A quantitative genetic model of reciprocal altruism: A condition for kin or group selection to prevail
(population genetics/quantitative inheritance/genotypic correlation/TIT FOR TAT)

KENICHI AKI
National Institute of Genetics, Mishima, Shizuoka-ken 441, Japan
Communicated by Moto Kimura, March 29, 1983

ABSTRACT A condition is derived for reciprocal altruism to evolve by kin or group selection. It is assumed that many additively acting genes of small effect and the environment determine the probability that an individual is a reciprocal altruist, as opposed to being unconditionally selfish. The particular form of reciprocal altruism considered is TIT FOR TAT, a strategy that involves being altruistic on the first encounter with another individual and doing whatever the other did on the previous encounter in subsequent encounters with the same individual. Encounters are restricted to individuals of the same generation belonging to the same kin or breeding group, but first encounters occur at random within that group. The number of individuals with which an individual interacts is assumed to be the same within any kin or breeding group. There are 1 + \(i\) expected encounters between two interacting individuals. On any encounter, it is assumed that an individual who behaves altruistically suffers a cost in personal fitness proportional to \(c\) while improving his partner’s fitness by the same proportion of \(b\). Then, the condition for kin or group selection to prevail is

\[
r > (c/b) - \left(1 - (c/b)\right)[C]/\left[1 + (1 - (c/b))\left[C + (T_b/V_b)\right]\right]
\]

if group size is sufficiently large and the group mean and the within-group genotypic variance of the trait value (i.e., the probability of being a TIT-FOR-TAT strategist) are uncorrelated. Here, \(C\), \(V_b\), and \(T_b\) are the population mean, between-group variance, and between-group third central moment of the trait value and \(r\) is the correlation between the additive genotypic values of interacting kin or of individuals within the same breeding group.

The right-hand side of the above inequality is monotone decreasing in \(C\) if we hold \(T_b/V_b\) constant, and kin and group selection become superfluous beyond a certain threshold value of \(C\). For small finite group sizes, the threshold value of \(C\) is also given in a kin-selection model.

Population genetic and game theoretic analyses (1–5) of reciprocal altruism (6) have indicated a threshold effect in which the trait is selectively disadvantageous while its frequency in the population is low but is advantageous when fairly common. Kin and group selection (7) are considered to be possible mechanisms for the initial spread of the trait (1, 2, 4, 5), although it is recognized that group structure becomes superfluous once the threshold has been crossed. (1 use the term group structure to mean subdivision of the population into temporary kin groups or persistent breeding groups.) However, these processes have not been explicitly modeled using population genetic methods. For example, Boorman and Levitt’s (2) analysis of the cascade effect invokes a group-structured population but not group selection. Recently, Brown et al. (3) suggested that reciprocation may increase when arbitrarily rare even in the absence of group structure. Their claim, although certainly true in the mathematical limit, remains to be proved realistic.

It is the purpose of this paper to present a quantitative genetic model of reciprocal altruism and derive an analytical formula for its evolution by kin or group selection. A behavioral trait as complex as reciprocal altruism is most probably determined by many genes and the environment and not monogenically as in the population genetic models of Boorman and Levitt (1, 2) and Brown et al. (3). The condition for kin or group selection to prevail is given in terms of the correlation, \(r\), between the additive genotypic values of interacting kin or of individuals within the same group (8–12). The advantage of this parameter is that, under certain assumptions, it is approximately equal to Wright’s coefficient of relationship (13) and, when so interpreted, its estimation in natural populations is straightforward. It is shown that, because of the threshold effect, the critical value of \(r\) can be substantially lower than that required for the evolution of unconditional altruism.

THE MODEL

For the moment, let us neglect group structure. The model presented below is quite similar to, but was developed independently of, that of Brown et al. (3).

Repeated opportunities for altruistic interaction between the same two individuals of the same generation probably occur in many species of animals. Let \(g(i)\) be the probability of \(i\) reencounters between two individuals after their first random encounter. On any encounter, an individual of the pair can behave altruistically toward the other, suffering a cost in personal fitness proportional to \(c\) while improving his partner’s fitness by the same proportion of \(b\). Or, alternatively, his behavior can be selfish, which in this case is defined as doing nothing. We assume that the number of individuals with which an individual interacts is the same and such that the product of this number and the proportionality constant is unity. The total fitness change to an individual is the sum of changes resulting from all encounters with all interactants [Hamilton’s model of benefit dispensation (8)]. The total fitness change can be substantial if the mean number of encounters, \(1 + \bar{t}\), is the mean of the probability distribution \(g(i)\), is large.

Axelrod and Hamilton (4) describe a strategy called TIT FOR TAT that, in my opinion, is an excellent model of reciprocal altruistic (or cooperative) behavior in man. This strategy is one of being altruistic on the first encounter and doing whatever the other individual did on the previous encounter on subsequent encounters with the same individual. Let \(C\) be a character value giving the probability of playing the strategy TIT FOR TAT as opposed to the alternative strategy of unconditional selfishness. Each individual “chooses” its strategy without regard to what others in the population are doing. Then, if we assume a sufficiently large population size for simplicity and first encounters to be at random, the fitness of an indi-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.
individual with character value \( C \) is given by
\[
W = W_0 + C\bar{C}(b - c)(1 + i) + C(1 - \bar{C})(-c) + (1 - C)\bar{C}b + (1 - C)(1 - \bar{C})0. \tag{1}
\]
Here, \( W_0 \) is an arbitrary standard fitness and \( \bar{C} \) is the mean character value. The second term on the right-hand side of Eq. 1 is derived by noting that the number of interactants who are TIT-FOR-TAT strategists is proportional to \( \bar{C} \) and that, in encounters with each such interactant, the altruistic individual will receive a net fitness increment proportional to \( b - c \) on each encounter (3, 14) for a total of \( 1 + i \) expected encounters. The meaning of the remaining three terms is also straightforward when we note that fitness change can occur only on the first encounter if either or both interactants are selfish. Eq. 1 reduces to
\[
W = W_0 + b\bar{C} + [c + (b - c)\bar{C}]. \tag{2}
\]

In the model presented here, each individual has perfect memory with regard to the altruistic or nonaltruistic nature of all previous interactants, restricting assistance to reciprocating individuals after the first encounter. But it cannot, as man appears to be capable of doing, learn as a bystander to the interactions between others. I have assumed that the number of interactants per individual is the same and such that the product of this number and the possible fitness change per encounter is a constant, being either \( b, b - c, 0, \) or \(-c\). The second half of the above assumption appears realistic because the resources available to each individual are limited. Later, when we consider group structure, I will assume that the number of interactants per individual is the same within each group although not necessarily so between groups. I will also explicitly consider the effect of finite group size in a later section.

Although it is possible to derive fitness in terms of the general-payoff matrix associated with reciprocal altruism (4, 6) (see ref. 3 for how this can be done), I use the two-parameter model (3, 14) for simplicity and to permit direct comparison of the results with Hamilton's condition for the evolution of unconditional altruism (8–11).

**ADDITIVE POLYGENIC FORMULATION**

Let \( C \) be determined by many genes of small effect acting additively within and between loci and an independent environmental deviation. We assume that the distribution of \( C \) lies entirely within the boundaries 0 and 1 at all times. A more general treatment would require us to define an underlying variable with thresholds at 0 and 1. Thus caution must be exercised when inferring properties of the two boundaries.

The change per generation in the population mean \( \bar{C} \) of the trait value under selection is
\[
\Delta \bar{C} = 2 \sum_k \sum_i p_i^{(k)} \gamma_i^{(k)} w_i^{(k)} / \bar{W}, \tag{3}
\]
where \( p_i^{(k)} \) is the frequency, \( \gamma_i^{(k)} \) is the average effect on the trait value, and \( w_i^{(k)} \) is the average excess in fitness of the \( k \)th allele at the \( k \)th locus contributing to the trait and \( \bar{W} \) is the mean Wrightian fitness of the population given by
\[
\bar{W} = W_0 + (b - c)\bar{C} + (b - c)\bar{C}^2. \tag{4}
\]
Eq. 3 is the discrete-time version of Crow and Nagylaki's (15) continuous-time formula for the change of a character correlated with fitness, restricted to the case of genes of additive effect (11). After straightforward algebraic manipulation (see ref. 11 for relevant methods), we obtain
\[
\Delta \bar{C} = (c - (b - c)\bar{C})V / \bar{W}. \tag{5}
\]
where \( V \) is the genotypic variance.

We note three properties of Eq. 5. (i) If \( i = 0 \), the model and hence Eq. 5 reduce formally to the case of unconditional altruism (11), provided we redefine \( b \) and \( c \) to be lifetime changes in fitness. (ii) A threshold exists if \( c < (b - c)i \); TIT FOR TAT is favored when
\[
\bar{C} > c / [(b - c)\bar{i}] \tag{6}
\]
and selected against when the opposite inequality holds. We are assuming that \( b > c \). The position of this threshold is lower for smaller values of \( c/b \) and greater values of \( i \). Inequality 6 is identical to equation 4 of Brown et al. (3) when the following appropriate correspondences are made: \( r = \bar{C}, \alpha \propto 1 + i, \beta \propto 1 + 1, T \propto b, R \propto b - c, P \propto 0, \) and \( S \propto -c \). When \( i \to \infty \), the threshold certainly converges to 0, as noted by Brown et al. (iii) If we set \( \bar{C} = 1 \), we have \( i \propto c / (b - c) \) as the condition for local stability of fixation of TIT FOR TAT. The method is valid, since we are in effect considering a small perturbation away from the boundary. If \( g(i) \) is the geometric distribution—i.e., \( g(i) = w(1 - w) - i \)—this inequality reduces to \( w \geq c/b \). The same condition results (14) on substitution of \( S \propto -c, \alpha \propto 1, \beta \propto 1 + 1, T \propto b, R \propto b - c, P \propto 0, \) and \( S \propto -c \) [Hamilton's model of benefit dispensation (3, 14)] in inequality 1 of Axelrod and Hamilton (4). However, note that whereas TIT FOR TAT is not an evolutionarily stable strategy if the number of reencounters is determinate (4)—i.e., \( g(i) = \delta_j \) for some \( j \)—fixation of TIT FOR TAT can still be locally stable in the genetic model presented here.

**KIN OR GROUP SELECTION**

Let us now consider a group-structured population with the size of each group and the number of such groups (either temporary kin or persistent breeding groups) assumed to be sufficiently large. We can define analogs of Eqs. 1–5 with the variables \( \bar{C}, V, \) and \( \bar{W} \) replaced by corresponding \( s \)-subscripted variables \( \bar{C}_s, V_s, \) and \( \bar{W}_s \); these equations describe the dynamics of selection within the \( st \)th group.

Selection between groups is assumed to occur by differential proliferation in proportion to mean fitness. That is, if \( P_s \) and \( P'_s \) are the relative sizes of the \( st \)th group before and after selection, then
\[
P'_s = P_s \bar{W}_s / \bar{W}, \tag{7}
\]
where \( \bar{W}_s \) is the mean fitness of the \( st \)th group given by the analog of Eq. 4 and
\[
\bar{W} = \sum_s P_s \bar{W}_s, \tag{8}
\]
is the mean fitness of the whole population.

The change per generation in \( \bar{C} \), the population mean of the trait value, due to the combined effects of within- and between-group selection is given by (11)
\[
\Delta \bar{C} = \sum_s P'_s \Delta \bar{C}_s + \sum_s \bar{C}_s \Delta P_s. \tag{9}
\]
Such an approach in which the effect of selection is divided into a within-group component and a between-group component has been shown by previous authors (9, 11, 16–18) to be equally applicable to kin and group selection.

In the kin-selection model (11), the kin groups are assumed to be formed anew at the beginning of each generation. On the other hand, in the group-selection model, the breeding groups are assumed to persist over generations. Breeding groups that grow too large are assumed to split, and those that become too small to go extinct, although these processes are not explicitly modeled. (See ref. 10 for details of the group-selection model.
considered here, in particular, migration between groups.) The model of group selection considered here is very simple. More elaborate models such as Wright's shifting balance theory (10) may be more realistic. The relevance of this model to reciprocal altruism is clear, because within each group locally stable equilibria exist if \( C = 0 \) and \( C = 1 \) and a threshold exists at some intermediate value of \( C \) provided \( c < (b - c)i \).

Substituting Eqs. 7 and 8 and the \( s \)-subscripted analogs of Eqs. 4 and 5 into Eq. 9, we derive

\[
W \Delta C = -cV_w + (b - c)V_b + \left[ \sum_i P_i \bar{C}_i V_i + \sum_i P_i \bar{C}^2_i - \bar{C} \right] \tag{10}
\]

where \( V_w = \sum_i P_i V_i \) and \( V_b = \sum_i P_i (\bar{C}_i - \bar{C})^2 \) are the within- and between-group variances. If \( i > 0 \) and \( C > 0 \), then the third term on the right-hand side of Eq. 10 is positive because \( \bar{C}_i \) and \( V_i \) are nonnegative quantities. Thus, reciprocal altruism can evolve—i.e., \( \Delta C > 0 \)—even when unconditional altruism cannot, if we assume that the ratio of \( V_w \) to \( V_b \) is not significantly different for either trait under selection (see below).

The correlation between genotypic trait values in a group is

\[
r = V_b/(V_b + V_w). \tag{11}\]

After some manipulation and in terms of \( r \), Eq. 10 can be rewritten as

\[
\Delta \bar{C} \propto \left[ 1 + (1 - c/b)[\bar{C} + T_b/(V_b)]r - (c/b) \right. \\
\left. + (1 - c/b)[\bar{C} + \text{Cov}(\bar{C}_n, V_n)/(V_b + V_w)] \right]. \tag{12}\]

where \( \text{Cov}(\bar{C}_n, V_n) = \sum_i P_i (\bar{C}_i - \bar{C})(V_i - \bar{V}_w) \) is the covariance of the means and the genotypic variances of each group and \( T_b = \sum_i P_i (\bar{C}_i - \bar{C})^2 \) is the third central moment of group means.

Although \( r \) is here defined as an intraclass correlation, it is approximately equal to the product moment correlation because group size is assumed to be sufficiently large; in fact, the latter is the relevant correlation as explicit consideration of finite group size shows (see ref. 11 and below). The coefficient \( r \) is a measure of the genetic similarity of individuals within a group or, alternatively, the genetic differentiation of groups. In particular, if, as we henceforth assume, per locus selection coefficients are small and linkage equilibrium holds between the relevant loci, it is approximately equal to Wright's coefficient of relationship (13).

As the positive values of \( r \) imply that first encounters between two individuals are assortative with respect to altruistic tendency in the whole population, although by assumption they are random within the group.

Henceforth, I put \( \text{Cov}(\bar{C}_n, V_n) = 0 \), an assumption often made in the quantitative genetics literature. I also assume that the magnitude of \( T_b/V_b \) is small (which is not unreasonable because we are assuming many loci of small effects) but will retain the term. Note that \( T_b \) is a measure of the skewness of the distribution of group means and is 0 for a symmetrical distribution such as the normal. From Eq. 12, the condition for reciprocal altruism to evolve becomes

\[
r > (c/b - 1 - (c/b)b)\bar{C}/ \left[ 1 + (1 - c/b)b(\bar{C} + T_b/V_b) \right]. \tag{13}\]

I note five properties of inequality 13. (i) On setting \( i = 0 \), inequality 13 reduces to Hamilton's rule (8-11), as should. (ii) If \( \bar{C} = 0 \) and \( T_b = 0 \), the right-hand side reduces to \( c/b \), indicating that in this case the condition for initial increase of reciprocal altruism is as stringent as for unconditional altruism.

Note that the inclusive fitness argument in evolutionary game theory gives the same result (4). (iii) However, when \( \bar{C} \) is small,

\[
T_b/V_b \text{ may not be negligible in comparison. In particular, if } T_b > 0 \text{ while } \bar{C} \text{ is small, this will promote kin and group selection. Note that } \bar{C} + (T_b/V_b) \text{ is nonnegative, so that the right-hand side is positive for small values of } \bar{C}. \]

(iv) The right-hand side of inequality 13 is monotone decreasing in \( \bar{C} \) when \( T_b/V_b \) is held constant. That is, as more and more individuals in the population become TIT-FOR-TAT strategists, less and less group differentiation as measured by \( r \) is required for positive selection. (v) When \( \bar{C} \) crosses the threshold given by inequality 6, group differentiation is no longer necessary (in the extreme case, \( r \) can be 0). Recall that the position of this threshold is lower for smaller values of \( c/b \) and greater values of \( i \). However, if \( \text{Cov}(\bar{C}_n, V_n) \) is not 0, the position of the threshold is not the same as in inequality 6.

In Fig. 1 are plotted critical values of \( r \) as a function of \( \bar{C} \) for various sets of parameter values \( c/b \) and \( i \). It is assumed here that \( T_b = 0 \).

###FINITE GROUP SIZE

The above formulae, which assume large group size, are good approximations when the groups are breeding units such as troops of primates whose membership averages several tens of individuals (20, 21). However, in higher vertebrates, temporary kin groups, in particular sibships, may comprise no more than two or three individuals. In such a case, it becomes necessary to explicitly consider finite group size. In the following, I restrict attention to kin groups persisting at most one generation, a policy that renders the simplifying assumption of same group size \( N \) (at the time of formation) fairly realistic.

The derivations involve a bit more algebra than above but are essentially identical and therefore are not shown. The difference arises because we must substitute \( C' \) for \( \bar{C} \) in Eq. 1, where

\[
T_b/V_b \text{ may not be negligible in comparison. In particular, if } T_b > 0 \text{ while } C' \text{ is small, this will promote kin and group selection. Note that } C' + (T_b/V_b) \text{ is nonnegative, so that the right-hand side is positive for small values of } C'. \]

Fig. 1. Critical values of \( r \) for reciprocal altruism to evolve by kin or group selection are plotted against \( C' \) for various sets of the parameters \( c/b \) and \( i \) (inequality 13 with \( T_b = 0 \)). \( r \), Correlation between the additive genotypic values of individuals within the same group; \( C' \), proportion of reciprocal altruists in the whole population; \( i \), mean number of reencounters between two individuals; \( c/b \), cost/benefit ratio for each altruistic act. The values of \( c/b \) are taken to be 1/4, 1/3, and 1/2; the intersection of a curve with the \( r \) axis corresponds to its \( c/b \) value. In each triplet of curves with the same \( c/b \) value, the values of \( i \) are 1 for the top, 2 for the middle, and 3 for the bottom curve.
\( \bar{C} \) is defined by
\[
C + (N - 1)\bar{C} = N\bar{C}. \tag{14}
\]
That is, \( \bar{C} \) is the mean trait value among all individuals in the group other than the individual whose fitness we are computing. There are two remarks worth making. First, it is necessary to assume that the environmental variance associated with each genotype is the same. Second, the intraclass correlation \( t \) and the product moment correlation \( r \) must be distinguished, where the two are related as
\[
1 + (N - 1)r = Nt. \tag{15}
\]
Corresponding to Eq. 12, we have
\[
\Delta \bar{C} \propto (1 + (c/b)\bar{C}) (T_b/V_b) - (T_w/V_w)\left| r - (c/b) + (1 - (c/b)\bar{C} + (1 - 2/N)\text{Cov}(\bar{C}, V_i)/V_b + N_w + (T_b/V_b - T_w/V_w)/N, \tag{16}
\]
where \( T_w \) is the average within-group third central moment of genotypic values. Note that Eq. 16 reduces to Eq. 12 when \( N \to \infty \). Also, if we are prepared to assume that \( \text{Cov}(\bar{C}, V_i) = T_b = T_w = 0 \), then the effect of finite group size disappears. (These are computed for the finite size.) In general, the dynamics may be more complicated and should be clarified by numerical or simulation studies.

**DISCUSSION**

Many authors have entertained the possibility that the group structure of early human populations was important in developing cooperative and altruistic behavior toward members of the group. In a recent study of genetic correlations in troops of primates based on variation in blood protein gene frequencies, it was suggested (unpublished data) that the best-informed guess of the value of \( r \) in bands of Pleistocene hominids was 10–15%. However, the genetic correlation among adult males may have been higher, if males were the sedentary sex as they are in the majority of contemporary human societies (22). In that study also, the possibility was tested that genetic correlations differ between adults of the two sexes in Japanese monkeys, where a sex asymmetry in dispersal is also observed. The difference was in the direction expected but not statistically significant. This is perhaps too low a correlation to have permitted the evolution of unconditioned group altruism. Based on purely theoretical considerations, Haldane (23) reached the similar conclusion that man does not possess many genes making for altruism of a general kind—i.e., altruism directed at nonspecific members of the group.

However, the theoretical results of this paper do suggest that group selection may have played some part in the evolution of reciprocal altruism among our ancestors. The curves shown in Fig. 1 suggest the following scenario. While the trait is rare, kin selection, especially between groups of sibs, may have to be invoked if the cost/benefit ratio is high. Alternatively, neutral alleles that increase in frequency by random genetic drift may acquire the novel function of determining altruistic tendency due to a sudden change in the environment and thus obviate the necessity for intense kin selection. When the proportion of TFT-FOR-TAT strategists in the whole population has increased somewhat, the consequent reduction in the requisite genetic correlation may permit the evolution by group selection of reciprocal cooperation between nonspecific members of the group. Finally, when the threshold has been crossed, it pays to be cooperative toward all individuals in the population. A point that cannot be stressed too strongly is that, once the frequency of TFT-FOR-TAT strategists has increased beyond the threshold, there is no going back in the evolutionary sense, even when, as is probably true of modern man, between-group genetic differences have greatly diminished due to increased migration.

Darwin invoked group selection to explain the existence of human heroism, which he regarded as a form of reciprocal altruism (24). His insight is, as usual, amazing.

This is a contribution no. 1,477 from the National Institute of Genetics. This work was supported in part by a Grant-in-Aid for Special Project Research on Biological Aspects of Optimal Strategy and Social Structure from the Japan Ministry of Education, Science and Culture.