to be a decrease in height, then the fitness change to the focal individual is actually a cost and could be written as $b_0 = -c$. Then the condition $p_0 b_0 + pb > 0$ for the mutant allele to increase in frequency can be written $Rb > c$ where $R = p/p_0$ is called the relatedness between owner and neighbour. This formula is known as Hamilton's Rule and it tells us that an altruistic act is selectively favoured if the benefit to others, weighted by relatedness, exceeds the cost.

We now look more carefully at the above calculation. Let's take a simple population structure, and suppose that individuals occur in patches of 5, and that an individual's behaviour affects only those on the same patch. Suppose that a single deviant individual changes its own fitness by b_0 and changes the fitness of its i th patchmate by b_i ($1 \leq i \leq 4$) where we use some suitable fitness measure such as contributions to the effective gamete pool of the next generation. Let this gamete pool contain a total of T gametes, D of which are mutant.

Now select a random copy of the mutant allele, which we will suppose causes its owner, the *focal* individual, to be deviant. We calculate the change in frequency of the mutant allele *due to this one deviant individual* and to do this we need to know the frequency of the mutant allele in each of the 5 inhabitants of the patch. Let these frequencies be p_i ($0 \leq i \leq 4$) and let $\bar{p} = D/T$ be the average frequency of the mutant allele in the population as a whole. Then the changes in D and T are:

$$\Delta D = \sum p_i b_i \quad . \quad (1)$$

$$\Delta T = \sum b_i \quad . \quad (2)$$

where the sums are over all 5 inhabitants of the patch ($0 \leq i \leq 4$). There is an assumption here of "fair meiosis" whereby the gametes produced by any individual contain its alleles in the same proportion as found in its genotype.

Now we take the average E of (1) and (2) over all copies of the mutant allele. To get the final expression in (3) we employ a technical assumption that the b_i and p_i variables are uncorrelated.

$$E(\Delta D) = E(\sum p_i b_i) = \sum E(p_i) E(b_i) \quad (3)$$

$$E(\Delta T) = \sum E(b_i) \quad (4)$$

For the average in (3) and (4) to make sense, there must be a natural way of assigning, for each focal individual, an $i=1$ patchmate and an $i=2$ patchmate etc. This is essentially an assumption of spatial homogeneity of the population, that each individual has the same geometric configuration of patchmates. For example, the five individuals might be equally spaced around a circle, so that each individual has two “near” neighbours ($i = 1, 2$) and two “far” neighbours ($i = 3, 4$).

If the mutant allele happens to be rare (as is often assumed), an increase in numbers will correspond to an increase in frequency, and we can expect the sign of $E(\Delta D)$ to predict frequency change, but otherwise the average change in frequency must be calculated as:

$$\Delta\bar{p} = \frac{D + E(\Delta D)}{T + E(\Delta T)} - \frac{D}{T} = \frac{D + \Sigma E(p_i)E(b_i)}{T + \Sigma E(b_i)} - \bar{p} = \frac{\Sigma[E(p_i) - \bar{p}]E(b_i)}{T + \Sigma E(b_i)} = K \Sigma \frac{E(p_i) - \bar{p}}{E(p_0) - \bar{p}} E(b_i) \quad (5)$$

where K is a positive constant. This motivates the definition of the *inclusive fitness effect* of the deviant behaviour:

$$\Delta W_{IF} = \Sigma R_i E(b_i) \quad (6)$$

where

$$R_i = \frac{E(p_i) - \bar{p}}{E(p_0) - \bar{p}} \quad (7)$$

is *the average relatedness* of the focal individual to its i th neighbour. Note that $i=0$ is one of the summands in (4) (as the focal individual is one of those affected by the deviant behaviour), and the relatedness R_0 of the focal individual to itself is 1.

According to (5), the inclusive fitness effect has the same sign as the average allele frequency change due to the deviant behaviour of a focal individual, where an individual is selected to be focal according to its mutant allele frequency p_0 . For example, in a diploid population, an individual with 2 mutant alleles will have twice the probability of being selected as an individual with 1 mutant allele.

Now recall that we have assumed that only one mutant copy acts at a time. But what happens if we allow them all to act together? First of all we note that there's no reason to expect the inclusive fitness effect to necessarily represent the resulting change in allele frequency, because it contains no information about the fitness effect of simultaneous deviant behaviour from several individuals, and that can certainly make a difference to allele frequency change. For example, for a trait such as a warning call, it might be that an expression by several individuals serves the group no better (and might even do worse!) than a single call. On the other hand, a complex task requiring cooperation might have little effect if performed by one, but a large effect if performed by two. Some assumption is required and the right one turns out to be additivity of effect. For example, if a deviant height in one individual increases the fitness of the actor by 5% and decreases the fitness of the four others in the patch by 3% each, then a deviant height in two individuals should increase the fitness of each actor by 2% ($2 = 5 - 3$) and decrease the fitness of the three others in the patch by 6% each.

The fundamental theorem of inclusive fitness. With this assumption of additive fitness effects between actors, the fundamental theorem says that the mutant allele will increase in frequency when the inclusive fitness effect is positive, and will decrease when it is negative.

This is a remarkable result. It tells us that an "average" deviant individual can tell whether the deviant behaviour will increase or decrease in frequency by adding up the effects of its deviation on all affected individuals, each one weighted by its relatedness to the individual, and look at the sign of that expression. Its applicability is much wider than our simple example might suggest. The genius behind this result is that of the British biologist W.D. Hamilton in a 1964 paper. It is no exaggeration that the appearance of this result revolutionized the evolutionary modeling literature, and it is a measure of the depth of the result that this revolution has unfolded gradually

over the past 40 years, each insight into the power of the method giving rise in due course to others. This “power” is two-fold—conceptual and computational. First of all, expression (6) provides a heuristic which gives us a powerful way to analyze behaviour and make intuitive guesses. Secondly it gives us a way to make calculations when a straightforward accounting of all the different possible local (neighbourhood) configurations in which an allele might find itself is completely infeasible.

To implement the result we have to be able to calculate the fitness effects b_i and the relatedness coefficients R_i and we look at these one at a time.

The fitness effects. Typically there are different kinds of interactions, some with a greater effect than others. To model this situation, we often group interactions into categories and express fitness in terms of average behaviour over a category. As a simple example, suppose the fitness $W = W(x, y, z)$ of a plant depends upon its own height x , the average height y of its two near neighbours, and the average height z of its two far neighbours. Here we have three categories, the focal individual and its near and far neighbours. Note that we expect W to increase with x and decrease with y and z .

Now W will not usually be linear in x , y and z , but we can always get the additive effects required by the inclusive fitness method (at least approximately) by assuming that the behavioural deviations are small enough that a differential approximation is valid. Thus, we let the focal individual have height $x^* + dx$ where x^* is the established height and the behavioural deviation dx is assumed small. Then the first-order fitness effects are:

$$\text{effect on the focal individual:} \quad b_0 = \frac{\partial W}{\partial x} dx \quad (8)$$

$$\text{effect on each near neighbour} \quad b_y = \frac{\partial W}{\partial y} dy = \frac{\partial W}{\partial y} \frac{dx}{2} \quad (9)$$

$$\text{effect on each far neighbour} \quad b_z = \frac{\partial W}{\partial z} dz = \frac{\partial W}{\partial z} \frac{dx}{2} \quad (10)$$

where the partial derivatives are evaluated at the normal height, $x = y = z = x^*$. In each of (9) and (10) we note that the focal individual’s effect on the average category height is $dx/2$. The inclusive fitness effect is:

$$\Delta W_{\text{IF}} = b_0 + 2R_y b_y + 2R_z b_z = \left[\frac{\partial W}{\partial x} + \frac{\partial W}{\partial y} R_y + \frac{\partial W}{\partial z} R_z \right] dx \quad (11)$$

where R_y and R_z are the relatedness of the focal individual to near and far patchmates. Note that equation (6) has 5 summands but (11) has only 3—as it is over categories of individuals grouped according to like effects.

Interestingly enough, the expression in the square brackets reminds us of the expression we’d get for dW if we treated y and z as functions of x :

$$dW = \frac{dW}{dx} dx = \left[\frac{\partial W}{\partial x} + \frac{\partial W}{\partial y} \frac{dy}{dx} + \frac{\partial W}{\partial z} \frac{dz}{dx} \right] dx. \quad (12)$$

This provides a useful heuristic for the inclusive fitness calculation—treat the average phenotype y of each category as a function of focal phenotype, take the total derivative of the fitness function with respect to focal phenotype, replace dy/dx and dz/dx by the category-specific relatedness, and we get an expression for inclusive fitness.

Example: Parasite virulence. Completely analogous to the plant height model is the question of the level of aggressiveness of a parasite within a host—too low and it can't compete against others, too high and the life of the host and the group as a whole is endangered. A simple model capturing this tension between within-group competition and prudent use of resources is

$$W(x,y) = \frac{x}{y} G(y)$$

where x determines individual competitiveness, y is the average value of y within the group, and $G(y)$ measures the resources available to the group, and is supposed to be a decreasing function of y . Taking $dx = 1$, equation 11, when simplified, gives us:

$$\Delta W_{IF} = \frac{\partial W}{\partial x} + \frac{\partial W}{\partial y} R = \frac{G(x^*)}{x^*} (1 - R) + G'(x^*) R. \quad (13)$$

Here, R is the average relatedness to the group as a whole, and everything is evaluated at $x = y = x^*$. The first term of (13) treats the total resources G as fixed and measures the gain to the focal individual of its increased competitiveness, less the amount that must therefore be lost by a random group member, and the second term measures the reduction in total resources available (note that G' is negative). A simple model for G is $G(z) = 1 - cz$, where c is the rate at which average group competitiveness reduces available resources. If we put this into (13), and set the expression to zero, we get the ESS to be $x^* = (1 - R)/c$.

Relatedness. The calculation of relatedness (7) requires that we find the $E(p_i)$, the average mutant frequency in the genotype of the i th neighbour of a random focal allele. Note that since this average is taken over all mutant alleles and a focal individual is deviant in proportion to its number of mutant alleles, we can interpret $E(p_i)$ as the probability that the i th neighbour of an individual of deviant height is deviant.

The calculation of the $E(p_i)$ might seem to require complete knowledge of the local distribution of the mutant allele. It is remarkable that we can usually manage with a simple recursive argument based on the concept of identity by descent (IBD). Two genes are said to be IBD if they have a common ancestor (with no intervening mutation) some number of generations back. It turns out that under mild technical assumptions, two genes chosen in a pair of interactants will be either IBD or probabilistically independent.

Now choose a random mutant copy and choose a copy in the i th neighbour of its owner ($0 \leq i \leq 4$). Note that $E(p_i)$ is the average probability that the second copy is mutant. Let g_i be the probability that the two copies are IBD. If they are not IBD they are independent so that $E(p_i) = g_i + (1 - g_i) \bar{p}$. If we put these into (7) we get

$$R_i = \frac{g_i}{g_0} \quad (14)$$

A striking feature of this formula is that it displays relatedness as independent of allele frequency. Recall that we selected the mutant allele at random among those segregating at the height locus, and we would certainly expect that its fitness might well depend on its frequency. But (14) tells us that under appropriate assumptions this is not the case, and the inclusive fitness effect is (allele-)frequency independent. For example, (14) tells us that the relatedness between sibs in an outcrossed diploid population is $R = (1/4)/(1/2) = 1/2$.

To calculate the g_i in our plant population, we must specify the amount of gene flow between patches. As a simple example, assume that the population is effectively infinite and all pollen disperses widely but fertilized seeds always remain on their native patch. The force of the word "widely" is that uniting gametes are never IBD. We assume also that the four patchmates of a focal individual are indistinguishable, so that the $E(p_i)$ have a common value for $1 \leq i \leq 4$.

We calculate the g_i recursively. First of all g_0 is clearly $1/2$ as uniting gametes are never IBD. For $1 \leq i \leq 4$, $g_i = g$ is the probability random genes in two different individuals are IBD. This can be the case only if they both came from previous generation ovules, and that has probability $1/4$. In this case they were random alleles in "two" plants in the parental patch which might have been the same plant (with probability $1/5$). If they were the same plant, they are IBD with probability $1/2$, and if they were different plants, they are IBD with probability g . This gives us the recursion:

$$g = \frac{1}{4} \left[\frac{1}{5} \frac{1}{2} + \frac{4}{5} g \right]$$

which solves to give $g = 1/32$. Hence, $R = 1/16$.

This possibility of calculating R with a simple recursive argument transforms inclusive fitness from being just an interesting concept to being an important analytic tool for the study of evolutionary stability.

Reproductive value.

There is one final aspect of inclusive fitness we have not touched upon and this arises when the population consists of individuals of different classes, for example, age classes (adults and juveniles), sex classes (male and female) role classes (resident and satellite males, breeder and helper females), etc. In this case we have the problem of how fitness effects on individuals in different classes are to be compared, and that's what reproductive value is designed to accomplish.

For example, in our plant population, we might have different age classes (for a perennial species) or sexes (for a dioecious species). In this case a mutant allele might have a different effect in different classes, and a deviant individual of a certain class might affect the fitness of neighbours of different classes differently, so that the analysis can become complex. We handle this by first calculating the inclusive fitness effect ΔW_j of the mutant allele in a class j individual and then ΔW_{IF} is the average of these weighted by the *class reproductive values*, the relative contribution of each class to the future gene pool of the population. Secondly for each class j , we generalize (6) as:

$$\Delta W_j = \sum v_i R_i E(b_i) \tag{15}$$

where v_i is the *individual reproductive value* of neighbour i , defined as its relative contribution to the future gene pool of the population. Here, to obtain a common fitness yardstick, the b_i are typically measured as percentage increases. Examples of these calculations are found in the literature.

Comments and notes.

Evolutionary stability. Inclusive fitness plays an important role in the analysis of evolutionary stability (ESS ref) where we typically must determine whether alleles for a certain trait will increase in frequency. In our patch example, for the established height to be evolutionarily stable it must be the case that alternative heights have reduced fitness, and for small deviations, that implies a negative inclusive fitness effect. From (11),

$$\left[\frac{\partial W}{\partial x} + \frac{\partial W}{\partial y} R_y + \frac{\partial W}{\partial z} R_z \right] dx \leq 0 \quad (16)$$

for small dx both positive and negative. This can be the case only when we have the *equilibrium condition*:

$$\frac{\partial W}{\partial x} + \frac{\partial W}{\partial y} R_y + \frac{\partial W}{\partial z} R_z = 0 \quad (17)$$

In practice we would begin the analysis of the model with this equation evaluated at $x = y = x^*$ and solve it to obtain the ESS value of x^* .

Direct fitness. Our approach here was to calculate the fitness effect of a random mutant allele on all neighbouring copies and then add these all up over all copies of the allele. This is the standard inclusive fitness approach, but it can be contrasted with an equivalent "direct fitness" approach which calculates the fitness effect *on* a random mutant allele *by* all neighbouring copies and adds *these* all up. Sometimes one approach works better than the other, but the difference is really in how the accounting is done.

Approximations. The calculations of inclusive fitness are approximate for several reasons. First of all, the assumptions of additive genetic effects on fitness are crucial to the argument. These are of two kinds: within and between individuals. Within individuals we have assumed a "semi-dominance," that a mutant homozygote is twice as likely to be deviant as a mutant heterozygote. And between individuals, we have assumed that the effects several deviant individuals acting together can be added. These are of course unlikely to hold in practice, but they might hold to a good approximation for small effects. Secondly, the action of the deviant allele will usually disturb the distribution of the allele but relatedness is calculated in the "neutral" distribution in which the deviant allele is not permitted to act. Again, this approximation can be expected to be good when the effects of the deviant behaviour are small.

Haplodiploidy. Many of the early applications of inclusive fitness concern the phenomenon of haplodiploidy (refs) which induces very unusual patterns of relatedness whereby, for example, a female is more closely related to her sister than to her own daughter. This is one explanation for the evolution of eusociality, a phenomenon in which some individuals help to raise their siblings rather than having any reproductive output of their own. The roots of this hypothesis lie in Hamilton's rule which, as mentioned earlier, implies that we expect greater altruism to be

directed towards more closely related individuals, in this case towards sisters rather than daughters."

The covariance form of relatedness. The power of the inclusive fitness approach lies in its ability to reduce a complex allelic distribution to simple relatedness coefficients R which can be calculated by a recursive argument. In our expression (7) for R , the expectations are over the set of mutant alleles. An alternative formulation of relatedness is obtained by taking the expectation over *all individuals*. We obtain this form from (7) if we weight each interaction by the genotype of the focal individual. What we get is the covariance form:

$$R_i = \frac{E(p_0 p_i) - \bar{p}^2}{E(p_0^2) - \bar{p}^2} = \frac{\text{Cov}(p_0, p_i)}{\text{Cov}(p_0, p_0)} \quad (18)$$

Connections with group selection. When the population is divided into patches or groups, a fruitful point of view is to juxtapose individual fitness with the fitness of the group as a whole. It's often the case that what's good for one is bad for the other and evolution seeks a right balance. For example, in our height model, tall individuals do better than their neighbours, but groups of tall individuals do worse than groups of short individuals because more energy into growth means less energy into reproduction. Inclusive fitness tells us that high within-group relatedness should shift the balance towards the needs of the group.

An intuitive interpretation of high within-group relatedness is that individual similarity within a group should be much greater than individual similarity between groups. If we let p be the average allele frequency within each group, then $\text{Var}(p)$ is called the between-group variance and it is not hard to show that if we take the average of (18) over all members of the group, we get

$$E(R_i) = \frac{\text{Var}(p)}{\text{Var}(p_0)}. \quad (19)$$

which is the ratio of within-group variance to total variance:

LITERATURE CITED

- Grafen, A. 1984. Natural selection, kin selection and group selection. Pp 62-84 in Krebs, J.R. & Davies, N.B. eds. Behavioural Ecology 2nd Ed. Blackwell, Oxford.
- Hamilton, W.D. 1996. *The Narrow Roads of Gene Land*. The collected papers of W.D. Hamilton. Vol 1. Evolution of Social Behaviour. W.H. Freeman, New York, Oxford University Press. 552 pages.
- Hamilton, W. D. and May, R. M. 1977. Dispersal in stable habitats. *Nature* **269**: 578-581.
- Michod, R. E. and Hamilton, W. D. 1980. Coefficients of relatedness in sociobiology. *Nature* **288**: 694-697.
- Queller, D. C. 1992. A general model for kin selection. *Evolution* 46: 376-380.
- Taylor, P.D. and Frank, S. How to make a kin selection argument. *Journal of Theoretical Biology* 180 (1996) 27-37.