



## Multivariate Analysis of Correlated Selection and Kin Selection, with an ESS Maximization Method

STEVEN A. FRANK\*

*Department of Ecology and Evolutionary Biology, University of California, Irvine,  
CA 92697-2525, U.S.A.*

*(Received on 27 March 1997, Accepted in revised form on 14 July 1997)*

Kin selection coefficients are used in two distinct ways. First, these coefficients measure phenotypic correlations that affect the marginal costs and benefits of behaviors. For example, the phenotypic correlation in sex ratio produced by two females in an isolated patch influences the favored sex ratio. Second, kin selection coefficients describe genotypic correlations that measure fidelity of transmission. For example, a female values daughters vs. nieces according to genotypic correlations.

It is widely known that kin selection coefficients may be interpreted as phenotypic or genotypic correlations in different contexts. However, these different interpretations have never been fully separated, and their different roles have not been clearly explained.

I provide proofs of a generic analytical approach. The technique automatically separates phenotypic correlations among social partners from genotypic components of transmission. The result is a general method that can be derived from first principles and applied to multivariate problems in social evolution. I emphasize a simple, practical maximization method that can be used to calculate equilibrium conditions for complex social interactions.

© 1997 Academic Press Limited

### Introduction

Current understanding of kin selection and inclusive fitness follows from Hamilton's (1970) classic derivations (Grafen, 1985). Hamilton's paper includes two distinct approaches.

Hamilton first summed the fitness effects on an individual caused by all the phenotypes of neighbors (including the individual itself). This is often called direct or neighbor-modulated fitness, because the approach tracks the total effect of social interaction with neighbors on the direct fitness of a recipient individual.

Hamilton then reversed the direction of causality by summing fitness effects on neighbors caused by the focal individual's (actor's) phenotype. The evolutionary consequences of the actor's phenotype depends on the fitness effect on neighboring recipients, and the

probability that the recipients transmit the actor's phenotype to the next generation. The transmission probability is the coefficient of relatedness between actor and recipient. Hamilton defined the term "inclusive fitness" as the transmission weighted sum of the actor's effects on neighbors.

It would seem, at first glance, that direct and inclusive fitness provide equivalent measures for the same social processes. Differences have been noted, however, and the extent of the differences depends partly on how one chooses to define terms (Cavalli-Sforza & Feldman, 1978; Maynard Smith, 1980; Michod, 1982; Grafen, 1985; Queller, 1985, 1992a,b).

One difference is illustrated by a simple example. Suppose that an individual's fitness is increased by the cooperative tendency of neighbors. The neighbors are either unrelated (genotypically uncorrelated) conspecifics or members of a different species. The direct fitness of a relatively altruistic individual will rise as

\* E-mail: safrank@uci.edu.

its phenotypic correlation with neighbors increases. This increase occurs because an altruistic individual has helpful partners according to the phenotypic correlation between partners.

Direct fitness always captures the role of correlated phenotypes in differential fitness (selection). The correlations may be caused either by shared genealogy (genetic kinship) or by other factors. Direct fitness is simply an example of the multivariate selection analysis of correlated characters developed by Lande & Arnold (1983).

Quantitative genetic models of direct fitness typically provide only a rudimentary analysis of transmission (e.g. Queller, 1992a,b). The standard models assume that each individual transmits its own phenotype according to a simple measure of heritability, which is sufficient to describe the evolutionary response to selection.

In contrast with direct fitness, typical inclusive fitness models fail to measure the consequences of genetically unrelated social partners. An actor that aided an unrelated recipient would not obtain any inclusive fitness benefit, because the transmission weighting is zero. Inclusive fitness does, however, provide a richer description of heritability components than direct fitness. In a simple case, a mother may be related differently to sons and daughters in some genetic systems. Inclusive fitness handles this difference by treating sons and daughters as different components of transmission. Multiple pathways of transmission are not handled by the simple direct fitness techniques currently available.

I provide a solution for complex social evolution in which fitness is affected by many characters and many different classes of social partners. Classes include siblings, neighboring relatives, genetically unrelated conspecifics, and different species. The most general method is an extension of direct fitness to account for multiple components of transmission. I also provide connections between direct and inclusive fitness methods. My techniques provide tools for calculating evolutionary consequences and for finding equilibrium conditions [extending Taylor & Frank (1996)].

### Selection and Transmission

To separate selection and transmission, it is useful to work with breeding values from quantitative genetics theory. Let a set of potential predictors for a character be  $\mathbf{x} = (x_1, \dots, x_n)^T$ . Then any character  $z$  can be written as  $z = \mathbf{b}'\mathbf{x} + \delta$ , where the  $\mathbf{b}'$  are partial regression coefficients for the slope of the character  $z$  on each predictor,  $x$ , and  $\delta$  is the

unexplained residual. The additive, or average, effect of each predictor,  $b_j x_j$ , is uncorrelated with the residual,  $\delta$ .

In genetics the standard predictors are the hereditary particles (alleles). We write our standard regression equation for the character  $z$  of the  $i$ th individual as

$$z_i = \sum_j b_j x_{ij} + \delta_i = g_i + \delta_i \quad (1)$$

where  $g_i = \sum_j b_j x_{ij}$ , is the breeding value or additive genetic value for the character  $z$ . The breeding value is the best linear fit for the set of predictors,  $\mathbf{x}_i$ , in the  $i$ th individual. Each  $x_{ij}$  is the number of copies of a particular allele  $j$ , in an individual  $i$ . If we add the reasonable constraint that the total number of alleles per individual is constant,  $\sum_j x_{ij} = K$ , then the degree of freedom 'released' by this constraint can be used among the  $b$ 's to specify the mean of  $z$ . Thus, we can take  $\bar{z} = \bar{g}$ , and  $\bar{\delta} = 0$ .

To study the change in trait value over time, we write character value in the next time period as  $z' = g' + \delta'$ , with  $\bar{z}' = \bar{g}'$ . Thus, the change in average trait value is  $\bar{z}' - \bar{z} = \Delta\bar{z} = \Delta\bar{g}$ . The Price (1970, 1972) equation provides a method to obtain an exact analysis of  $\Delta\bar{g}$  with separation of selection and transmission (Frank, 1995). I use a modified version here that is convenient for my purposes.

Before proceeding it will be useful to rearrange the scheme for indexing individuals. Social evolution is conveniently studied by dividing the population into different behavioral classes, for example, mothers, daughters, nieces, social partners, and so on. We will be concerned with individuals that are members of a particular social class. Let  $i$  index social class, and  $ik$  be members of social class  $i$  with genotype  $k$ . The frequency of the  $k$ th type in the  $i$ th class is  $q_{ik} = q_i p_{ik}$ , where  $q_i$  is the abundance of the  $i$ th class and  $p_{ik}$  is the abundance of the  $k$ th genotype within the  $i$ th class. The standard identities for frequencies hold, in particular,  $\sum_{ik} q_{ik} = 1$  and  $\sum_k p_{ik} = 1$ .

We start with the simple definition

$$\Delta\bar{g} = \bar{g}' - \bar{g} = \sum q'_{ik} g'_{ik} - \sum q_{ik} g_{ik}.$$

I use the peculiar definitions of the Price equation with respect to the indices  $ik$ . The value of  $q'_{ik}$  is not obtained from the frequency of elements with index  $ik$  in the descendant population, but from the proportion of the descendant population that is derived from the elements with index  $ik$  in the parent population. If we define the fitness of element  $ik$  as  $w_{ik}$ , the contribution to the descendant population from type  $ik$  in the parent population, then

$q'_{ik} = q_{ik}w_{ik}/\bar{w}$  where  $\bar{w}$  is the mean fitness of the parent population.

The assignment of breeding values  $g'_{ik}$  also uses indices of the parent population. The value of  $g'_{ik}$  is the average breeding value contributed to descendants by parents with index  $ik$ . The change in breeding value for descendants of  $ik$  is defined as  $\Delta g_{ik} = g'_{ik} - g_{ik}$ .

The advantage of these definitions is that we can define parental classes,  $i$ , in any convenient way, and we can assign members of the descendant population to any parental class without the need to respect lineal descent. This is particularly useful for kin selection, in which one often assigns a fitness component of a neighbor to an actor whose phenotype controls the neighbor's fitness component. This will be made clear later.

We can use these definitions to write an exact expression for the change in character value

$$\begin{aligned}\Delta \bar{g} &= \sum_{ik} q'_{ik} g'_{ik} - \sum_{ik} q_{ik} g_{ik} \\ &= \sum_{ik} q_{ik} (w_{ik}/\bar{w}) g'_{ik} - \sum_{ik} q_{ik} g_{ik} \\ &= \text{Cov}(w, g')/\bar{w} + \sum_{ik} q_{ik} (g'_{ik} - g_{ik}) \\ &= \beta_{wg'} V_{g'}/\bar{w} + D_g.\end{aligned}\quad (2)$$

The term  $D_g$  is the change in the effect of alleles between the parent and offspring generations. I will assume  $D_g = 0$ . It is important to consider what assumptions this requires

$$\begin{aligned}D_g &= \sum_{ik} q_{ik} (g'_{ik} - g_{ik}) \\ &= \sum_i q_i \sum_k p_{ik} (g'_{ik} - g_{ik}) \\ &= \sum_i q_i (\tilde{g}'_i - g_i) \\ &= \tilde{g}' - \bar{g}.\end{aligned}$$

The second line is obtained by the prior definition,  $q_{ik} = q_i p_{ik}$ . The third line defines the average breeding value for the character among class  $i$  parents as  $g_i$ , and average breeding value among offspring assigned to class  $i$  parents as  $\tilde{g}'_i$ . Note that  $\tilde{g}'_i$  is defined with respect to parental frequencies,  $p_{ik}$ . Thus descendant values are weighted equally for all parents, ignoring

selection and differential fitness among parents. The final line defines  $\tilde{g}'$  as the average breeding value among descendants, taken with respect to parental frequencies.

Thus  $D_g$  summarizes the change in breeding value between ancestor and descendant. Recall that we may assign descendants to nonlinear ancestors. Although there are many ways for  $D_g$  to be equal to zero, two general assumptions capture the main issues.

First, if there is no variation in  $g_i$  among parental classes  $i$ , then the pattern by which descendants are assigned to parental class has no effect on  $D_g$ . This assumption is reasonable when the definition of class (e.g. sister, brother) is uncorrelated with the average breeding value of the class. When breeding value for the character is associated with class definition, then the particular details of the problem should make it clear how to calculate  $D_g$ .

The second assumption for  $D_g = 0$  is that the average effect of a particular genotype does not change between parent and offspring. Changes in environment or changes in allele frequency with non-additive allelic interactions can change the effect of genotype between parent and offspring. Changes in environment are never fully predictable. Changes in average effect can be calculated for particular assumptions about non-additivity, but the calculations are often tedious. Average effects are approximately constant over time when the population has little genetic variance and allele frequencies change by a small amount. This constancy of additive effects is equivalent to linearization of a dynamical system within a small analytical region, the standard assumption of local equilibrium analysis.

When  $D_g = 0$ , the direction of evolutionary change is completely summarized by the sign of the regression coefficient,  $\beta_{wg'}$  in eqn (2), thus

$$\text{sign}(\Delta \bar{g}) = \text{sign}(\beta_{wg'}) = \text{sign}(\text{Cov}(w, g')/\bar{w}). \quad (3)$$

The next section analyses factors that influence the direction of evolutionary change.

### Regression Equations for Direct Fitness

This section follows Fig. 1 to partition the regression  $\beta_{wg'}$  into components of correlated selection and kin selection. This method analyses variation in the fitness of class  $i$  members,  $w_{ik}$ , by starting with variation in the descendant genotype of class  $i$  members,  $g'_{ik}$ . Fitness is affected by phenotypes  $z_{ij}$ , which may be controlled by other classes. The method therefore tracks the direct fitness of each class as influenced by the behavioral phenotypes of social partners.

Total fitness in the population is  $w = \sum q_{ik} w_{ik} / \bar{w}$ , where dividing by  $\bar{w}$  normalizes total fitness to one. Each  $w_{ik}$  measures contribution to the following generations for genotype  $k$  of class  $i$ , and thus implicitly includes reproductive value weightings. We can separate reproductive value and reproductive success by  $w_{ik} / \bar{w} = v_i w_{ik}$ , where  $v_i$  is the reproductive value of an individual of class  $i$ . The frequency of parents in class  $i$  can be combined with the reproductive value of each member of class  $i$  as  $c_i = q_i v_i$ , with the  $v$ 's normalized so that  $\sum c_i = 1$ . The  $c$ 's are therefore the class reproductive values, the total contribution of class  $i$  to the following generations (Taylor & Frank, 1996).

The term  $W_{ik}$  is the reproductive success of genotype  $k$  of class  $i$ , normalized so that the class average,  $W_i$ , is the same for all  $i$  when there is no genetic variation. I assume that each  $W_i$  refers to a fitness component with a common reproductive value weighting. Taylor & Frank (1996) provide methods for handling complex demography and for assigning reproductive value to fitness components.

Expanding eqn (3),

$$\begin{aligned} \text{Cov}(w, g') / \bar{w} &= \sum_{ik} q_i p_{ik} (w_{ik} / \bar{w}) (g'_{ik} - \tilde{g}') \\ &= \sum_{ik} q_i p_{ik} (v_i W_{ik}) (g'_{ik} - \tilde{g}') \end{aligned} \quad (4)$$

where the average effect of parental alleles in the offspring generation is  $\tilde{g}' = \sum q_i p_{ik} g'_{ik}$ , with the summation over parental frequencies.

Two regression equations are required to complete Fig. 1

$$W_{ik} = \alpha_w + \sum_j \beta_{ij} z_{ijk} + \epsilon_w$$

$$z_{ijk} = \alpha_z + \rho_{ij} g'_{ik} + \epsilon_z.$$

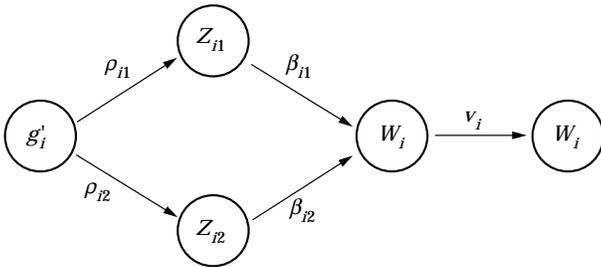


FIG. 1. Causal chain for the association between offspring breeding value,  $g'$ , and fitness,  $w$ . The pathway between  $g'$  and  $w$  is repeated for each  $i$ .

If we assume that all unspecified error terms in Fig. 1 are uncorrelated, then we can combine the regressions with eqn (4) to obtain

$$\text{Cov}(w, g') / \bar{w} = \sum_i q_i \left( v_i \sum_j \beta_{ij} \rho_{ij} \right) \sum_k p_{ik} g'_{ik} (g'_{ik} - \tilde{g}'). \quad (5)$$

Genotypic variation can be expanded as

$$\begin{aligned} \sum_k p_{ik} g'_{ik} (g'_{ik} - \tilde{g}') &= \sum_k p_{ik} g'_{ik} (g'_{ik} - g'_i + g'_i - \tilde{g}') \\ &= \sum_k p_{ik} g'_{ik} (g'_{ik} - g'_i) + g'_i (g'_i - \tilde{g}') \\ &= \sigma_i + g'_i (g'_i - \tilde{g}'). \end{aligned}$$

As discussed above under the assumptions for  $D_g = 0$ , I assume no variation among classes,  $g'_i - \tilde{g}' = 0$ . When there is variation among classes, the terms  $g'_i (g'_i - \tilde{g}')$  must be retained to describe selection among classes. Dropping the among class terms, and using the identity given above for reproductive value,  $c_i = q_i v_i$ , eqn (5) is

$$\text{Cov}(w, g') / \bar{w} = \sum_i c_i \sum_j \beta_{ij} \rho_{ij} \sigma_i. \quad (6)$$

The term  $\sigma_i$  is a measure of variance among the offspring of class  $i$ , taken with respect to parental frequencies. If we assume that the distribution of genetic variance is uncorrelated with the division of fitness components into classes, then  $\sigma_i$  is a constant with respect to  $i$ . This is reasonable because genetic variance is often the same within behaviorally defined classes such as sisters or brothers. If  $\sigma_i$  is constant with respect to  $i$ , then the direction of evolutionary change is

$$\text{sign}(\text{Cov}(w, g') / \bar{w}) = \text{sign} \left( \sum_i c_i \sum_j \beta_{ij} \rho_{ij} \right). \quad (7)$$

The regression terms,  $\beta$  and  $\rho$ , may change with directional evolution of the character under study. Thus the condition is primarily used for describing the instantaneous direction of change for a given set of assumed or measured regression parameters, or for providing equilibrium conditions.

### Regression Equations for Inclusive Fitness

Social evolution is commonly studied by inclusive fitness. The analysis begins with the individuals that

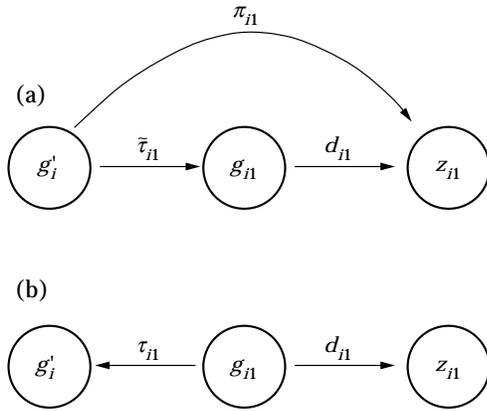


FIG. 2. Association between offspring breeding value,  $g'_i$ , and a character that influences parental fitness,  $z_{ij}$ . These diagrams expand the regression described by  $\rho_{ij}$  in Fig. 1. (a) Introduction of an intermediate explanatory variable,  $g_{ij}$ . The term  $g_{ij}$  is the breeding value for the character under study in an individual (actor) that influences the  $j$ th character of the  $i$ th parental class. (b) The inclusive fitness pathway. Inclusive fitness arguments usually ignore effects not associated with the actor, thus the  $\pi$  pathway is dropped. In order to place the actor at the center of causal explanation, the direction of the regression is changed between the actor and the recipient offspring. Thus,  $\tau$  is the component of heritability, or relatedness coefficient, given as slope of breeding value of recipient offspring on actor breeding value.

control phenotype. This point of view partitions into components the  $\rho_{ij}$  regression coefficients introduced in Fig. 1.

Two regression equations, summarized by the diagram in Fig. 2(a), can be combined to partition  $\rho_{ij}$  of Fig. 1

$$\begin{aligned} z_{ijk} &= \alpha_z + r_{ij}g_{ijk} + \pi_{ij}g'_{ik} + \epsilon_z \\ g_{ijk} &= \alpha_g + \tilde{\tau}_{ij}g'_{ik} + \epsilon_g \end{aligned}$$

yielding  $\rho_{ij} = \tilde{\tau}_{ij}d_{ij} + \pi_{ij}$ , under the assumption that  $\epsilon_g$  and  $\epsilon_z$  are uncorrelated.

The  $\rho_{ij}$  of Fig. 1 simply summarizes the association between a phenotype,  $z_{ij}$ , and offspring genotype,  $g'_i$ , without specifying how the phenotype is determined. The  $g_{ij}$  of Fig. 2 is the genotype of the actor that controls the phenotype  $z_{ij}$ , which in turn influences the fitness of the recipient. The explicit role of controlling genotype introduces a natural aspect of causal analysis. The difficulty with Fig. 2(a) is that the causal flow in the regression is from offspring genotype on the left to character on the right, via the controlling genotype as an intermediary. Inclusive fitness takes the controlling genotype's point of view, so that both phenotype,  $z_{ij}$ , and offspring genotype,  $g'_i$ , are expressed by regressions on the controlling genotype,  $g_{ij}$ .

Figure 2(b) shows the analysis taken fully from the controlling genotype's point of view. There are two

differences from Fig. 2(a). First, the term  $\pi_{ij}$  is dropped, ignoring extrinsic factors that cause an association between  $z$  and  $g'$ . Second, the direction of the regression between  $g_{ij}$  and  $g'_i$  is reversed, so that the new regression is expressed as offspring of recipient genotype on controlling genotype. This assigns variations in the abundance of descendant genotypes to the classes that control variations in phenotype. Put another way, pathways follow phenotypic cause rather than lineal descent.

When is the flip in the direction of regression valid in Fig. 2(b)? Substituting the regressions of Fig. 2(a) into eqn (6) and dropping  $\pi$  we have

$$\text{sign}(\text{Cov}(w, g')/\bar{w}) = \text{sign}\left(\sum_i c_i \sum_j \beta_{ij} d_{ij} \tilde{\tau}_{ij} \sigma_i\right). \quad (8)$$

We use the definition of regression to switch the direction of the  $\tau$  coefficient,  $\tau_{ij}\psi_{ij} = \tilde{\tau}_{ij}\sigma_i$ , where  $\psi_{ij}$  is the genetic variance within the class that controls phenotype  $j$  of class  $i$ . The genetic variance is always with respect to the character under study, which may differ from the phenotype  $z_{ij}$ . Substituting into eqn (8) yields

$$\text{sign}(\text{Cov}(w, g')/\bar{w}) = \text{sign}\left(\sum_i c_i \sum_j \beta_{ij} d_{ij} \tau_{ij} \psi_{ij}\right). \quad (9)$$

In the direct fitness formulation, I assumed that the variance among recipient offspring within class  $i$ ,  $\sigma_i$ , is independent of  $i$ , that is, we can take  $\text{Var}(g'_i)$  over individuals indexed by  $k$  as constant with respect to  $i$ . Convenient analysis of inclusive fitness requires that the genetic variance within controlling classes for the character under study is the same in all classes, that is,  $\psi_{ij} = \text{Var}(g_{ij})$  over individuals indexed by  $k$  is constant with respect to  $ij$ . If one takes these variance terms as constants, then the condition in eqn (9) can be simplified to

$$\text{sign}(\text{Cov}(w, g')/\bar{w}) = \text{sign}\left(\sum_i c_i \sum_j \beta_{ij} d_{ij} \tau_{ij}\right). \quad (10)$$

This provides the direction of evolutionary change by inclusive fitness.

Let us review the meaning of each term on the right side of eqn (10). The term  $c_i$  is the class reproductive value for the  $i$ th fitness component. The term  $\tau_{ij}$  is the slope of the transmitted genotypic value  $g'_i$  on the genotypic value of the controlling class,  $g_{ij}$ . This regression is a component measure of transmission fidelity, or heritability, which is frequently defined as a kin selection coefficient [for reviews, see Michod (1982), Grafen (1985), Queller (1992a,b); see Frank

(1997) for comparison of heritability components and kin selection coefficients]. The term  $\beta_{ij}$  is the slope of  $W_i$ , the  $i$ th fitness component, on  $z_{ij}$ , the  $j$ th character affecting  $W_i$ . This is the commonly used regression of fitness on multiple characters used by Lande & Arnold (1983). The term  $d_{ij}$  is the slope of  $z_{ij}$  on controlling genotype,  $g_{ij}$ . Genotype  $g_{ij}$  is measured with respect to the character under study, which may differ from the phenotype  $z_{ij}$ .

The inclusive fitness equation has advantages and disadvantages when compared with the direct fitness form in eqn (7). On the positive side, the  $\tau$  coefficients measure what can be thought of as components of heritability, all taken consistently from the causal actor's point of view. This is matched with the central role of actors in causing phenotypes of the recipient classes.

On the negative side, the inclusive fitness formulation excludes  $\pi$  in Fig. 2(a). This pathway measures the partial association between  $g'$  and  $z$  not mediated directly by actors who possess measurable breeding values for the character under study. For example,  $z_{ij}$  may be controlled by a different species or a different character from the one under study. It generally does not make sense to measure  $g$  on a different character or different species, thus  $\pi$  is the only pathway by which associations between  $g'$  and  $z$  can be measured. The inclusive fitness formulation also requires that genetic variances be the same within all actor classes.

These problems may not be serious within the context of inclusive fitness analysis. First, if the only goal is to measure the consequences of behavior by genetic relatives, then excluding  $\pi$  is acceptable but may prevent complete analysis of evolutionary consequences. An example is given below. Second, many simple inclusive fitness formulations study the behavior of a single actor class, thus the requirement that genetic variances are the same within all actor classes is irrelevant.

### Maximization Method

Equation (7) provides a direct fitness solution for the direction of evolutionary change. Equation (10) provides a matching inclusive fitness condition. These equations can be a bit tedious to apply when pursuing theoretical analysis. For example, the fitness component,  $W_i$ , may be expressed as a complex, nonlinear function of the traits  $\{z_{ij}\}$ , where the function is constructed as a natural description of the biology under study. Application of the regression solutions requires that the nonlinear interactions be reduced to a sum of linear components expressed in terms of the  $\beta$  coefficients.

I show that a simple maximization method can be used to obtain the direction of evolutionary change [extending Taylor & Frank (1996)]. This allows one to start with expressions for direct fitness components given in terms of all the characters affecting that class. Following eqn (7) maintains the direct fitness point of view, whereas eqn (10) transforms the direct fitness expressions into a summary of inclusive fitness.

For direct fitness, we begin with the definition of total fitness given earlier,  $w = \sum c_i W_i$ . We then differentiate  $w$  with respect to a randomly chosen allele in the descendant population. The population is assumed to be genetically monomorphic, except for rare genotypic variants of small average effect. The method thus defines

$$\frac{dw}{dg'} = \sum c_i \frac{dW_i}{dg'} = \sum_i c_i \sum_j \frac{\partial W_i}{\partial z_{ij}} \frac{dz_{ij}}{dg'}. \quad (11)$$

The left side has a natural interpretation as a slope of  $w$  on  $g'$ , matching the statement in eqn (3) that the direction of evolutionary change is determined by the sign of the regression  $\beta_{wg'}$ . Differentiation is an accurate measure of a regressed slope when variation is rare. The right side matches Fig. 1, with

$$\frac{\partial W_i}{\partial z_{ij}} = \beta_{ij}$$

$$\frac{dz_{ij}}{dg'} = \rho_{ij}.$$

Thus the sign of the derivative  $dw/dg'$  is sufficient for analysis of the direct fitness condition in eqn (7).

A maximization method for the inclusive fitness condition in eqn (10) can also be obtained. The steps for deriving the inclusive fitness form are a bit awkward, as in the regression model, but the causal point of view of actors is often valuable. Starting with eqn (11)

$$\frac{dw}{dg'} = \sum_i c_i \sum_j \frac{\partial W_i}{\partial z_{ij}} \frac{dz_{ij}}{dg'} = \sum_i c_i \sum_j \frac{\partial W_i}{\partial z_{ij}} \frac{dz_{ij}}{dg_{ij}} \frac{dg_{ij}}{dg'} \quad (12)$$

where the last line matches Fig. 2(a) with

$$\frac{dz_{ij}}{dg_{ij}} = d_{ij}$$

$$\frac{dg_{ij}}{dg'} = \tilde{\tau}_{ij}.$$

The goal is to define a differentiation operator that leads to eqn (10), without worrying about the steps that get there. Thus, following the transition from

Fig. 2(a) to Fig. 2(b), we define differentiation with respect to actor genotype by rearranging eqn (12) as

$$\frac{dw}{dg} = \sum_i c_i \sum_j \frac{\partial W_i}{\partial z_{ij}} \frac{dz_{ij}}{dg_{ij}} \tau_{ij} \quad (13)$$

where the right side matches Fig. 2(b) and eqn (10). The term  $g_{ij}$  is defined as follows. Randomly choose an individual of class  $i$ , and focus on the character  $z_{ij}$  that influences the individual's fitness. Then  $g_{ij}$  is a randomly chosen actor from the class that controls  $z_{ij}$  in the focal individual. The value of  $g_{ij}$  is the breeding value of the actor for the character under study, not the genotypic value affecting the character  $z_{ij}$ .

Equilibrium is obtained by analysing  $dw/dg' = 0$  for direct fitness, or  $dw/dg = 0$  for inclusive fitness, evaluated at a point with no genetic variation. Thus, at equilibrium, we take  $z_{ijk} = z_{ij}^*$  in the derivative and solve for the equilibrium values of  $z^*$ , checking that the condition provides a local maximum. A local equilibrium is often called an evolutionarily stable strategy, or ESS (Maynard Smith, 1982).

### Sex Ratio Example

I use a sex ratio example to illustrate the maximization methods and the difference between direct and inclusive fitness. The model is the standard analysis of local mate competition, described in general terms by Taylor & Frank (1996, and references in that paper). The two classes, or recipient fitness components, are male and female offspring. I use  $y$  for the sex ratio phenotype of a mother, and  $z$  for the average sex ratio phenotype of a local group. Sex ratio is the frequency of males per brood.

Each mother has fitness components for male and female offspring

$$W_m = \frac{y}{z} (1 - z)$$

$$W_f = 1 - y.$$

When there is no variation in phenotype at equilibrium,  $y = z = z^*$ , the fitness components have a normal value of  $1 - z^*$ . For direct fitness, equilibrium is analysed by studying  $dw/dg' = 0$  evaluated at  $y = z = z^*$ . Total fitness is  $w = c_m W_m + c_f W_f$ , with the  $c$ 's denoting class-specific reproductive values for males and females. Differentiating the components yields

$$\begin{aligned} \frac{dW_m}{dg'_m} &= \frac{\partial W_m}{\partial y} \frac{dy}{dg'_m} + \frac{\partial W_m}{\partial z} \frac{dz}{dg'_m} \\ &= \tilde{r}_m \left( \frac{1 - z^*}{z^*} \right) - \tilde{r}_m \left( \frac{s}{z^*} \right) \end{aligned}$$

and

$$\begin{aligned} \frac{dW_f}{dg'_f} &= \frac{\partial W_f}{\partial y} \frac{dy}{dg'_f} \\ &= -\tilde{r}_f \end{aligned}$$

where Fig. 3 shows the new parent-offspring terms,  $\tilde{r}_m = dy/dg'_m$  and  $\tilde{r}_f = dy/dg'_f$ , and the association between a male offspring and a random mother in the group,  $s\tilde{r}_m = dz/dg'_m$ . Solving  $dw/dg' = 0$  yields the equilibrium for the direct fitness model

$$z^* = \frac{c_m \tilde{r}_m (1 - s)}{c_m \tilde{r}_m + c_f \tilde{r}_f}. \quad (14)$$

I discuss this result after obtaining the equilibrium by the inclusive fitness method.

For inclusive fitness, the operator  $dg$  is interpreted as drawing a random individual from the recipient class, focusing on a phenotype that affects fitness, and picking the actor class that controls the phenotype. The breeding value of the actor is measured with respect to the character under study, not the phenotype affecting fitness. Thus, controlling genotype  $g_{ij}$  identifies breeding value in the actor that controls character  $j$  in fitness component (recipient class)  $i$ . Analysis by inclusive fitness of  $dw/dg = 0$  is summarized in Fig. 4. The analysis begins with

$$\begin{aligned} \frac{dW_m}{dg} &= \frac{\partial W_m}{\partial y} \frac{dy}{dg_{my}} \tau_{my} + \frac{\partial W_m}{\partial z} \frac{dz}{dg_{mz}} \tau_{mz} \\ &= r_m \left( \frac{1 - z^*}{z^*} \right) - R_m \left( \frac{1}{z^*} \right). \end{aligned}$$

The term  $dy/dg_{my} = 1$  is the slope of individual phenotype on individual breeding value. The term  $dz/dg_{mz}$  is the slope of a random individual on the breeding value that controls its phenotype. In this

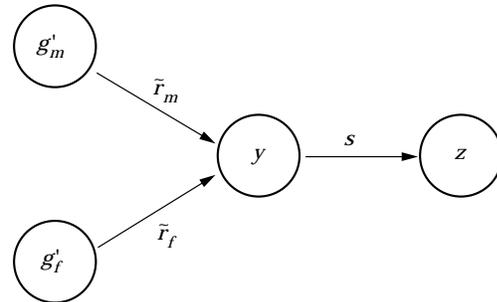


FIG. 3. The direct fitness model for sex ratio. The analysis begins with transmitted genotypic value,  $g'$ . The associations with mothers phenotype,  $y$ , are given by  $\tilde{r}$ . Direct fitness also requires a measure of association between  $g'$  and average group phenotype,  $z$ . I have assumed that this association can be expressed as the product of the association between  $g'$  and  $y$ , and the association between  $y$  and  $z$ .

case, each individual's phenotype is controlled by its own breeding value, so  $dz/dg_{mz} = 1$ . The term  $\tau_{my} = r_m$  is the slope of transmitted genotypic value through males on maternal genotype, and  $\tau_{mz} = R_m$  is the slope of a mother's transmitted genotypic value through males on a randomly chosen maternal genotype in the group.

For the female component

$$\begin{aligned} \frac{dW_f}{dg} &= \frac{\partial W_f}{\partial y} \frac{dy}{dg_{fy}} \tau_{fy} \\ &= -r_f \end{aligned}$$

where  $dy/dg_{fy} = 1$  is the slope of individual phenotype on individual breeding value, and  $r_f = \tau_{fy}$  is the slope of transmitted genotypic value through females on maternal genotype. Solving  $dw/dg = 0$  yields the equilibrium for the inclusive fitness model

$$z^* = \frac{c_m(r_m - R_m)}{c_m r_m + c_f r_f}. \quad (15)$$

The direct fitness and inclusive fitness models differ in the way genotypic value transmitted to males,  $g'_m$ , is associated with aspects of group phenotype. For direct fitness, the proper measure is the slope  $s\tilde{r}_m = dz/dg'_m$ . This associates group phenotype with offspring genotype. For inclusive fitness, the proper measure is  $R_m = dg'_m/dg_{mz}$ , the slope of offspring genotype on a random actor in the group.

The direct fitness model is more general. For example, the phenotypes of mothers may be correlated because of common environment or other factors not included in the breeding value. The direct fitness expression incorporates that additional association. In terms of Fig. 2(a), the direct fitness model retains  $\pi$ , or, in terms of Fig. 3, the association  $s$  measures both shared genetic and non-genetic factors. The direct fitness analysis shows that sex ratio evolution is controlled by two factors in Fig. 3: the phenotypic association between social partners,  $s$ , and the genotypic measures of transmission,  $\tilde{r}_m$  and  $\tilde{r}_f$ . Strong phenotypic associations,  $s$ , favor a low frequency of males independently of whether the association is caused by common genotype or extrinsic factors.

### Relatedness Coefficients

I have defined  $g'$  as the breeding value transmitted by a particular parent to its offspring. This is often equivalent to the breeding value carried by the parent's gametes. The general problem is, however, more complicated. For example, in the sex ratio model, the recipients of behavior are the mothers who

produce male and female offspring. These mothers also transmit the gametic value of their mates. Thus, a phenotype that influences the success of a mother through her sons also affects in the same way the success of the mother's mates through sons.

This technical problem of the joint effect on mothers and mates can be handled in one of two ways. First, we can maintain the strict definition of  $g'$  as the breeding value transmitted by a recipient rather than the breeding value of the whole offspring. In this case, each phenotype must be evaluated for its effect on the breeding values transmitted directly by mothers, and for its effect on the breeding values transmitted directly by fathers. The second approach treats  $g'$  as the breeding value of the whole offspring, which includes the contribution from the mother and her mate. This automatically accounts for the joint effect on mothers and mates without the need to bring fathers into the analysis. The second method is commonly used in the literature, and I used it implicitly in the previous section.

The same problem arises whenever a behavior influences the fecundity of a recipient. The behavior then has the same effect on the recipient and the recipient's mates. By contrast, a behavior that influenced the mating success of a male, but not his fecundity, would have no influence on the fitness of the male's mates. In this latter case it is important to use the strict definition of  $g'$  as transmitted breeding value rather than breeding value of whole offspring.

### The Paired-suicide Model

Charlesworth (1978) introduced a model of sibling interaction. He assumed that each individual has a probability of becoming sterile or committing a suicidal act, while simultaneously providing aid to a sibling. Clearly, if all siblings express the trait, then there is no one to receive the benefit.

This type of interaction is described by

$$W = 1 - y + b(1 - y)z$$

where  $y$  is the probability of the suicidal act by the focal individual,  $b$  is the benefit received by a non-suicidal individual that has a suicidal partner, and  $z$  is the partner's probability of the suicidal act.

Equilibrium is obtained by the standard application of eqn (11)

$$\frac{\partial W}{\partial y} + \rho \frac{\partial W}{\partial z} = 0$$

where  $\rho$  is the regression of partner phenotype,  $z$ , on the focal individual's transmitted genotype. Because the slope of the focal individual's phenotype on its

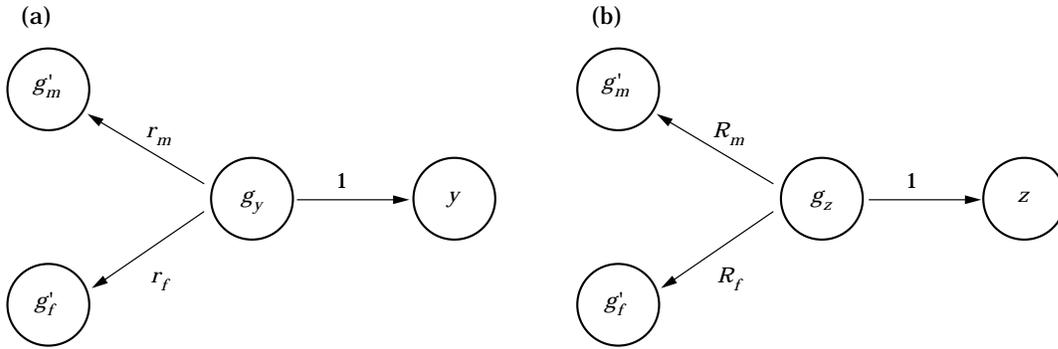


FIG. 4. The inclusive fitness model for sex ratio. (a) The term  $g_y$  is the genotypic value of a mother, with measures of transmission (relatedness) to sons and daughters. (b) Group phenotypes are studied by choosing at random a member of the group. The term  $g_z$  is the genotype that controls the phenotype of the chosen individual. The terms  $R_m$  and  $R_f$  are the transmission measures (relatedness) of the controlling genotype to random offspring in the group.

own genotype is one, the term  $\rho$  is a measure of phenotypic association between partners. Under the assumption that  $y$  and  $z$  have the same genetic basis, at equilibrium  $y = z = z^*$ , and the solution is

$$z^* = \frac{\rho b - 1}{b(1 + \rho)}.$$

This model shows that the evolution of sterility is limited when, in all individuals, the trait is controlled by the same genes and expressed in the same way. Some form of conditional expression is likely, in which a certain class of individuals expresses sterility more or less than a partner class. One simple approach is to assume that each individual is paired with one partner, and each partner expresses a different trait. For example, the smaller individual (class one) expresses  $y$  and the larger individual (class two) expresses  $z$ . The problem is now the joint evolution of two traits, in which the traits may be correlated.

The reproductive-value weighted fitnesses for the two classes are

$$c_1 W_1 = 1 - y + b(1 - y)z$$

$$c_2 W_2 = 1 - z + b(1 - z)y.$$

Standard application of eqn (11) gives the condition for the increase in  $y$  as

$$c_1 \frac{\partial W_1}{\partial y} + c_2 \rho_{yg_2} \frac{\partial W_2}{\partial y} + c_1 \rho_{zg_1} \frac{\partial W_1}{\partial z} + c_2 \rho_{zg_2} \frac{\partial W_2}{\partial z} > 0$$

where  $g'_i$  terms are the breeding value for trait  $y$  in class one and class two offspring. For simplicity, I assume the variance in breeding value for  $y$  and  $z$  is the same in all classes. The diagrams in Fig. 5 define the  $\rho$ 's in terms of genetic associations, under the assumption that all correlations are mediated by genetic pathways. From Fig. 5, the term  $\tau$  is the

genetic correlation between class one and two partners for a particular trait. This is a traditional form of the kin selection coefficient based on shared genotype. The term  $\lambda$  is the genetic association between  $y$  and  $z$  within individuals, a measure of linkage disequilibrium or pleiotropy.

When genetic variation is small, the condition for  $y$  to increase is given by  $dw/dg' > 0$ , and a similar condition can be obtained for the increase of  $z$ . The conditions for  $y$  and  $z$  to increase are, respectively,

$$\begin{aligned} -(1 + bz) + \tau b(1 - z) + \tau \lambda b(1 - y) \\ - \lambda(1 + by) > 0 \\ -(1 + by) + \tau b(1 - y) + \tau \lambda b(1 - z) \\ - \lambda(1 + bz) > 0. \end{aligned}$$

The natural equilibria of interest occur when one class always expresses sterility and the other class never

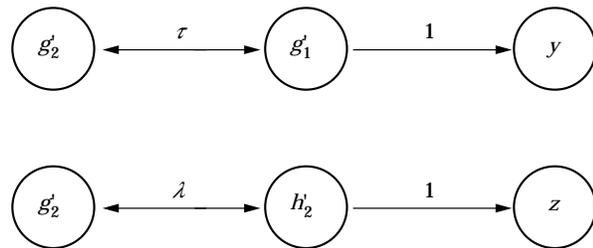


FIG. 5. Regressions between offspring breeding values and phenotypes for the paired-suicide model. I have assumed that all genetic variances are the same. Thus two-headed arrows are drawn between breeding values to show that the connection can be used in either direction. All regressions are mediated through offspring breeding values; other factors influencing association are ignored. The top shows  $\rho_{yg_2} = \tau$  where  $g'_i$  is the breeding value for  $y$  in class one and two offspring. The bottom shows  $\rho_{zg_2} = \lambda$ , where  $h'_i$  is the breeding value for  $z$  in class two offspring. Other regressions can be calculated from these assumptions, for example,  $\rho_{zg_1} = \rho_{yg_2} = \tau \lambda$ . The symmetry assumptions about variances and associations for each character are not required, and are made here for convenience to reduce the number of parameters.

expresses sterility, in particular, the points  $(y^*, z^*)$  of  $(0, 1)$  or  $(1, 0)$ . These corners are globally attracting when  $\tau b - 1 > 0$  and  $(\tau b - 1)/(b + 1) > \lambda$ . The first condition is a traditional Hamilton's rule form, with relatedness given by  $\tau$ , and the cost equal to one.

A strong, positive genetic correlation between traits, high  $\lambda$ , can prevent the corner equilibria because high expression in one class is matched to high expression in the other class. Another interesting point is that  $(\tau b - 1)/(b + 1) > \lambda$  is sufficient for local stability of these opposite corners. For example, if the standard kin selection coefficient is zero,  $\tau = 0$ , then  $-1/(b + 1) > \lambda$  is required for local stability. This occurs because a strong negative correlation between characters can cause class one to express fully the sterility trait,  $y$ , even when it benefits a genetically unrelated partner. The benefit to the partner, expressing  $z = 0$ , is sufficient to maintain a high value of  $y$  by genetic association. These simple roles of genetic correlation could be deduced intuitively. But the explicit method provides a clear connection between verbal reasoning about social evolution and analytical techniques.

### Conclusion

Relatedness turns out to be two separate phenomena. The first is the role of correlation among social partners. As the trait of an individual changes, its partners' traits change at a rate given by regression coefficients. Thus, we can calculate how a change in individual trait value is associated with changing social environment, and how social environment affects an individual's fitness. The regression (kin selection) coefficients summarize statistical information about partners.

The second aspect of relatedness concerns the fidelity of transmission. The value of a fitness component assigned to a particular individual must be weighted by the fidelity by which traits are passed

via that fitness component. This is necessary to measure the way in which trait value is associated with future share of the gene pool.

This paper provided a formal set of tools for the analysis of correlated selection and fidelity of transmission. The method can be applied to complex social problems with multiple actors and multiple traits. The differentiation techniques provide a useful approach for obtaining equilibrium conditions or approximating the direction of evolutionary change.

My research is supported by NSF grants DEB-9057331 and DEB-9627259.

### REFERENCES

- CAVALLI-SFORZA, L. L. & FELDMAN, M. W. (1978). Darwinian selection and "altruism". *Theor. Pop. Biol.* **14**, 268–280.
- CHARLESWORTH, B. (1978). Some models of the evolution of altruistic behaviour between siblings. *J. theor. Biol.* **72**, 297–319.
- FRANK, S. A. (1995). George Price's contributions to evolutionary genetics. *J. theor. Biol.* **175**, 373–388.
- FRANK, S. A. (1997). The Price equation, Fisher's fundamental theorem, kin selection, and causal analysis. *Evolution* (in press).
- GRAFEN, A. (1985). A geometric view of relatedness. *Oxf. Surv. Evol. Biol.* **2**, 28–89.
- HAMILTON, W. D. (1970). Selfish and spiteful behaviour in an evolutionary model. *Nature* **228**, 1218–1220.
- LANDE, R. & ARNOLD, S. J. (1983). The measurement of selection on correlated characters. *Evolution* **37**, 1212–1226.
- MAYNARD SMITH, J. (1980). Models of the evolution of altruism. *Theor. Pop. Biol.* **18**, 151–159.
- MAYNARD SMITH, J. (1982). *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.
- MICHOD, R. E. (1982). The theory of kin selection. *Ann. Rev. Ecol. Syst.* **13**, 23–55.
- PRICE, G. R. (1970). Selection and covariance. *Nature* **227**, 520–521.
- PRICE, G. R. (1972). Extension of covariance selection mathematics. *Ann. Hum. Genet.* **35**, 485–490.
- QUELLER, D. C. (1985). Kinship, reciprocity and synergism in the evolution of social behavior. *Nature* **308**, 366–367.
- QUELLER, D. C. (1992a). A general model for kin selection. *Evolution* **46**, 376–380.
- QUELLER, D. C. (1992b). Quantitative genetics, inclusive fitness, and group selection. *Am. Nat.* **139**, 540–558.
- TAYLOR, P. D. & FRANK, S. A. (1996). How to make a kin selection model. *J. theor. Biol.* **180**, 27–37.