Unifying the Theories of Inclusive Fitness and Reciprocal Altruism

Jeffrey A. Fletcher[1,2,*] **and Martin Zwick**[2,†]

1. Department of Zoology, University of British Columbia, #2370-6270 University Boulevard, Vancouver, British Columbia V6T 1Z4, Canada;
2. Systems Science PhD Program, Portland State University, Portland, Oregon 97207

Submitted December 19, 2005; Accepted May 31, 2006; Electronically published July 14, 2006

Online enhancement: appendix.

abstract: Inclusive fitness and reciprocal altruism are widely thought to be distinct explanations for how altruism evolves. Here we show that they rely on the same underlying mechanism. We demonstrate this commonality by applying Hamilton's rule, normally associated with inclusive fitness, to two simple models of reciprocal altruism: one, an iterated prisoner's dilemma model with conditional behavior; the other, a mutualistic symbiosis model where two interacting species differ in conditional behaviors, fitness benefits, and costs. We employ Queller's generalization of Hamilton's rule because the traditional version of this rule does not apply when genotype and phenotype frequencies differ or when fitness effects are nonadditive, both of which are true in classic models of reciprocal altruism. Queller's equation is more general in that it applies to all situations covered by earlier versions of Hamilton's rule but also handles nonadditivity, conditional behavior, and lack of genetic similarity between altruists and recipients. Our results suggest changes to standard interpretations of Hamilton's rule that focus on kinship and indirect fitness. Despite being more than 20 years old, Queller's generalization of Hamilton's rule is not sufficiently appreciated, especially its implications for the unification of the theories of inclusive fitness and reciprocal altruism.

Keywords: conditional behavior, Hamilton's rule, iterated prisoner's dilemma, kin selection, mutualism, synergy.

* E-mail: fletcher@zoology.ubc.ca.

† E-mail: zwickm@pdx.edu.

Am. Nat. 2006. Vol. 168, pp. 252–262.

More than 40 years ago, Hamilton developed an explanation for the evolution of altruism among relatives based on the idea of inclusive fitness (Hamilton 1963, 1964, 1970, 1972, 1975). His most famous result, known as Hamilton's rule (HR), is usually interpreted as specifying the conditions under which the indirect fitness of altruists (due to helping relatives have more offspring) sufficiently counterbalances the immediate self-sacrifice of altruists. In this way, the altruistic trait can increase overall. This mechanism is also known as kin selection (Maynard Smith 1964).

Twenty-five years ago, Axelrod and Hamilton (1981) launched a still vigorous area of research (for reviews, see Dugatkin 1997; Sachs et al. 2004; Doebeli and Hauert 2005) in which computer-based models of the iterated prisoner's dilemma (IPD) are used to study the evolution of cooperation via reciprocal altruism (Trivers 1971). In their article, Axelrod and Hamilton suggest that there are these two alternative explanations for the evolution of cooperative traits: when the benefits of altruism fall to relatives, cooperation can evolve by inclusive fitness, and when benefits fall to nonrelatives, cooperation can evolve by reciprocal altruism.[1] A third theory for the evolution of altruism, based on multilevel (or group) selection (Wilson 1975; Wade 1978), is not addressed directly by Axelrod and Hamilton (or in this article), but several researchers have demonstrated the underlying unity between inclusive fitness and multilevel selection theories (Wade 1980; Breden 1990; Queller 1992*b*; Frank 1998; Sober and Wilson 1998). For reasons discussed below, reciprocal altruism is often left out of these unification efforts. Here we focus on this missing piece: the unification of the theories of inclusive fitness and reciprocal altruism.

It is understandable that Axelrod and Hamilton (1981) do not suggest that HR could apply to reciprocal altruism and their IPD models. In addition to assuming that players are unrelated, these models involve conditional strategies (genotype/phenotype differences) and nonadditive fitness

[1] We use cooperation and altruism synonymously because cooperation in our models involves an immediate altruistic sacrifice in fitness that provides relative fitness benefits to recipients.

functions that HR could not accommodate. Yet just a few years later, Queller (1985) developed a version of HR that handles both of these issues, and in his article, Queller suggests that his version could apply to reciprocal altruism. More recently, Sober and Wilson (1998) suggest a unification of inclusive fitness and reciprocal altruism theories. They show how additive versions of the prisoner's dilemma (PD) correspond to fitness functions used in inclusive fitness models, but they do not address the two critical issues mentioned above: genotype/phenotype differences when behaviors are conditional and nonadditivity. Frank (1994, 1998) notes that regression coefficients between species in his model of mutualism measure reciprocity and are similar to coefficients of relatedness in inclusive fitness models, but he also does not address these two issues. The emphasis of his analysis (Frank 1998) is on partitioning selection and transmission and on unifying quantitative genetic and population genetic approaches.

Despite these suggestions that inclusive fitness and reciprocal altruism theories are related, unification of these two theories requires that HR be effectively applied to reciprocal altruism. However, until now, there has been no direct and successful demonstration of using Queller's more general version of HR in reciprocal altruism models. In fact, in an expansion of his original results, Queller (1992*a*, 1992*b*) drops any mention of its applicability to reciprocal altruism.

Here we demonstrate that Queller's equations do indeed provide a foundation for the unification of inclusive fitness and reciprocal altruism theories. Our approach differs from Nee's (1989) application of a version of HR to an IPD model of reciprocal altruism. In Nee's work, the generality of Queller's equation was not utilized; instead, Queller's version was brought back to the shared genotype level by adding an additional term that related the phenotype of others to their common genotype with focal altruists. Here we use Queller's equation in its full generality by simply taking it at face value: it is the phenotype of others that is crucial, not their genotype. This more encompassing viewpoint allows us to include heterospecific interactions in mutualistic symbiosis, where altruists and recipients are clearly genetically unrelated and nonidentical.

Our models and analysis suggest that, rather than being fundamentally different mechanisms, inclusive fitness and reciprocal altruism are alternative ways to satisfy a common single requirement for self-sacrificing traits to increase in a population. This requirement can be stated as follows: there must be sufficient positive assortment between individuals with the altruistic genotype in question and the helping phenotypes of others they interact with, such that on average those with the focal genotype benefit from the helping behaviors of others more than the costs they incur for their own helping behaviors. The required combination of genotype-phenotype assortment, benefit to cost ratio, and any nonadditive effects is given by Queller's generalization of HR. This rule applies whether the source of positive assortment is interactions among relatives (the original application), conditional behaviors among nonrelatives, or reciprocal interactions across species (mutualistic symbiosis). This single requirement governed by Queller's version of HR brings unity to the separate theories of inclusive fitness and reciprocal altruism.

We begin by briefly reviewing HR, Queller's contributions, and the original IPD experiments (Axelrod and Hamilton 1981; Axelrod 1984). We then use Queller's version of HR for genotype/phenotype differences to analyze an experiment involving an additive IPD. Using Queller's nonadditive version of HR, we also show how additive behavior that is iterated within generations gives fitness consequences similar to those of synergistic behavior where a pairing is a single interaction per generation. (Synergy is defined below as positive nonadditivity.) Finally, we extend this model so that there are conditional cooperator and defector types in each of two mutualistic species that interact. Here the particular conditional strategy, benefit level provided, cost paid for cooperative behaviors, and any nonadditive effects can be different in each of the species. For each species, we use Queller's version of HR to accurately predict whether the cooperative trait will increase in that species. Finally, we discuss the implications of these results for understanding and unifying the theories of inclusive fitness and reciprocal altruism.

The Progressive Generalization of Hamilton's Rule

Hamilton's rule (Hamilton 1963, 1964) gives the condition necessary for an altruistic trait to increase in a population in the next generation and is deceptively simple:

$$rb > c, \tag{1}$$

where b is usually interpreted as the average fitness benefit to a recipient of the altruistic behavior and c is the average cost to an altruist for this behavior. Complications arise in the meaning of the r term, which has been progressively generalized over the years. Originally thought of as a simple measure of relatedness via descent (Hamilton 1963, 1964), Hamilton (after interacting with Price [1970]) broadened the meaning of r to be a measure of the assortment of genetic types regardless of relatedness by descent (Hamilton 1970, 1972, 1975):

$$\frac{\mathrm{Cov}\,(G_A, G_O)}{\mathrm{Var}\,(G_A)}\,b > c, \tag{2}$$

Table 1: Progressive generalization of Hamilton's rule illustrated by situations for which different versions (eqq. [1]–[4]) apply

Equation	Kin interactions	Nonkin genetic similarity	Genotype-phenotype differences ($G \neq P$)	Nonadditive fitness functions ($d \neq 0$)
(1)	Yes	…	…	…
(2)	Yes	Yes	…	…
(3)	Yes	Yes	Yes	…
(4)	Yes	Yes	Yes	Yes

where G_A is the genotype (or breeding value) with respect to the altruistic trait for each potential actor (subscript A) and G_O is the average genotype value of others (subscript O) that interact with each potential actor. Queller (1985) further generalized Hamilton's r term by explicitly including the consequences of the phenotype (behaviors) of actors and others on selection for a genetic trait rather than focusing on the effect of genotypes directly. This yields

$$\frac{\text{Cov}(G_A, P_O)}{\text{Cov}(G_A, P_A)} b > c, \tag{3}$$

where P_A is the phenotype of the actor and P_O is the average phenotype of others interacting with each actor. In the appendix in the online edition of the *American Naturalist*, we provide more details about the generalization of Hamilton's rule as well as the mathematical details of our models and analysis.

These covariance ratio expressions (eqq. [2], [3]) may be convenient in comparing different versions of r, but cross multiplying results in a more easily interpreted form of Queller's equation. For example, equation (3) can be written as

$$\text{Cov}(G_A, P_O)b > \text{Cov}(G_A, P_A)c. \tag{3a}$$

This says that the genotype represented by G_A will increase in the population if the covariance between its presence in each potential actor and the helping behaviors (phenotypes) of others, scaled by the benefit of this help, is more than the covariance between its presence and the helping behaviors of actors themselves, scaled by the cost of these behaviors. Simply put, the genotype will increase if on average individuals carrying it receive more fitness benefits than they pay out. Note that it is the phenotype or behaviors of others (P_O) that is critical, not their genotype; there is no G_O term in this equation. This has consequences for the usual indirect fitness interpretation of HR. As Frank (1998, p. 68) points out, "The directionality of Queller's relatedness coefficient … is opposite to the directionality of Hamilton's inclusive fitness coefficient." The benefit (b) term in Queller's version is best interpreted as the contribution to the direct fitness of those with the altruistic genotype from the behavior of others. In the indirect fitness concept, b is seen as the contribution by the actor to the fitness of others. This is a possible interpretation of HR when these two benefits are the same, but when the amount given and the amount received by focal altruists differ, as in our model of symbiosis below, only the interpretation suggested by Queller's version works correctly.

When there is a deviation (d) from fitness additivity for mutual cooperation (as there is in many IPD models), then an additional term is needed that specifies the degree to which the focal genotype covaries with mutual cooperation,[2] scaled by the amount of deviation (Queller 1985):

$$\text{Cov}(G_A, P_O)b + \text{Cov}(G_A, P_A P_O)d > \text{Cov}(G_A, P_A)c. \tag{4}$$

This deviation value can be positive (representing synergy), negative (representing diminishing returns), or 0 (representing additivity). Note that dividing both sides of equation (4) by $\text{Cov}(G_A, P_A)$ results in the same r term as in equation (3), plus an additional covariance ratio related to the deviation from additivity. There are other versions of HR (for reviews, see Pepper 2000; West et al. 2002), but equations (1)–(4) represent significant steps in a progressive generalization of HR that are summarized in table 1.

Queller's versions (eqq. [3], [4]) apply to all situations covered by Hamilton's versions (eqq. [1], [2]), plus they handle additional situations that may not allow a recursive analysis (Grafen 1985), such as when the frequency of cooperative behavior depends on both genotype and environmental factors (e.g., the behaviors of others). To see that each equation above is more general than the previous ones, note that if fitness functions are additive ($d = 0$), then equation (4) becomes equation (3); if phenotype frequencies equal genotype frequencies ($G_O = P_O$ and $G_A = P_A$), then equation (3) becomes (2); and if the sim-

[2] The product $P_A P_O$ represents mutual cooperation, where cooperate (C) behaviors have a phenotype of 1 and defect (D) behaviors a phenotype of 0. This is explained further in the following sections.

ilarity in genotype between actors and others is solely due to interactions within kin groups, then equation (2) can become equation (1), where r represents relatedness by descent. Queller's version (eq. [4]) is the most general in that it works without these restrictions or assumptions.

The Iterated Prisoner's Dilemma Model of Reciprocal Altruism

The PD captures a fundamental problem of social life: individually rational behavior may lead to a collectively irrational and deficient outcome. In n-player versions, this dilemma is also known as a "tragedy of the commons" (Hardin 1968) or a freeloader (free rider) problem (McMillan 1979; Avilés 2002). Typical two-player PD fitness values for the actor, given its own and its opponent's behaviors, are shown in table 2. Here behaviors are either cooperate (C) or defect (D). Table 2 also shows parameters that decompose these fitness payoffs in terms of the benefit (b) provided to an opponent by a C behavior, the cost (c) paid by the cooperator, the base fitness (w_0) that is independent of cooperation, and the deviation (d) from additivity when cooperation is mutual. In the PD, each player has a dominant strategy to defect (D), but if they both cooperate, both can receive more (in this case, three instead of one) than if they both defect. It is presumed that players exhibit their behaviors simultaneously and that there is no knowledge or guarantee about what the other player will do. The dilemma is that cooperation makes a player vulnerable to exploitation; in the case of mixed behaviors, the defector gets the highest payoff (five), while the cooperator gets the lowest (zero).

Note that while a two-player fitness matrix can be represented in terms of these four parameters, other parameterizations are also possible. Also note that this typical PD matrix, which is the one used by Axelrod (1984), is nonadditive: it cannot be achieved without a nonzero d term. In this case, there are diminishing returns for mutual cooperation ($d = -1$), but synergistic ($d > 0$) matrices that still define a PD are also possible. While Axelrod purposely chose a nonadditive PD to ensure that tournament results did not depend on additivity (R. Axelrod, personal communication, 2005), the nonadditive nature of these common PD fitness values is not generally appreciated. Because of nonadditivity, this PD cannot be analyzed with a pre-Queller HR.

Although in a PD situation it is individually rational to defect in each single play of the game, Axelrod and Hamilton (1981) provided early support for reciprocal altruism theory (Trivers 1971) by showing that conditional cooperative strategies can do well when interactions (PD games) are iterated. The most successful strategy in Axelrod's tournaments (submitted by social scientist Anatol Rapoport) was also one of the simplest. Called "tit for tat" (TFT), this strategy always cooperates with an opponent in the first interaction (PD game) and, in all subsequent interactions, simply plays whatever the opponent did in the last game. This conditional behavior allowed TFT to minimize exploitation by defecting opponents, such as "always defect" (ALLD), while taking advantage of mutual cooperation when it met other "nice" strategies. Since these original experiments more than 25 years ago, much research has been done on the IPD (Dugatkin 1997; Sachs et al. 2004; Doebeli and Hauert 2005).

Table 2: Typical prisoner's dilemma fitness values for an actor, given its behavior (phenotype P_A) and its opponent's behavior (phenotype P_O)

	Opponent's behavior	
Actor's behavior	C ($P_O = 1$) contributes b	D ($P_O = 0$) contributes 0
C ($P_A = 1$) sacrifices c	3 $w_0 + b - c + d$	0 $w_0 - c$
D ($P_A = 0$) sacrifices 0	5 $w_0 + b$	1 w_0

Note: These fitness values can be represented as the result of additive benefit contributions (b) from its partner, its own sacrifice or cost (c), the base fitness value uncorrelated with C and D behaviors (w_0), and the deviation from additivity for mutual cooperation (d). For the shown fitness payoff values, $b = 4$, $c = 1$, $w_0 = 1$, and $d = -1$.

From the perspective of Queller's version of HR (eq. [3]), the combination of iterated games and conditional play can create positive assortment among the helping behaviors (phenotypes) of others and the conditionally cooperative genotype (e.g., TFT), even when there is no positive assortment among genotypes. That is, if one calculates Hamilton's r using only genotypes (eq. [2]), it will be 0 in the case of random binomial pairing. Therefore, the traditional version of HR cannot be satisfied, and it appears as if the increase in cooperation (e.g., the increase in TFT types) observed in IPD models of reciprocal altruism is not due to inclusive fitness as measured by HR. This is not in fact the case; as we will see, Queller's version of HR predicts exactly when conditional cooperation will increase in these models.

Hamilton's Rule Applied to a Classic Reciprocal Altruism Model

Confirming Queller's Version

Here we offer a simple example where Queller's version of HR (eq. [3]) is applied to reciprocal altruism using a population consisting of the two classic types mentioned above, TFT and ALLD. This kind of population has been used previously to apply a modified HR to an IPD (as mentioned above; Nee 1989), to classify types of altruism

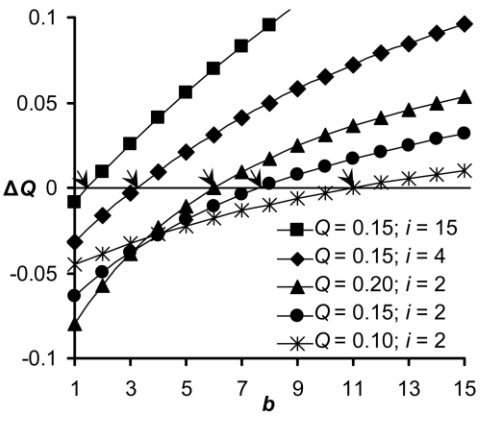

Figure 1: Change in the proportion of TFT players (ΔQ) after one generation as a function of benefit (b) level in a population of TFT and ALLD players with binomial pairing. Results shown for various starting TFT proportions (Q) and numbers of iterated games (i), where other parameters for each game are held constant at $c = 1$, $w_0 = 1$, and $d = 0$. Arrows indicate where $\Delta Q = 0$. These equilibrium points are predicted by Queller's version of HR (see table A2 in the online edition of the *American Naturalist*).

(Kerr and Godfrey-Smith 2002), and to study the evolution of altruism in finite populations (Nowak et al. 2004). Because one of the types (TFT) uses conditional behaviors, we must measure genotype and phenotype frequencies separately. To calculate the covariances needed in equation (3), we take TFT as our focal genotype (G) and assign it a value of 1 and ALLD a value of 0. For phenotype (P), the cooperate (C) behavior has a value of 1, and the defect (D) behavior a value of 0. In a population of these two types, there will be three possible parings (TFT-TFT, TFT-ALLD, ALLD-ALLD), each with predictable values for G_A, P_O, and P_A, given the number of iterated games (i) within generations (table A1 in the online edition of the *American Naturalist*).[3] Again, subscript A indicates the focal actor, and subscript O means others (in this case, the focal actor's opponent). We can now calculate the change in TFT frequency in the population (ΔQ), assuming fitness payoffs are proportional to offspring representation. Given the population averages for the frequency of C behaviors and the initial frequency of TFT types (Q), we can also calculate whether Queller's version of HR is satisfied for any given parameter settings (see appendix).

We begin with an additive PD game ($d = 0$). Figure 1 shows the change in the proportion of the TFT type (ΔQ) after one generation of random pairing and asexual reproduction as a function of the benefit value (b) for different values of the initial TFT frequency (Q) and number of iterations (i). For convenience, other parameters are held constant at $c = 1$ and $w_0 = 1$. Notice that, all else being equal, more iterations or higher starting Q make it easier for ΔQ to increase. Arrows in figure 1 indicate the predicted equilibrium points ($\Delta Q = 0$) using Queller's (eq. [3]) version of HR and the same parameters used in figure 1 (appendix). If instead of equation (3), which involves P_O, we use the more restricted equation (2), which involves G_O, $r = 0$ for this binomial population structure, and we do not correctly predict the increase in the proportion of TFT (Q). In contrast, Queller's version exactly predicts these "tipping points," that is, the value of benefit, b, beyond which TFT increases in each population.

Iterations and Synergy

Queller's version of HR (eq. [4]) suggests two ways to enhance the evolution of cooperation, given random grouping: if cooperative behaviors toward altruists are more frequent than the frequency with which they are grouped with other altruists or if there is nonadditive synergy for mutual cooperation. Here we show that these two effects can have equivalent fitness consequences. Cooperation evolves in the populations plotted in figure 1 because conditional behavior positively assorts cooperation with TFT genotypes, but an alternative analysis is possible. Assuming additive PD parameter values of $b = 4$, $c = 1$, $w_0 = 1$, and $d = 0$ and iterations of $i = 10$, the cumulative intergenerational fitness consequences of different pairings are shown in table 3 as if they were the result of a single interaction between the players. This resulting game matrix, no longer a PD, can be decomposed into our four parameters. We use primes to distinguish parameters of the resulting game from those of the original. The

Table 3: Cumulative fitness values for pairings of tit-for-tat (TFT) and always defect (ALLD) players that last for $i = 10$ iterated games

	Opponent's behavior(s)	
Actor's behavior	TFT	ALLD
TFT	40 $w_0' + b' - c' + d'$	9 $w_0' - c'$
ALLD	14 $w_0' + b'$	10 w_0'

Note: For each game, $b = 4$, $c = 1$, $w_0 = 1$, and $d = 0$. The shown fitness payoff values interpreted as the result of only a single interaction can be decomposed with $b' = 4$, $c' = 1$, $w_0' = 10$, and $d' = 27$. This game is not a prisoner's dilemma; in the game theory literature, it is called "assurance" or "stag hunt."

[3] For mathematical convenience, i represents a fixed number of games in each interaction. Similar results hold for games of average length i, and the simple players in this model are not capable of using knowledge of the number of games for backward induction.

parameters for table 3 are $b' = 4$, $c' = 1$, $w_0' = 10$, $d' = 27$ (appendix).

Now suppose that we do not know the fitness consequences of each social interaction (game) or how often interactions (iterations) occur within each generation. Instead, we see only who is paired with whom and the resulting fitness consequences to each type at the end of each generation. The fitness consequences are the same as in the original situation, but from this "black box" perspective, there are no iterations and no difference in genotype versus phenotype frequencies. There are just binomial single-interaction ($i = 1$) pairings but a strong fitness synergy when TFT meets TFT. Analyzing this synergistic ($d' = 27$) situation with equation (4) and no genotype/phenotype differences gives the exact same inequality as assuming additivity ($d = 0$) and genotype/phenotype differences due to iterations ($i = 10$) and conditional play (appendix). This perspective provides an alternative explanation for how conditional strategies such as TFT evolve. From this point of view, it is not so much that iterations and conditional play "solve" the PD itself (in which defection is favored) but that they effectively change the game into one in which mutual cooperation (in table 3 labeled TFT) has the highest fitness payoff. Note, however, that the "assurance" game that the PD has been converted into is not itself dilemma free: ALLD still receives more than TFT in all heterogeneous pairings (table 3), and therefore a maximin strategy results in a Nash equilibrium of mutual defection, which is Pareto nonoptimal.

Hamilton's Rule Applied to Cooperation across Species

Mutualism Model

Here we use Queller's version of HR to analyze a simple model of mutualistic symbiosis in which there are two interacting species (labeled 1 and 2), each with two different types: ALLD and (usually) a conditional cooperator type such as TFT. For convenience, the cooperative behaviors of interest take place only heterospecifically. For instance, the behaviors between cleaner fish and their hosts (Bshary and Grutter 2002) are strictly heterospecific: cleaner fish do not clean conspecifics, and hosts are not cleaned by other hosts. We also assume random (binomial) pairings, but conditional behavior will provide the asymmetry in benefits within species necessary for mutualists to increase (Ferriere et al. 2001). Note that these interactions, unlike our within-species cooperation examples above, can be asymmetric between species in terms of costs, benefits, and deviation from additivity (Frank 1994; Sachs et al. 2004). Figure 2 illustrates these fitness parameters between the two species. The benefit and cost for species 1 exhibiting a C behavior are b_1 (benefit to species 2) and c_1 (its own cost) and similarly for species 2. We label the benefits by their source to highlight the fact that the benefit received by a member of one species comes from a completely unrelated member of the other species. Nonadditive effects, which have their source in both species, are subscripted with the species whose fitness is directly affected. The proportion of the cooperative type in each species is given by Q_1 and Q_2, respectively. In general, whether the focal genotype (e.g., TFT) of species 1 increases in the next generation is predicted by Queller's version of HR (eq. [4]), where A (actor) is a member of species 1 and O (other) is a member of species 2:

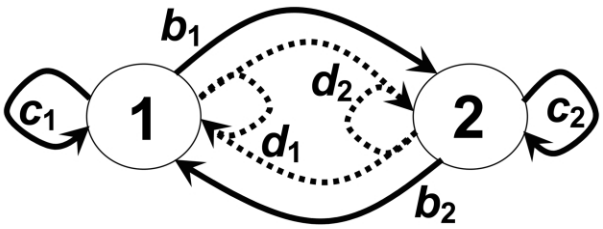

Figure 2: Fitness relationships between two interacting species (1 and 2). The origin of each arrow corresponds to a cooperate behavior (C) by the species at its origin, and the termination of each arrow indicates which species' fitness is directly affected by this behavior. For instance, a C behavior exhibited by a member of species 1 has two definite effects—decrement of c_1 in its own fitness and an increment of b_1 to its species 2 partner's fitness—as well as one potential "interaction" effect (indicted by dashed lines), a change of d_1 in its own fitness only if there is also a simultaneous C behavior by its species 2 partner. We label benefits by their source to emphasize that help is given heterospecifically.

$$\mathrm{Cov}(G_1, P_2)b_2 + \mathrm{Cov}(G_1, P_1P_2)d_1 > \mathrm{Cov}(G_1, P_1)c_1. \quad (5)$$

A symmetric equation (where all subscripts are switched) predicts whether the focal type in species 2 increases or not. We will refer to these two instances of Hamilton's rule as HR_1 and HR_2, respectively (appendix).

This model of symbiosis has some similarities to one developed by Frank (1994), but his model assumes that phenotype and genotype frequencies are the same and that fitness functions are additive. While this simple model does not capture all types of mutualisms (Bronstein 2001), as far as we know, the analysis presented here is the first example of the use of Queller's version of HR to analyze cooperation across species in which behavior can be conditional ($G \neq P$).

Dynamic Simulations

Figure 3 illustrates the coupled dynamics in our model, where cooperation can reach saturation in both species

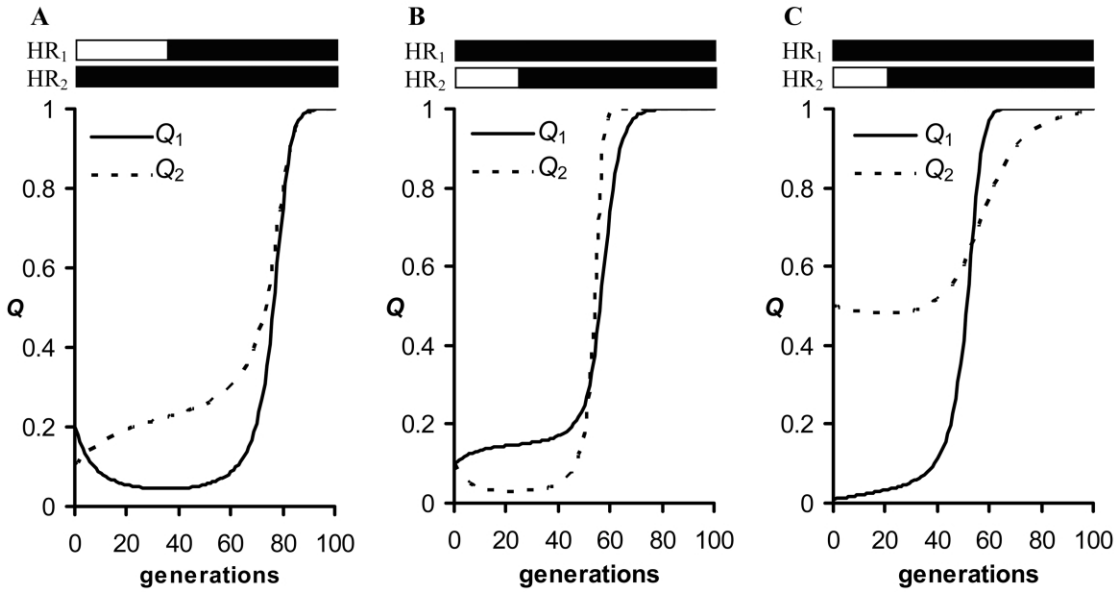

Figure 3: Dynamics between two species, each with a cooperative type and an ALLD type for various parameter settings. *A*, Cooperation evolves even though in species 1 cooperation costs more than the benefit it produces. Parameters are $i = 4$, $b_1 = 1.5$, $c_1 = 2$, $d_1 = 0$, initial $Q_1 = 0.2$, $b_2 = 5$, $c_2 = 0.1$, $d_2 = 0$, and initial $Q_2 = 0.1$; in both species, the cooperative type is TFT and $w_0 = 2$. *B*, All parameters are the same for both species, except that the cooperative type in species 1 is TF2T and in species 2 is Pavlov. Parameters are $i = 80$, $b_1 = b_2 = 4$, $c_1 = c_2 = 1$, $d_1 = d_2 = 0$, initial $Q_1 = Q_2 = 0.1$, and $w_0 = 1$. *C*, Cooperative type is unconditional (ALLC) in species 1 and TFT in species 2. Other parameters are $i = 100$, $b_1 = 2$, $c_1 = 1$, $d_1 = 0$, initial $Q_1 = 0.01$, $b_2 = 2.2$, $c_2 = 0.1$, $d_2 = 1.3$, initial $Q_2 = 0.5$, and in both species $w_0 = 1$. The bars above each graph indicate at which generations Queller's version of HR (e.g., eq. [5]) is satisfied (*solid bar*) for each species respectively (HR_1 and HR_2). Note that satisfying HR corresponds to when the cooperative type in each species increases.

even under some surprising conditions. In figure 3*A*, mutual symbiotic cooperation evolves even though cooperation costs members of species 1 more than the benefit they provide ($c_1 > b_1$). This inefficient form of altruism is not generally thought to evolve, but it can evolve here because of the high benefits of cooperation provided by the other species; that is, b_2 is sufficiently greater than c_1 (eq. [5]). In this case, fitness is additive, $d_1 = d_2 = 0$. In these dynamic numerical simulations, HR_1 and HR_2 are calculated each generation, and horizontal bars above each graph in figure 3 indicate when these respective inequalities are satisfied in each species. In all cases, Queller's version of HR accurately predicts when the cooperative type will increase. For instance, at the start of the run depicted in figure 3*A*, HR_1 (eq. [5]) is not satisfied and the TFT type decreases in species 1, but because the fraction of the TFT type in species 2 (Q_2) is simultaneously increasing, Q_1 is eventually pulled up in this coupled system.

Note that if the benefit provided by cooperative behaviors in species 1 (b_1) is used in equation (5) instead of b_2 for comparison with c_1, HR_1 does not accurately predict the fate of TFT in this species. To work, the b term must be the benefit received, not the benefit provided, by carriers of the focal genotype (e.g., TFT). The criterion for an increase in altruism is not that the benefit given by altruists sufficiently exceed their costs but rather that, on average, the benefit received by those with the altruistic genotype exceed their costs.

In the run shown in figure 3*B*, all parameters are the same for both species and additive, but the cooperative strategies differ: in species 1, it is tit for two tats (TF2T), which plays C unless the previous two plays by its opponent were both D, and in species 2, it is Pavlov (Nowak and Sigmund 1993), which initially cooperates but switches its behavior if it did not get one of the two highest payoffs in the last game. The terms Q_1 and Q_2 give the fraction of the more cooperative type in each species, respectively. Initially, when both species are dominated by ALLD types ($Q_1 = Q_2 = 0.1$), Pavlov looses ground in species 2 as it alternates C and D behaviors when paired with the ALLD type in species 1; it never gets one of the two highest payoffs and therefore keeps switching. The TF2T type in species 1 fares better because it cooperates only in the first two games when it meets an ALLD. Eventually, the fact that the TF2T type in species 1 increases provides more opportunities for the Pavlov type in species 2 to end up in a mutually cooperative interaction, and it too eventually increases.

Finally, in figure 3*C*, the cooperative type in species 1 is unconditional always cooperate (ALLC) and in species 2 is TFT. In this case, even having conditional behaviors in only one species can be enough for cooperation to evolve in both. Here species 1 faces a PD in each interaction, experiences no synergy for mutual cooperation ($d_1 = 0$), and starts with only 1% of its population being the ALLC type ($Q_1 = 0.01$). We would expect this naive cooperator to be selected out of species 1 under random pairing, but instead it steadily increases to saturation. Here species 2 starts with an even mixture of TFT and ALLD types and experiences synergy for mutual cooperation ($d_2 = 1.3$).

Of course, many other parameter settings are possible, including where cooperation goes extinct. Here we just illustrate that conditional behavior with iterations and/or nonadditivity can allow a cooperative symbiotic relationship to evolve under random grouping even if one species is inefficient ($c > b$) in its help (fig. 3*A*), less effective in avoiding exploitation by ALLD (fig. 3*B*), or even unconditionally cooperative (fig. 3*C*). Our different strategy types are not meant to represent any particular symbiosis examples in nature but to show that the evolution of mutualisms needs not depend on particular strategies and that strategies can vary between mutualistic partners. Note that in this simple model with its obligatory heterospecific interactions, the fates of the cooperative types in each species are ultimately tied together, and the system reaches an equilibrium of either all cooperation ($Q_1 = Q_2 = 1.0$) or all defection ($Q_1 = Q_2 = 0.0$). At each generation along the way, Queller's version of HR accurately predicts the direction of selection for the (conditionally) cooperative type in each species.

Discussion

A recent review article (Sachs et al. 2004), with the same title as Axelrod and Hamilton's (1981) seminal article and Axelrod's (1984) book, echoes the traditional view that inclusive fitness and reciprocal altruism are fundamentally distinct explanations for the evolution of altruism. For instance, these authors state that these theories differ because reciprocal altruism can operate "between nonrelatives and between species" (Sachs et al. 2004, p. 139) and that inclusive fitness is unique because "the cooperative individual need not benefit from its act" (Sachs et al. 2004, p. 143).

In this article, we demonstrate that, on the contrary, the distinction between inclusive fitness and reciprocal altruism is not sharp. We show that Hamilton's inclusive fitness rule (in Queller's generalized form) applies to reciprocal altruism. Analysis of Hamilton's rule also reveals that the evolution of altruism by inclusive fitness involves reciprocity (even if nonconditional): on average, carriers of the altruistic genotype must receive direct benefits. While this reciprocated benefit can be asymmetric, on average, it must sufficiently exceed focal carriers' costs, where the meaning of "sufficiently" is captured by Queller's version of HR. Inclusive fitness is also broadened beyond the notion of genes helping other copies of themselves in recipients (Williams 1966; Dawkins 1976). This "selfish gene" interpretation of HR holds only in the special case where helping behaviors are predicted by the common alleles between donor and recipients. A broader notion of inclusive fitness, as Queller (1985) argued for, is fitness augmented by help from others regardless of their genotype.

There are two important and related ideas here that are reflected in the most general form of HR (eq. [4]; table 1), which can encompass not only kin selection but also reciprocal altruism and symbiosis. For its most general application, first, the direct cost of behaving altruistically should be compared with the direct benefit gain to carriers of the altruistic genotype from others (rather than to an indirect benefit via the enhanced fitness of other carriers), and second, the direct fitness benefit depends on the behaviors (phenotypes) of others, not their genotypes. The first point is plainly evident in our symbiosis results, which show that, in the general case where benefit provided differs from benefit received, only an interpretation of HR based on direct benefit received by carriers gives a correct result. Rather than employing multiple interpretations of HR—one for relatives, one for nonrelatives having common alleles, and one for different alleles (including in different species)—it is more parsimonious for a theory of altruism to be based simply on the most general interpretation of HR. From this perspective, the equation (1) version (for relatives) of HR is just a special case of the equation (2) version, which does not depend on relatedness by descent, and the genotype of others (G_O) in the equation (2) version is just a stand-in for the phenotype of others (P_O) in the most general form of the rule, that is, Queller's equations (3) and (4). It is also more parsimonious to see HR as measuring whether the fitness gains to carriers are sufficiently greater than their costs rather than in terms of indirect fitness.

Note that explaining how, on average, benefits to carriers of the altruistic genotype end up exceeding their costs does not affect the definition of altruism at the individual level (Kerr et al. 2004). The conventional perspective in which an individual altruist incurs cost and gives benefit remains essential to defining altruism. Given that individuals have no guarantees about their partner's present and future behaviors, cooperation (C) in any given PD interaction (game) is altruistic because, compared with the alternative behavior (D), a cooperator gives benefit to others at a cost to itself. Even summing over iterations, TFT can

be seen as altruistic from a relative fitness perspective (Wilson 2004) in that TFT never does better than its paired opponent (Rapoport 1991; Sober and Wilson 1998) and its opponent does better than if it had been paired with ALLD. The fact that pairs of TFT do better than pairs of ALLD for particular parameters helps explain how altruism evolves overall, and this is captured in Queller's version of HR and in the multilevel selection framework where groups are of size 2 (Sober and Wilson 1998).

We have intentionally used simple models of reciprocal altruism in order to illustrate our point about unification, but Queller's version of HR can be applied to a much broader array of circumstances. This includes values of b, c, and d that fall outside the definition of a PD, group sizes greater than two, diploid genetics, other population structures besides binomial random grouping, degrees of cooperation rather than just all C or D, and other forms of conditional behavior beyond those based on just the past behavior of others.

The Role of Synergy

Models of the evolution of altruism typically assume various combinations of random interactions, additive fitness functions, and a one-to-one correspondence between genotype and phenotype frequencies. Dropping one or more of these assumptions can make the evolution of cooperation more likely. Traditional inclusive fitness models focus on nonrandom interactions due to kinship while leaving the other assumptions in place. Traditional reciprocal altruism models assume random encounters but use conditional behavior to break the correspondence between genotype and phenotype.

What has been less explored but is explicitly addressed by Queller's version of HR (eq. [4]) is the role of nonadditive synergy, something quite different from the potentially additive benefit of mutual cooperation (see Hauert et al. 2006). Its significance will of course depend on fitness consequences in particular interactions. In the accompanying commentary on Queller's (1985) original article, Grafen (1985, p. 311) argued that "for genes of small effect, additivity is restored and the correctness of Hamilton's rule is restored with it." While HR in its additive form may be a good approximation when fitness effects are small, cooperative traits may have strong synergistic effects. From cooperative hunters that bring home spoils greater than they could get alone (Packer and Ruttan 1988), to cooperatively swimming sperm that reach the egg faster than individual swimmers (Moore et al. 2002), to the potential synergistic benefits of mutualisms (Herre et al. 1999; Bronstein 2001; Ferriere et al. 2001), the natural world is full of potentially superadditive situations (Wright 2000; Michod and Nedelcu 2003).

At the same time, cooperative interaction may create new opportunities for defection (Michod and Nedelcu 2003), for example, the free-riding hunter or swimming sperm that expends less energy than average but still reaps the benefit of others' cooperation. The relationship between synergy and exploitation by defectors is difficult to appreciate in the paired interactions modeled here. This is because when there is one C behavior, there is no synergy, and when there are two C behaviors, there are no defectors to do the exploiting. More generally, where synergistic benefits are an increasing function of the proportion of cooperators in a group (and benefits are shared among all group members), there are necessarily more cooperators in situations with the highest synergistic payoffs, while defectors are at a relative advantage within each group because they do not pay the cost.

Synergy may also be important in addressing recent arguments that reciprocal altruism rarely occurs in nature (Hammerstein 2003). Some of these assessments are based on the low frequency of repeated interactions, but as we have shown, fewer iterations are required in the presence of nonadditive synergy. Elsewhere, we showed that multiple generations within groups similarly result in nonadditivity (Fletcher and Zwick 2004). Though not explored here, nonadditivity may also be negative, as in the typical PD (table 2) or other cases of diminishing returns (Foster 2004; Hauert et al. 2006).

A General Theory with Many Specific Mechanisms

In a review on the evolution of mutualism, Herre et al. (1999, p. 52) lament that "there is no general theory of mutualism that approaches the explanatory power that 'Hamilton's rule' appears to hold for the understanding of within-species interactions." In this article, we have shown that in fact HR itself, in Queller's generalized form, provides a general theoretical basis for understanding the evolution of cooperation across species. Fundamentally, the evolution of altruism (within or between species) depends on sufficient positive assortment between individuals with the altruistic genotype of interest and the helping behaviors (i.e., phenotypes) of others and/or sufficient synergistic effects of mutual cooperation (eq. [4]).

Sufficient association between cooperators and cooperation from others or synergistic effects can be created in a variety of ways. These include interactions in spatially structured populations among kin (Hamilton 1964) or across species (Doebeli and Knowlton 1998), iterated and conditional behavior based on the past behaviors (Trivers 1971; Axelrod and Hamilton 1981; Axelrod 1984; Dugatkin 1997) or reputations (Nowak and Sigmund 1998, 2005; Panchanathan and Boyd 2003) of others, policing (Frank 1995, 2003), punishment of nonaltruists (Boyd and Rich-

erson 1992; Fehr and Gächter 2002; Boyd et al. 2003), the constraint of social norms (Bowles et al. 2003), foraging in nonlinearly renewed heterogeneous resource distributions (Pepper and Smuts 2002), periodic environmental disturbances (Mitteldorf and Wilson 2000), the presence of fixed (Hauert et al. 2002) or conditional (Aktipis 2004) nonparticipants, the coevolution of group joining and synergistic cooperative behaviors (Avilés et al. 2004), multigenerational groups (Fletcher and Zwick 2004), and even recognition and coevolution of arbitrary tags (Riolo et al. 2001; Axelrod et al. 2004). And of course more than one of these less proximate mechanisms can be operating simultaneously.

Summary

In summary, we have argued for a common mechanism by which altruism evolves that is fundamental to both inclusive fitness and reciprocal altruism theories. To support the unification of these two theories, we have demonstrated how a more general form of Hamilton's inclusive fitness rule developed by Queller (1985) can be used to analyze a classic model of reciprocal altruism. In addition, we have shown how Queller's version of HR accurately predicts the evolution of cooperation between two different species in a simple model of symbiosis. This highlights the fact that kinship, genetic similarity, or common species identity between donors and recipients is not fundamental to the workings of Hamilton's rule. What this rule requires is that those carrying the altruistic genotype receive direct benefits from the phenotype (behaviors) of others (adjusted by any nonadditive effects) that on average exceed the direct costs of their own behaviors. Kinship interactions or conditional iterated behaviors are merely two of many possible ways of satisfying this fundamental condition for altruism to evolve.

Acknowledgments

For helpful comments on the manuscript, we thank C. Hauert, D. Queller, P. Salazar, D. Wilson, and an anonymous reviewer. For helpful discussions, we are grateful to I. Angerson, L. Avilés, A. Blachford, R. Blok, M. Doebeli, B. Kerr, L. Nunney, J. Purcell, C. Spenser, and J. Tyeman. For support, J.A.F. thanks the National Science Foundation International Fellowship Program.

Literature Cited

Aktipis, C. A. 2004. Know when to walk away: contingent movement and the evolution of cooperation. Journal of Theoretical Biology 231:249–260.

Avilés, L. 2002. Solving the freeloaders paradox: genetic associations and frequency-dependent selection in the evolution of cooperation among nonrelatives. Proceedings of the National Academy of Sciences of the USA 99:14268–14273.

Avilés, L., J. A. Fletcher, and A. Cutter. 2004. The kin composition of groups: trading group size for degree of altruism. American Naturalist 164:132–144.

Axelrod, R. 1984. The evolution of cooperation. Basic Books, New York.

Axelrod, R., and W. D. Hamilton. 1981. The evolution of cooperation. Science 211:1390–1396.

Axelrod, R., R. A. Hammond, and A. Grafen. 2004. Altruism via kin-selection strategies that rely on arbitrary tags with which they coevolve. Evolution 58:1833–1838.

Bowles, S., J.-K. Choi, and A. Hopfensitz. 2003. The co-evolution of individual behaviors and social institutions. Journal of Theoretical Biology 223:135–147.

Boyd, R., and P. J. Richerson. 1992. Punishment allows the evolution of cooperation (or anything else) in sizable groups. Ethology and Sociobiology 13:171–195.

Boyd, R., H. Gintis, S. Bowles, and P. J. Richerson. 2003. The evolution of altruistic punishment. Proceedings of the National Academy of Sciences of the USA 100:3531–3535.

Breden, F. 1990. Partitioning of covariance as a method for studying kin selection. Trends in Ecology & Evolution 5:224–228.

Bronstein, J. L. 2001. The exploitation of mutualisms. Ecology Letters 4:277–287.

Bshary, R., and A. S. Grutter. 2002. Asymmetric cheating opportunities and partner control in a cleaner fish mutualism. Animal Behaviour 63:547–555.

Dawkins, R. 1976. The selfish gene. Oxford University Press, New York.

Doebeli, M., and C. Hauert. 2005. Models of cooperation based on the prisoner's dilemma and the snowdrift game. Ecology Letters 8:748–766.

Doebeli, M., and N. Knowlton. 1998. The evolution of interspecific mutualisms. Proceedings of the National Academy of Sciences of the USA 95:8676–8680.

Dugatkin, L. A. 1997. Cooperation among animals: an evolutionary perspective. Oxford University Press, New York.

Fehr, E., and S. Gächter. 2002. Altruistic punishment in humans. Nature 415:137–140.

Ferriere, R., J. L. Bronstein, S. Rinaldi, R. Law, and M. Gauduchon. 2001. Cheating and the evolutionary stability of mutualisms. Proceedings of the Royal Society of London B 269:773–780.

Fletcher, J. A., and M. Zwick. 2004. Strong altruism can evolve in randomly formed groups. Journal of Theoretical Biology 228:303–313.

Foster, K. R. 2004. Diminishing returns in social evolution: the not-so-tragic commons. Journal of Evolutionary Biology 17:1058–1072.

Frank, S. A. 1994. Genetics of mutualism: the evolution of altruism between species. Journal of Theoretical Biology 170:393–400.

———. 1995. Mutual policing and repression of competition in the evolution of cooperative groups. Nature 377:520–522.

———. 1998. Foundations of social evolution. Princeton University Press, Princeton, NJ.

———. 2003. Perspective: repression of competition and the evolution of cooperation. Evolution 57:693–705.

Grafen, A. 1985. Hamilton's rule OK. Nature 318:310–311.

Hamilton, W. D. 1963. The evolution of altruistic behavior. American Naturalist 97:354–356.

———. 1964. The genetical evolution of social behavior. I, II. Journal of Theoretical Biology 7:1–52.
———. 1970. Selfish and spiteful behavior in an evolutionary model. Nature 228:1218–1220.
———. 1972. Altruism and related phenomena, mainly in social insects. Annual Review of Ecology and Systematics 3:193–232.
———. 1975. Innate social aptitudes of man: an approach from evolutionary genetics. Pages 133–155 *in* R. Fox, ed. Biosocial anthropology. Wiley, New York.
Hammerstein, P. 2003. Why is reciprocity so rare in social animals? Pages 83–93 *in* P. Hammerstein, ed. Genetic and cultural evolution of cooperation. MIT Press, Cambridge, MA.
Hardin, G. 1968. The tragedy of the commons. Science 162:1243–1248.
Hauert, C., S. De Monte, J. Hofbauer, and K. Sigmund. 2002. Volunteering as Red Queen mechanism for cooperators in public goods games. Science 296:1129–1132.
Hauert, C., F. Michor, M. A. Nowak, and M. Doebeli. 2006. Synergy and discounting of cooperation in social dilemmas. Journal of Theoretical Biology 239:195–202.
Herre, E. A., N. Knowlton, U. G. Mueller, and S. A. Rehner. 1999. The evolution of mutualisms: exploring the paths between conflict and cooperation. Trends in Ecology & Evolution 14:49–53.
Kerr, B., and P. Godfrey-Smith. 2002. Individualist and multi-level perspectives on selection in structured populations. Biology and Philosophy 17:477–517.
Kerr, B., P. Godfrey-Smith, and M. W. Feldman. 2004. What is altruism? Trends in Ecology & Evolution 19:135–140.
Maynard Smith, J. 1964. Group selection and kin selection. Nature 201:1145–1147.
McMillan, J. 1979. The free-rider problem: a survey. Economic Record 55:95–107.
Michod, R. E., and A. M. Nedelcu. 2003. On the reorganization of fitness during evolutionary transitions in individuality. Integrative and Comparative Biology 43:64–73.
Mitteldorf, J., and D. S. Wilson. 2000. Population viscosity and the evolution of altruism. Journal of Theoretical Biology 204:481–496.
Moore, H., K. Dvoráková, N. Jenkins, and W. Breed. 2002. Exceptional sperm cooperation in the wood mouse. Nature 418:174–177.
Nee, S. 1989. Does Hamilton's rule describe the evolution of reciprocal altruism? Journal of Theoretical Biology 141:81–91.
Nowak, M. A., and K. Sigmund. 1993. A strategy of win-stay lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. Nature 364:56–58.
———. 1998. Evolution of indirect reciprocity by image scoring. Nature 393:573–577.
———. 2005. Evolution of indirect reciprocity. Nature 437:1291–1298.
Nowak, M. A., A. Sasaki, C. Taylor, and D. Fudenberg. 2004. Emergence of cooperation and evolutionary stability in finite populations. Nature 428:646–650.
Packer, C., and L. Ruttan. 1988. The evolution of cooperative hunting. American Naturalist 132:159–198.
Panchanathan, K., and R. Boyd. 2003. A tale of two defectors: the importance of standing for evolution of indirect reciprocity. Journal of Theoretical Biology 224:115–126.
Pepper, J. W. 2000. Relatedness in trait group models of social evolution. Journal of Theoretical Biology 206:355–368.
Pepper, J. W., and B. B. Smuts. 2002. A mechanism for the evolution of altruism among nonkin: positive assortment through environmental feedback. American Naturalist 160:205–213.
Price, G. R. 1970. Selection and covariance. Nature 277:520–521.
Queller, D. C. 1985. Kinship, reciprocity and synergism in the evolution of social behavior. Nature 318:366–367.
———. 1992*a*. A general model for kin selection. Evolution 46:376–380.
———. 1992*b*. Quantitative genetics, inclusive fitness, and group selection. American Naturalist 139:540–558.
Rapoport, A. 1991. Ideological commitments and evolutionary theory. Journal of Social Issues 47:83–99.
Riolo, R. L., M. D. Cohen, and R. Axelrod. 2001. Evolution of cooperation without reciprocity. Nature 414:441–443.
Sachs, J. L., U. G. Mueller, T. P. Wilcox, and J. J. Bull. 2004. The evolution of cooperation. Quarterly Review of Biology 79:135–160.
Sober, E., and D. S. Wilson. 1998. Unto others: the evolution and psychology of unselfish behavior. Harvard University Press, Cambridge, MA.
Trivers, R. L. 1971. The evolution of reciprocal altruism. Quarterly Review of Biology 46:35–57.
Wade, M. J. 1978. A critical review of the models of group selection. Quarterly Review of Biology 53:101–114.
———. 1980. Kin selection: its components. Science 210:665–667.
West, S. A., I. Pen, and A. S. Griffin. 2002. Cooperation and competition between relatives. Science 296:72–75.
Williams, G. C. 1966. Adaptation and natural selection. Princeton University Press, Princeton, NJ.
Wilson, D. S. 1975. A theory of group selection. Proceedings of the National Academy of Sciences of the USA 72:143–146.
———. 2004. What is wrong with absolute individual fitness? Trends in Ecology & Evolution 19:245–248.
Wright, R. 2000. Non-zero: the logic of human destiny. Pantheon, New York.

Associate Editor: Peter D. Taylor
Editor: Jonathan B. Losos