

# The evolution of altruism: Game theory in multilevel selection and inclusive fitness

Jeffrey A. Fletcher<sup>a,b,\*</sup>, Martin Zwick<sup>a</sup>

<sup>a</sup>*Systems Science Ph.D. Program, Portland State University, Portland, OR 97207, USA*

<sup>b</sup>*Department of Zoology, The University of British Columbia, 2370-6270 University Blvd., Vancouver, BC, Canada V6T 1Z4*

Received 23 January 2006; received in revised form 27 September 2006; accepted 27 September 2006

Available online 4 October 2006

## Abstract

Although the prisoner's dilemma (PD) has been used extensively to study reciprocal altruism, here we show that the  $n$ -player prisoner's dilemma (NPD) is also central to two other prominent theories of the evolution of altruism: inclusive fitness and multilevel selection. An NPD model captures the essential factors for the evolution of altruism directly in its parameters and integrates important aspects of these two theories such as Hamilton's rule, Simpson's paradox, and the Price covariance equation. The model also suggests a simple interpretation of the Price selection decomposition and an alternative decomposition that is symmetrical and complementary to it. In some situations this alternative shows the temporal changes in within- and between-group selection more clearly than the Price equation. In addition, we provide a new perspective on strong vs. weak altruism by identifying their different underlying game structures (based on absolute fitness) and showing how their evolutionary dynamics are nevertheless similar under selection (based on relative fitness). In contrast to conventional wisdom, the model shows that *both* strong and weak altruism can evolve in periodically formed random groups of non-conditional strategies if groups are multigenerational. An integrative approach based on the NPD helps unify different perspectives on the evolution of altruism.

© 2006 Elsevier Ltd. All rights reserved.

**Keywords:** Hamilton's rule;  $n$ -player prisoner's dilemma; Price covariance equation; Simpson's paradox; Strong versus weak altruism

## 1. Introduction

The evolutionary mechanisms by which altruistic behaviors evolve have been vigorously debated over the last several decades. The most prominent theories are reciprocal altruism (Axelrod and Hamilton, 1981; Trivers, 1971), inclusive fitness (Hamilton, 1964, 1970, 1975), and multi-level selection (Wade, 1978; Wilson, 1977, 1997). The iterated prisoner's dilemma (PD) naturally lends itself to the study of reciprocal altruism (Axelrod, 1984; Dugatkin, 1997), yet real-world biological and social systems often do not involve pair-wise interactions or knowledge of past actions, and consequences of cooperation and defection may be distributed among many individuals. The  $n$ -player prisoner's dilemma (NPD) is a model that captures the

diffuse harm to the common good that self-interested behaviors may cause. It encompasses *both* (Hardin, 1971) problems of exploitation of a common resource ("tragedy of the commons" (Hardin, 1968)), and problems of inequitable contributions towards a common good ("free-rider problem" (Avilés, 2002; McMillan, 1979)). The minimal conditions for the evolution of altruism are captured in our NPD model and are: non-zero-sum fitness functions for altruistic behaviors and sufficient population assortment (variance among groups) with respect to these behaviors. Heritability is assumed.

The simplicity of our model allows us to connect explicitly the NPD to the other two related theories of altruism evolution (Frank, 1998; Queller, 1985, 1992; Sober and Wilson, 1998; Wade, 1980): inclusive fitness and multilevel selection, a connection which has not previously been made formally, despite hints in the literature (e.g. Bowles et al., 2003; Frank, 1995, 1998; Hauert et al., 2002; Leigh, 1999; Michod, 1999). George Price was one of the

\*Corresponding author. Tel.: +1 604 822 3224; fax: +1 604 822 2416.

E-mail addresses: [fletcher@zoology.ubc.ca](mailto:fletcher@zoology.ubc.ca) (J.A. Fletcher), [zwickm@pdx.edu](mailto:zwickm@pdx.edu) (M. Zwick).

first to recognize the similarities between social dilemmas and levels of selection: “the cases discussed where individual selection decreases group fitness are closely and deeply analogous to economic effects recently discussed by Hardin in a paper entitled *The tragedy of the commons...*” (1969, quoted in Frank, 1995). It was later shown that the tragedy of the commons is equivalent to an NPD (Hardin, 1971). Surprisingly, this connection is still not widely appreciated. For instance, a recent issue of *Science* (Kennedy, 2003) devoted entirely to the tragedy of the commons fails to even mention the PD.

The NPD model parameters relate simply to Hamilton’s rule (from inclusive fitness theory) and Simpson’s paradox and the Price covariance equation (which are central to understanding multilevel selection theory). The fundamental role of assortment is captured in Hamilton’s rule; in multilevel selection theory this rule describes the variance in group composition, where groups with a high proportion of altruists are more productive than those with less. This variance in productivity drives between-group selection favoring altruist-dominated groups, while within-group selection favors non-altruists within every group.

The model also suggests a simple interpretation of the Price selection decomposition and a symmetrical alternative version presented here. This alternative decomposition, when contrasted with the Price equation, highlights the assumptions, usefulness, and limitations of both. We also identify the different game structures that distinguish weak and strong altruism and show how these two types of altruism behave similarly under selection. Specifically, in contrast to conventional wisdom, *both* strong and weak altruism can be selected for and maintained using periodically formed random groups of non-conditional strategies if groups are multigenerational. Our game-theoretic model and analysis thus offer a framework for unifying different approaches to the evolution of altruism.

## 2. The NPD model

In the simplest form of the model there are only two groups with no dispersal. In order to illustrate fully the potential course of opposing between- and within-group selective forces acting over multiple generations, we begin by keeping these two groups completely isolated. This results in the population eventually reaching an equilibrium of mutual defection. In Section 5, we relax this constraint and model populations composed of many groups where random mixing occurs after varying numbers of generations within groups. In both versions of the model, for each group  $i$  there are  $a_i$  cooperators (altruists) and  $s_i$  defectors (selfish) with group size  $n_i = a_i + s_i$  and the fraction of cooperators  $q_i = a_i/n_i$ . Note that there are no strategies other than always-cooperate and always-defect. We follow the frequency of cooperators in each group and across the whole population. Fig. 1 illustrates a simple NPD with parallel linear fitness functions,  $w_a$  and  $w_s$ , that give the fitness *per individual* cooperator (altruist) or

defector (selfish) in the vertical axis and where  $q_i$  is the horizontal axis. These parallel lines are the simplest fitness curves that satisfy the NPD.

There are two parameters in this model: the slope,  $b$ , of the fitness functions and,  $c$ , their difference in intercept. (For simplicity we add a base fitness  $w_0 = c$  to both  $w_a$  and  $w_s$  so that fitness payoff values are never negative and the cooperator’s intercept is 0.) The cost of being a cooperator vs. a defector is the intercept difference  $c$ . The benefit provided by each cooperator to the group is  $b$ . To see this, note that the added benefit to *each* group member (including the focal player) in having one additional cooperator in the group (vs. a defector) is  $b/n_i$  (the change in  $q_i$  is  $1/n_i$ ) and therefore the *total* benefit produced by a cooperator for all group members is  $n_i(b/n_i) = b$ . If  $b$  is sufficiently bigger than  $c$  and/or group size ( $n_i$ ) is small enough, such that  $b/n_i > c$ , then this is no longer a PD. In this case we have weak altruism (Wilson, 1975, 1990) as opposed to the strong altruism that corresponds to the PD. Note that even if  $b$  and  $c$  are held constant, different groups within the same population can exhibit either strong or weak altruism depending on their size.

For both types of altruism, the defector’s fitness line dominates the cooperator’s at all  $q_i$  and therefore cooperation always involves an altruistic *relative* sacrifice of fitness compared to defection. The deficient outcome captured in this model is the fact that the fitness to defectors when all players in a group defect ( $q_i = 0.0$ ) is *lower* than the fitness to cooperators when all group members cooperate ( $q_i = 1.0$ ), i.e.,  $b > c$ . This is a necessary condition for an NPD. It is also the condition for beneficial non-zero-sumness, i.e., the benefit created by a cooperator exceeds the cost to the cooperator and the average fitness line ( $w_{av}$ ) has a positive slope. Since defection dominates cooperation, this deficient outcome is an attractor of the dynamics within each group. Thus for both strong and weak altruism within-group selection acting alone *does not maximize* individual fitness. We explore the strong vs. weak altruism issue more fully in Sections 4 and 5.

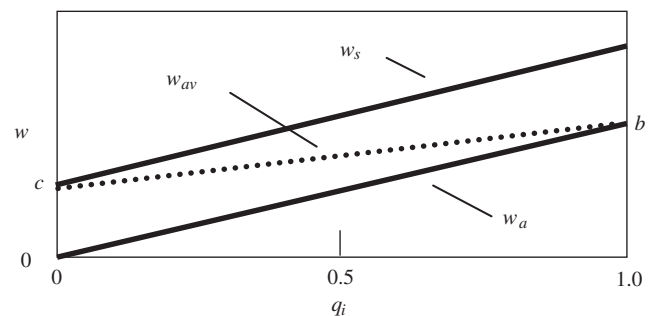


Fig. 1. Simple n-player prisoner’s dilemma. Fitness functions for individual cooperators ( $w_a$ ) and defectors ( $w_s$ ) given as a function of the fraction of cooperators in a group ( $q_i$ ). The two solid lines have slope  $b$ . The dashed line indicates the average fitness ( $w_{av}$ ), which has a positive slope. The intercept difference of the two functions is given by  $c$ , where for convenience a base fitness equal to  $c$  has been added to both fitness functions.

At each generation the number of cooperators and defectors within each group is increased proportional to fitness payoffs:

$$a'_i = a_i[1 + w_a(q_i)], \quad (1)$$

$$s'_i = s_i[1 + w_s(q_i)], \quad (2)$$

where primed terms represent values after reproduction. These fitness functions can be interpreted as overlapping generations or as discrete generations where the fitness independent of the altruistic trait is one offspring per individual. Here fitness is fecundity and offspring counts (including fractional components) are determined by the fitness functions from Fig. 1:

$$w_a(q_i) = bq_i - c + w_0, \quad (3)$$

$$w_s(q_i) = bq_i + w_0. \quad (4)$$

To aid in comparisons among runs, in each generation the *total* population size is proportionally scaled to its original size, preserving each group's  $q_i$  value. Competition between groups is implicit as more productive groups comprise a larger proportion of the total population in subsequent generations. For convenience we define total population variables  $A = \sum a_i$ ,  $S = \sum s_i$ ,  $N = A + S$  and  $Q = A/N$ .

In the next section, we show the analytic connection of our NPD model to Hamilton's inclusive fitness rule. We then (Section 2.2) use the minimal two group version of this model to illustrate Simpson's paradox dynamics where each of the two groups begins at the same size (500 individuals) but is different in composition. We also use this version of our NPD model to illustrate its connection to the Price covariance equation and our alternative selection decomposition (Section 3). After using the model to illustrate different classifications of altruism (Section 4), we switch to a version where there are 100 groups of initial size 10 and as above group size is allowed to vary with group productivity, but at regular periods groups are randomly reformed again to groups of size 10 (Section 5). Similar results hold for a broad range of parameter combinations (Fletcher and Zwick, 2004; Wilson, 1987). We do not fully explore this range here; our purpose is to illuminate and unify some fundamental concepts with illustrative examples.

### 2.1. Hamilton's rule

In this simple model, the condition for an increase in the overall frequency of cooperators from one generation to the next,  $Q' > Q$ , can be derived starting with our fitness functions Eqs. (3) and (4) (see Appendix A) and results in a form of Hamilton's rule (Hamilton, 1964, 1970, 1975):  $rb > c$ . The  $r$  value in this derivation can be expressed as the between-group over total variance in the cooperate trait,  $r = \text{var}_B(q_i) / \text{var}_T(Q)$ , where  $\text{var}_B(q_i)$  is the weighted between-group variance and  $\text{var}_T(Q)$  the total variance

among individuals in this trait. (We refer to this expression as the "variance ratio".) This is consistent with previous work showing that for altruists that benefit the whole group (Pepper, 2000) as in the model above,  $r$  can be expressed in terms of the variance ratio (Breden, 1990; Frank, 1998; Queller, 1992). According to Hamilton's rule, for altruism to evolve, the benefit  $b$  must not only be greater than the cost  $c$  (the minimum NPD condition), but the benefit must be greater than the cost even when the benefit is discounted by the variance ratio. The more structured the population with regard to cooperative interactions (i.e. the closer the variance ratio is to 1), the less non-zero-sumness is required (the smaller  $b-c$  can be).

The meaning of  $r$  has changed over the years from a simple measure of relationship via descent (Hamilton, 1964) to various statistical measures of similarity (Frank, 1998; Hamilton, 1970, 1975; Queller, 1985, 1992; Wade, 1980). For instance,  $r$  is often calculated using the covariance between the genotype of each potential actor and the average genotype of their recipients (Hamilton, 1975)—regardless of whether this assortment is due to kinship or, as we demonstrate later, even to random grouping. Hamilton's rule, thus, predicts the conditions under which altruism can increase not only via inclusive fitness, but also via multilevel selection. Note that when the benefits provided by an altruist are divided among only *others* in the altruist's group, then the average genotype of others does not include the actor, but when altruism is whole-group the actor *is* included in this average when calculating  $r$ . Thus  $r$  has a different meaning (and value) for other-only vs. whole-group altruism (Hamilton, 1972; Pepper, 2000). Note also that which individuals interact is not necessarily determined by distinct physical groups. Wilson (1975) coined the term "trait groups" to capture the idea that the interactions, with regard to the trait of interest, determine the assortment for that particular trait. It is this assortment between genotypes that interact that is traditionally measured in the  $r$  term of Hamilton's rule.

However, while the traditional interpretation of Hamilton's rule relies on genetic similarity between altruists and recipients, it is actually the phenotypes of others (not their genotypes) that determines what evolves. In simple models, including ours, where genotype completely determines phenotype, either formulation can be used. But in situations where behavior is conditional (e.g. traditional iterated PD experiments) the more general version of Hamilton's rule (Queller, 1985) using the phenotypes of others must be used (Fletcher and Zwick, 2006).

### 2.2. Simpson's paradox

Even though  $Q$  can increase, i.e.  $Q' > Q$ , when Hamilton's rule is satisfied, the frequency of cooperators decreases in every group, i.e.  $q'_i < q_i$  for all groups. This is an example of Simpson's paradox (Simpson, 1951), which is key to understanding the role of population structure in the evolution of altruism (Sober and Wilson, 1998) and is a

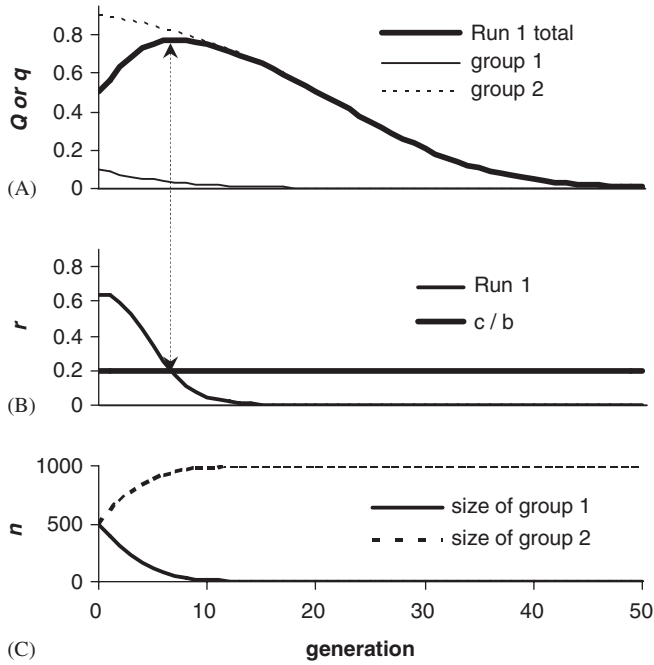


Fig. 2. Dynamics in  $Q$  or  $q$ ,  $r$ , and  $n$  for a typical NPD run (Run 1) with two evenly sized groups that vary in their initial frequency of cooperators. Parameter values are shown in Table 1: (A) frequency of cooperators vs. generation for the total population ( $Q$ ) and for each group ( $q_1$  and  $q_2$ ); (B) the between-group over total variance ratio in cooperation frequency ( $r$ ) vs. generation. The  $c/b$  value is also shown. A vertical dashed line with arrows indicates the critical point in Run 1—when  $r$  drops below  $c/b$  in (B),  $Q$  begins to decline in (A); and (C) shows how the size of each group changes over the run.

simple emergent of our NPD model. Fig. 2 shows a run (Run 1) in this model where Simpson’s paradox is evident. All runs used in Figs. 2–4 have a total population of 1000 divided into two groups with varying NPD parameters and initial population structures (Table 1). To help illustrate Simpson’s paradox all these runs are started at non-equilibrium strong altruism (NPD) conditions and allowed to progress to their natural equilibrium of 100% defection, given no migration or group reformation. Later we show how periodically reforming groups can yield an equilibrium of 100% cooperation. In Run 1 the overall cooperator frequency  $Q$  is initially 0.5 and the group sizes are equal, but the group cooperator frequencies differ:  $q_1 = 0.1$  and  $q_2 = 0.9$ . This population structure gives a variance ratio of 0.64, well above the  $c/b$  ratio of 0.2 for this run (Fig. 2B), and therefore  $Q$  increases in accord with Hamilton’s rule, even though  $q_1$  and  $q_2$  both decrease monotonically (Fig. 2A). Here, the Simpson’s paradox effect is due to group 2 (cooperator dominated) rapidly expanding, while group 1 (defector dominated) is shrinking, which is shown in Fig. 2C. At the peak of total cooperation in Run 1, group 2 comprises over 95% of the total population.

As mentioned, the Simpson’s paradox effect is transient unless, as we demonstrate later, mechanisms exist for reestablishing variation among groups. The changes in

Table 1  
Parameters for Runs 1–5 used in Figs. 2–4

Run	$a_1$	$s_1$	$a_2$	$s_2$	$c$	$b$	$c/b$
1	50	450	450	50	0.20	1.00	0.2
2	50	450	450	50	1.00	5.00	0.2
3	50	450	450	50	0.03	0.15	0.2
4	9	1	1	989	0.10	0.50	0.2
5	60	50	840	50	0.10	0.50	0.2

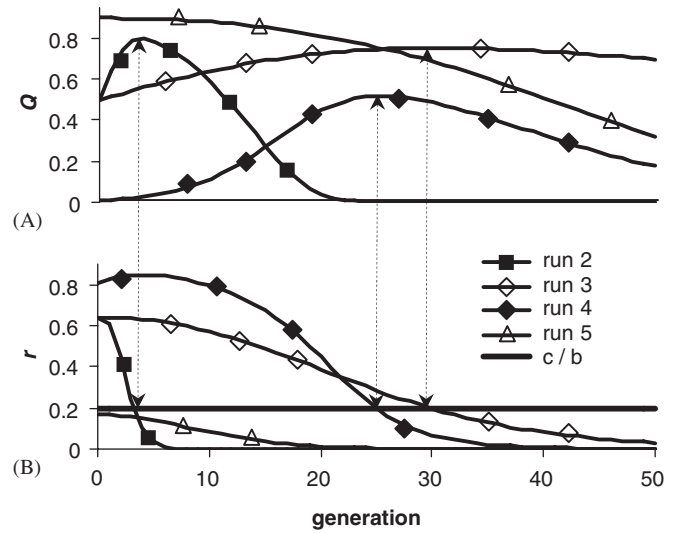


Fig. 3. Results for four Runs with various parameters (see Table 1). Run 2 ( $b = 5.00$ ,  $c = 1.00$ ) and Run 3 ( $b = 0.15$ ,  $c = 0.03$ ) demonstrate, respectively, the effect of increasing and decreasing the magnitude of  $b$  and  $c$  compared to Run 1 ( $b = 1.00$ ,  $c = 0.20$ ). Run 4 demonstrates a marked increase in cooperation despite a low initial cooperator frequency ( $Q = 0.01$ ); Run 5 demonstrates a steady decline in cooperator frequency despite a high initial cooperator frequency ( $Q = 0.9$ ); (A) frequency of cooperators ( $Q$ ) vs. generation and (B) the between-group over total variance ratio in cooperation frequency ( $r$ ) vs. generation. For all runs the  $c/b$  value is the same (0.20) and is shown. Vertical dashed lines with arrows indicate corresponding points in runs—when  $r$  drops below  $c/b$  in (B),  $Q$  begins to decline in (A).

group size and composition affect the variance ratio ( $r$ )—in the case of Run 1, the ratio decreases steadily (Fig. 2B). The generation when the variance ratio drops below  $c/b$  is precisely the point when the overall cooperator frequency begins to decline. A vertical dashed line with arrows indicates this corresponding point for Run 1 in Fig. 2.

Figs. 3A and B are similar to Figs. 2A and B, but four additional runs with a variety of NPD parameters are compared. The parameters from all five runs are given in Table 1. Fig. 3A shows the overall cooperation frequency  $Q$ , while Fig. 3B gives the variance ratio for each of these runs. To aid in comparing runs the  $c/b$  ratio is chosen arbitrarily to be the same for all runs and is shown by a thick unadorned horizontal line in both Figs. 2B and 3B. Many other parameter values give similar results.

Runs 2 and 3 show the effects of varying the magnitude of  $b$  and  $c$ , while keeping  $c/b$  and initial population structure (variance ratio) the same as in Run 1. In Run 2

with higher magnitudes the increase and subsequent decrease in  $Q$  happens more quickly; in Run 3 with lower magnitudes, the pattern is stretched out over many more generations. Run 4 demonstrates that even with low initial  $Q$  values ( $Q = 0.01$ ), a sufficiently high variance ratio can lead to a dramatic increase in cooperators. Run 4 also shows that the variance ratio need not always decrease (Fig. 3B). Changes in group size and composition can cause the variance ratio to increase transiently without external causes or mixing. Finally, Run 5 makes the point that even with an initial high frequency of cooperators,  $Q = 0.9$ , cooperators will not increase without a sufficient variance ratio. Here the variance ratio is less than  $c/b$  and therefore  $Q$  decreases monotonically. In summary, the NPD model given appropriate group structure yields Simpson’s paradox. However, the effect is transient without mechanisms for maintaining the necessary population structure, which we demonstrate further on.

### 3. An alternative to the Price selection decomposition

In the runs discussed so far, the transient increase and subsequent decrease in cooperator frequency highlights competing forces—the overall frequency of cooperators,  $Q$ , increases while the *between-group* selective force dominates and decreases when the *within-group* force becomes stronger. Here we present an alternative to the Price (1970) selection decomposition, which is symmetrical to it, but gives different results. We realized this alternative was possible while exploring the underlying idealizations implicit in the Price decomposition, idealizations which are usually not stated or recognized.

#### 3.1. The Price equation

Price (1970) introduced a covariance equation which allows us to partition the change in overall cooperator frequency,  $\Delta Q = Q' - Q$ , into within- and between-group components:

$$\Delta Q = \frac{\text{cov}(w_i, q_i)}{E(w_i)} + \frac{E(w_i \Delta q_i)}{E(w_i)} \quad (5)$$

or equivalently as  $\Delta Q = \Delta Q_B + \Delta Q_W$ , where  $w_i$  is a measure of group fitness, here the growth rate of each group ( $n'_i/n_i$ ), and  $\Delta Q_B$  and  $\Delta Q_W$  are the Price between- and within-group components of change in overall cooperator frequency, respectively. To highlight the underlying assumptions of the Price equation, these components of change can be rewritten as:  $\Delta Q_B = Q^* - Q$  and  $\Delta Q_W = Q' - Q^*$ , where  $Q^* = \sum[q_i n'_i]/N'$  (Appendix B). The  $Q^*$  term in  $\Delta Q_B$  has a simple interpretation: it plays the role of an idealized  $Q'$  in which the *before-selection*  $q_i$  values are applied to the *after-selection* group sizes,  $n'_i$ . The corresponding within-group expression corrects for the ignored changes in cooperator frequency within groups.

#### 3.2. An alternative selection decomposition

This simple interpretation of the idealization implicit in the Price between-group component suggests a symmetric alternative decomposition where the *within-group* component contains the  $Q'$  idealization and the between-group component is the correction term. This idealization assumes that the frequency of cooperators within each group changes, but that the relative size (fitness) of groups does not—we use the *after-selection*  $q'_i$  values and the *before-selection* group sizes,  $n_i$ . This alternative  $Q'$  idealization we denote as:  $Q^\# = \sum[q'_i n_i]/N$  and the alternative components of selection can be labeled:  $\text{alt} \Delta Q_W = Q^\# - Q$  and  $\text{alt} \Delta Q_B = Q' - Q^\#$ .

Fig. 4 shows the change from the initial  $Q$  value for Run 1 of Fig. 2. Also shown are the selection components of Run 1 given by the Price decomposition and the alternative. Notice that the two decompositions give quite different results. In the Price decomposition, the equilibrium state ( $Q = 0.0$ ) consists of a balance between a strong between-group force, 0.4, even though group 1 has disappeared (see Fig. 2C), and a strong within-group force,  $-0.9$ , even though cooperators have disappeared. In contrast, the alternative decomposition more intuitively says that the between-group selection force rises as group 2 initially increases over group 1, but that this force goes to zero as the first group disappears. The alternative decomposition thus more accurately captures the fact that the between-group selective force peaks at an intermediate number of generations within groups (see Fletcher and Zwick, 2004; Wilson, 1987 for other examples). In contrast, the Price decomposition incorrectly indicates that the between-group component of selection never diminishes. Which of these two decompositions is more appropriate will depend on the situation being studied. In the runs illustrated here where the within-group selection force eventually dominates, the alternative decomposition provides more insight.

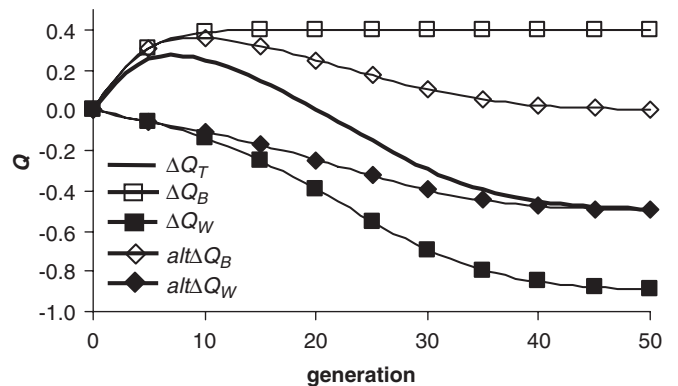


Fig. 4. Change in cooperator frequency ( $\Delta Q_T$ ) vs. generation for Run 1 of Fig. 2 along with the between- and within-group components of selection given by the Price decomposition ( $\Delta Q_B$  and  $\Delta Q_W$ , respectively) and the alternative decomposition ( $\text{alt} \Delta Q_B$  and  $\text{alt} \Delta Q_W$ ). The population consists of two isolated groups of 500 each where initial conditions are  $q_1 = 0.1$  and  $q_2 = 0.9$ .

### 3.3. The two decompositions in relation to the NPD

These two distinct decompositions relate directly to our NPD model. In the Price equation idealization  $Q^* = \sum [a_i(1 + w_{av}q_i)]/n_i'$  (Appendix B), where  $w_{av}q_i$  gives values for the average fitness line shown in Fig. 1. The slope of this average fitness line is  $b-c$  (see Fig. 1) or the degree of non-zero-sumness in the NPD model. To assess what between-group selection would do on its own, we did a version of this run (all parameters held the same) that neutralizes within-group selection by giving *both* cooperators and defectors the same average group fitness within each of the two groups. When we do this, we find that the change in  $Q$  over generations matches the Price between-group component,  $\Delta Q_B$ , *exactly*. Thus the average slope of the NPD fitness functions (degree of non-zero-sumness) is a measure of the between-group selection force predicted by the Price equation, i.e., groups with higher fractions of cooperators get proportionally more of the non-zero-sum advantage and outproduce groups with fewer cooperators.

We do a symmetrical assessment which isolates the effect of within-group selection by setting the slope  $b$  to zero. Here, there is a difference  $c$  between cooperator and defector fitness within groups but no difference between what these types get in different groups. In this case, there is no between-group selection. The resulting actual  $\Delta Q$  for this run exactly matches the alt  $\Delta Q_W$  component given by the alternative decomposition. Thus the parameter  $c$  in our model is a measure of the within-group selective force indicated by our alternative decomposition. The fact that between-group selection acting alone matches the Price between-group component, and within-group selection acting alone matches our alternative within-group component confirms the different idealization made in each selection decomposition as discussed above.

Note that neither decomposition is necessarily accurate when there is a mixture of between- and within-group selective forces acting simultaneously. While both decompositions assume the forces can be decoupled, in reality they affect each other. The Price decomposition posits in its between-group term that all change is due to between-group selection and then assumes the difference between this assumption and the actual change in  $Q$  is due to the counterbalancing force of within-group selection. The alternative decomposition takes the opposite tack. When the two approaches give roughly the same answer (as in the first few generations shown in Fig. 4), then the forces are roughly decomposable. But as change over longer periods is compared the values given by the two approaches diverge (e.g. compare values at generation 30 in Fig. 4). This is an indication that over this time period the strength of between-group selection has been affected by changes in group compositions (caused by within-group selection); and that the strength of within-group selection has been affected by group size changes (caused by between-group selection). The alternative decomposition presented here is a symmetrical complement to Price's decomposition. It

highlights the non-independence of the between- and within-group selective forces and may give more intuitive results than the Price equation in some situations.

### 4. Game theory and strong vs. weak altruism

Although evolutionary game theory (Maynard Smith, 1982; Maynard Smith and Price, 1973) has been very useful in helping biologists reason about evolutionary outcomes, unnecessary confusion and controversy has been generated by the way game-theoretic ideas have sometimes been applied to natural selection. In a classic game-theoretic analysis rational self-interest is defined by the action that produces the highest utility (fitness) to a player—regardless of the effect this behavior has on the utility of other players (i.e., positive or negative externalities). Categorizing behavior without regard to externalities is also a feature of the “byproduct mutualism” concept (Dugatkin, 2002). From the classic game-theoretic viewpoint, dynamics are said to be governed by individuals maximizing their own absolute utility. But what drives natural selection are *differences* in offspring number in subsequent generations, i.e., *relative* fitness, and fitness differences are also important in defining altruism (Boyd, 1988). Applying the classic game-theoretic (absolute fitness) viewpoint to systems under selection can lead to incorrect predictions. This is illustrated below.

We stated earlier that a condition for an NPD in our model requires  $c > b/n_i$ . That is, the cost of being a cooperator must be greater than the cooperator's share of the benefit it creates for the group. This condition, based on whether there is an absolute fitness cost or only a relative cost, also marks the boundary between strong and weak altruism (Wilson, 1979, 1990). While the NPD corresponds to strong altruism, weak altruism corresponds to a game called “Spite” in the game theory literature (Hamburger, 1979; Rapoport and Guyer, 1978) or described more recently in terms of a tandem bicycle contest (Kerr and Godfrey-Smith, 2002).<sup>1</sup> Here we will simply refer to this game as weak altruism. Again note that the same behavioral trait (with the same values of  $b$  and  $c$ ) can change between strong and weak altruism depending on changes in group size (Pepper, 2000).

In terms of Fig. 1, the distinction between strong (NPD) and weak altruism is determined by the intercept difference in the fitness lines ( $c$ ) and the slope of these lines ( $b$ ), for a given group size ( $n_i$ ). Weak altruism occurs if the fitness lines are either close enough together (small  $c$ ) or steep

<sup>1</sup>The name “Spite” refers to the idea that only a spiteful player would defect in this situation because cooperation is a dominant strategy that maximizes absolute utility (fitness) regardless of what others do. The highest payoff requires mutual cooperation and this is emphasized in the tandem bicycle formulation of this game, where both players must cooperate to win the race. Defecting when cooperation is a dominant strategy makes sense if enhancing relative rather than absolute fitness is the decision criteria, but the importance of *relative* fitness is not emphasized in the stories that give rise to these game titles.

enough (large  $b$ ) that hypothetically switching one D player to a C moves  $q_i$  enough to the right on the  $x$ -axis that the absolute fitness of this individual would increase. From a classic game theory perspective, the expected dynamics in the weak altruism game is towards mutual cooperation which is the dominant strategy, but under selection the dynamics of weak altruism move towards mutual defection within groups. Again, this is because under selection differences in fitness—not the maximization of absolute fitness—drive the dynamics (Wilson, 2004).

Problems with the distinction between strong vs. weak altruism can be illustrated in our model for the runs used earlier (Figs. 2 and 3) where all runs begin under strong altruism (NPD) conditions (see Table 1). For example, note that in Run 1 the first group's size shrinks below 5 between generation 15 and 16 (Fig. 2C), and thus crosses the boundary from strong to weak altruism ( $b = 1$ ,  $c = 0.2$ , and altruism is weak when  $n_i < b/c$  or  $n_i < 5$ ). Under the classic view (focused on absolute fitness), cooperation should now be favored within this group because cooperation is the dominant strategy with the highest absolute fitness payoff for each player. Therefore cooperation should increase when  $n_i < 5$  and decrease when  $n_i > 5$  for these model parameters. Yet, in reality, cooperation (now weak altruism) continues to be steadily selected against and we observe that the equilibrium for this group is mutual defection. This is because even a weak altruist helps every other group member to gain more fitness than any increase in fitness that it gives itself.

Still some continue to argue for a fundamental distinction based on absolute fitness, preferring to reserve the word “altruism” only for cases where there is an absolute fitness sacrifice (e.g., Foster et al., 2006; Lehmann and Keller, 2006; Maynard Smith, 1998; Nunnery, 1985, 2000). For instance, Nunnery (2000) prefers the term “benevolence” instead of “weak altruism” and claims that: “Benevolent traits spread by individual selection and are not vulnerable to cheating”. This conclusion is based on models where random groups are reformed every generation, but when these conditions do not hold (as in our model above) we see that weak altruists do *not* spread by within-group selection and *are* vulnerable to exploitation by defectors.

Several studies have reported “surprising” results when animal or human subjects choose self-sacrificing behaviors that do not maximize absolute fitness or utility, but instead seem more concerned with relative fitness or fairness (Boyd et al., 2003; Brosnan and de Waal, 2003; Fehr and Gächter, 2002). Yet these results are not surprising from the relative, rather than absolute, perspective of selection. In fact more than 35 years ago Hamilton (1971) in discussing the PD cautioned: “Natural selection ... seems to give one clear warning about situations of this general kind. When payoffs are connected with fitness, the animal part of our nature is expected to be more concerned with getting ‘more than the average’ than with getting ‘the maximum possible’”. We illustrate the dynamic similarities

of strong and weak altruism under selection in the next section.

## 5. Maintaining between-group selection

So far with our NPD model we have explored the simplest case of multilevel selection where the within-group level is represented by only two alternative strategies and the between-group level is represented by only two distinct groups. As we have seen, in this case the increase in cooperators is transient because once one group dominates there is no longer a between-group selection force (the variance ratio goes to zero) and the within-group more fit defect strategy takes over this single group. In order to illustrate the critical role which variance among groups plays in the evolution of altruism, we modify our model to include periodic random redistributions of the population into multiple groups, these redistributions occurring after varying numbers of generations within groups. Between redistribution events reproduction takes place and group sizes vary with cooperator and defector fitness as previously explained. This modification implements a population structure intermediate between two classic models: the Hamilton (1975) group selection model and the Maynard Smith (1964) haystack model.

In Hamilton's model, an infinite well-mixed population is formed into evenly sized groups based on a random binomial distribution. Interactions take place within groups where the benefit altruists give is divided equally among other group members. After one generation in groups, the population is again well mixed and groups randomly reformed. In this model cooperators are selected against because, as Hamilton showed, the variance among groups after random group formation is not enough for between-group selection to be stronger than within-group selection. In contrast, in Maynard Smith's haystack model groups exist for many generations until each becomes completely fixed for either cooperation or defection, after which there is global mixing and group reformation. In this model altruism is also selected against, but here it is because the advantage that cooperator-dominated groups have over defector-dominated groups is lost by waiting until all mixed groups are taken over by defectors (Wilson, 1987).

While natural populations do not match either of these extremes, both Hamilton's and Maynard Smith's conclusions that altruism cannot evolve in their models have been widely cited. Our simple model with intermediate numbers of generations in-between random group reformation events also does not match any particular natural population, but it illustrates what can happen if between-group selection has more time to act than in Hamilton's model, but within-group selection has less time to operate than in Maynard Smith's model. The role of an intermediate number of generations within groups is explored more fully elsewhere (Fletcher and Zwick, 2004; Wilson, 1987). Here our purpose is to use our simple model to illustrate the

dynamic similarities between strong and weak altruism when we move away from the extreme of random mixing every generation. To guarantee that only strong altruism is operating (regardless of changes in group size) we use *other-only* altruism (Pepper, 2000) in which an altruist gives nothing to itself and its benefit is divided evenly among the *others* in its group. In the other-only case the  $x$ -axis in Fig. 1 would be the fraction of *others* cooperating in a group and the  $q_i$  value for calculating  $a'_i$  in Eq. (1) becomes  $(a_i - 1)/(n_i - 1)$  and for calculating  $s'_i$  in Eq. (2) becomes  $a_i/(n_i - 1)$ ; i.e., the fraction of cooperators in others.

We contrast this other-only (strong) altruism with the whole-group altruism we have used thus far. Fig. 5 illustrates the relationship between the strong/weak and other-only/whole-group distinctions. As mentioned earlier, the value of  $r$  from Hamilton's rule differs for these two types (Pepper, 2000) in that the calculation of the average genotype of "others" that each focal individual interacts with includes the actor in the whole-group case. For other-only altruism the expected value of  $r$  when groups are formed at random is zero (Hamilton, 1975) so on average no positive values of  $b$  and  $c$  will satisfy Hamilton's rule ( $rb > c$ ). For whole-group altruism, randomly formed groups of size  $n_i$  produce an average  $r$  value of  $1/n_i$  (Wilson, 1979, 1990) so Hamilton's condition becomes  $b/n_i > c$  which is the definition of weak altruism; hence the consensus that strong altruism cannot evolve via randomly formed groups (Hamilton, 1975; Maynard Smith, 1998; Nunney, 1985, 2000; Sober and Wilson, 2000; Wilson, 1975, 1990). Elsewhere we have shown this conclusion is incorrect when groups in classic models of altruism are made multigenerational (Fletcher and Zwick, 2004).

We contrast the dynamics of other-only (strong) and whole-group (weak) altruism for individual runs in Fig. 6 and then give aggregate results in Table 2. By definition, the other-only runs are guaranteed to involve strong altruism whereas the whole-group runs begin with weak altruism conditions for comparison. For the parameters used here, altruism becomes strong even for the whole-group type of benefit distribution within individual groups where  $n_i > 20$ , but remains weak in groups smaller than 20. Our purpose here is to illustrate that the distinction between strong and weak altruism is not fundamental (Fletcher and Doebeli, 2006), i.e., that the dynamic trajectories can be similar (except for an initial transient)

whole-group (actor gives itself $x$ )		other-only (actor gives itself nothing)
weak altruism ( $x > c$ )	strong altruism ( $x < c$ ) ; ( $x = 0$ )	

Fig. 5. Relationship between whole-group vs. other-only altruism (defined by whether the altruist receives a share ( $x$ ) of the benefit it gives its group, or not) and weak vs. strong altruism (defined by whether  $x$  is greater or less than the altruist's cost ( $c$ )).  $x = b/n_i$  when the benefit ( $b$ ) an altruist gives is shared equally among a group of size  $n_i$ . In the case of other-only altruism  $x = 0$ .

if interactions are not always completely random. Here this deviation from randomness is caused by multiple generations within groups before random mixing.

Fig. 6 shows that the  $Q$  values for both other-only and whole-group altruism follow the same familiar hump-shaped pattern seen in Figs. 2 and 3 except that in the other-only runs  $Q$  decreases *initially*. The similar general dynamics of the other-only and whole-group runs is another indication of the similarities between strong and weak altruism under selection—they both loose ground to defectors within groups and can only increase due to the Simpson's paradox effect. That is, cooperation increases because altruist-dominated groups contribute proportionally more offspring to the overall population, even while the fraction of cooperators decreases in every group. The initial decrease in  $Q$  for other-only runs is expected—on average  $r = 0$  after groups are formed in these runs and

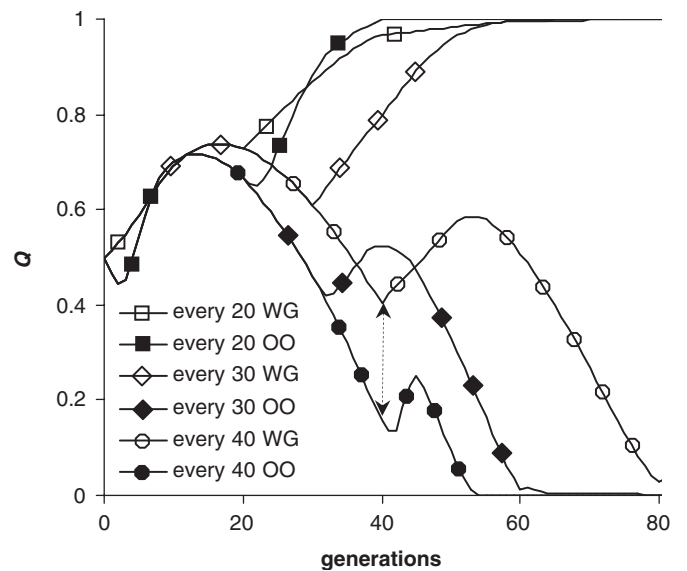


Fig. 6. Frequency of cooperators ( $Q$ ) vs. generation for six runs where an initial population of 500 cooperators and 500 defectors is randomly assigned to 100 groups and periodically randomly redistributed into groups. Three runs each are done for whole-group (WG) and other-only (OO) altruistic benefit, for redistribution frequencies of 20, 30, and 40 generations. For all runs  $b = 20$  and  $c = 1.0$ . The vertical dashed line with arrows highlights that WG vs. OO run dynamics are initially different immediately after group reformation (here at generation 40), but this difference soon disappears and all the OO runs show the same hump-shaped dynamics as the WG runs.

Table 2

Percentage of runs ending in  $Q = 1.0$ , where 100 runs were done for both whole-group (WG) and other-only (OO) benefit distribution for each reformation frequency (same parameters as Fig. 6)

	Reformation frequency					
	1	10	20	30	40	50
WG (%)	100	100	100	66	12	12
OO (%)	0	100	65	10	8	7

thus within-group selection is stronger than between-group selection. An example of this initial dynamical difference immediately after group reformation for other-only vs. whole-group runs is highlighted by the dashed line with arrows in Fig. 6. Although altruism initially decreases in the other-only (strong altruism) runs, within a few generations after group reformation,  $Q$  surprisingly begins to increase. It then follows the familiar hump-shaped pattern seen in the whole-group runs until another group reformation event. This behavior is due to the  $r$  value increasing while multiple generations are spent within groups, even as  $Q$  decreases (Fletcher and Zwick, 2004). In runs where groups are reformed close to when  $Q$  peaks (e.g. reformation every 20 generations), cooperation can ratchet up to saturation, whereas in runs where groups are reformed long after the peak in  $Q$  (e.g. reformation every 40 generations), cooperation tends to be eliminated. This is true for both the strong and weak altruism runs in each pair.

The results in Fig. 6 show typical individual runs done with the same initial random number seed for comparison. We also did 100 runs at each reformation frequency with the same parameters but different seeds. In this model, we allow fractional counts of cooperators and defectors from generation to generation, but do not use counts less than one in group reformation events. This tends to weed out residual fractions and because there is no mutation, the extremes of  $Q = 1.0$  and  $0.0$  act as attractors and intermediate values do not persist indefinitely. In all cases runs were done until either cooperator or defector saturation was reached. Table 2 shows the percentage of whole-group and other-only runs reaching cooperator saturation for these parameter conditions as well as shorter and longer periods between group reformations. These results support the trends discussed above while also confirming that for the same parameter values, weak altruism evolves more readily than strong (Wilson, 1979, 1990).

## 6. Summary

By embodying non-zero-sumness, population structure (assortment), and heredity in their most basic forms, this NPD model offers a simple framework for understanding the paradoxical nature of the evolution of altruism, integrating such central concepts as Hamilton's rule, Simpson's paradox, and the Price covariance equation. It also suggests an alternative selection decomposition which is more intuitive in some situations and helps emphasize the coupled nature of within- and between-group selection acting over multiple generations. We show that a game-theoretic approach is also useful in understanding the similarities between weak and strong altruism undergoing selection. We contrast other-only (strong) and whole-group (initially weak) versions of the NPD model to highlight both their initial differences immediately after random group formation and their

overall dynamical similarities. Finally, this also illustrates, in contrast to conventional wisdom, that *both* strong and weak altruism can evolve in periodically randomly formed groups as long as they are multigenerational.

Recently, game-theoretic models have been demonstrated where cooperation increases even without reciprocity. In these cases, social interactions are clumped by various mechanisms including, the presence of non-players (Aktipis, 2004; Hauert et al., 2002), the need for sufficiently similar arbitrary tags (Riolo et al., 2001), and social institutions for conformity within groups (Bowles et al., 2003; Boyd et al., 2003). From the perspective of the model presented here, we would expect these results with their various cost, benefit, and population structure parameters also to conform to Hamilton's rule, although this kind of analysis is not usually undertaken in such papers (Bowles et al., 2003 is an exception). In addition, recent studies have demonstrated mechanisms that can tip the balance of opposing levels of selection towards altruism, including the stochasticity inherent in finite populations (Fletcher and Zwick, 2004; Nowak et al., 2004), resource heterogeneity (Pepper and Smuts, 2002), imitation of social norms (Boyd and Richerson, 2002), and lotteries among non-kin that reduce competition within groups (Avilés et al., 2004). Understanding how fitness payoff structure, levels of selection, and population assortment interact can also elucidate how levels of biological hierarchy evolve (Margulis, 1993; Maynard Smith and Szathmáry, 1995; Michod, 1997).

The evolution of altruism does not require either reciprocity or kinship. What is essential is only sufficiently non-zero-sum benefits for heritable altruistic behaviors, and sufficiently non-uniform interactions among individuals with these behaviors. This is captured in our simple NPD model. As demonstrated here, the necessary combination of non-zero-sumness and population structure is specified by Hamilton's rule. This overall framework can help social science researchers who emphasize game-theoretic models to see their results in the context of Hamilton's rule (which applies to both inclusive fitness and multilevel selection theories), while also enabling biology researchers who focus on relatedness to recognize the inherent game-theoretic character of their models.

## Acknowledgments

We are grateful to L. Avilés, N. Lehman, G. Lendaris, M. Murphy, and D. Wilson for their careful review of an earlier draft, as well as to three anonymous reviewers. For useful discussions on this topic we additionally thank M. Doebeli, R. Gadagkar, P. Hammerstein, A. Joshi, L. Keller, B. Kerr, L. Nunney, J. Pepper, E. Szathmáry, and M. Woyciechowski. Financial support was provided to J.A.F. by the National Science Foundation International Research Fellowship Program.

**Appendix A. Derivation of Hamilton’s rule from NPD fitness functions**

Here we derive Hamilton’s rule starting with the basic NPD fitness functions and the condition that the fraction of cooperators in the whole population increases.

Starting with  $Q' > Q$  we get

$$\frac{A'}{N'} > \frac{A}{N}. \tag{A.1}$$

Using the assumption that  $w_0 = c$ , we simplify Eqs. (3) and (4) to

$$w_a(q_i) = bq_i \tag{A.2}$$

and

$$w_s(q_i) = bq_i + c, \tag{A.3}$$

respectively. The critical factor is the difference  $c$  between the fitness of altruistic (C) and selfish (D) individuals. Setting the base fitness to  $c$  just allows us to add this difference to  $w_s$  rather than subtracting it from  $w_a$ , and thereby avoid the possibility of negative fitness values. Substituting with these equations and summing over all groups  $i$  we get

$$\frac{A + \sum a_i q_i b}{N + \sum a_i q_i b + \sum s_i (q_i b + c)} > \frac{A}{N}. \tag{A.4}$$

Cross multiplying and isolating  $c/b$  on the right side yields

$$\frac{N(\sum a_i q_i - QA)}{NQS} > \frac{c}{b}. \tag{A.5}$$

We cancel  $N$  and substitute terms to give

$$\frac{\sum a_i q_i - 2AQ + AQ}{A(1 - Q)} > \frac{c}{b} \tag{A.6}$$

and then

$$\frac{\sum n_i q_i^2 - 2Q \sum n_i q_i + Q^2 \sum n_i}{A - 2AQ + AQ} > \frac{c}{b}. \tag{A.7}$$

This we rewrite as

$$\frac{\sum n_i (q_i^2 - 2Qq_i + Q^2)}{A - 2AQ + NQ^2} > \frac{c}{b} \tag{A.8}$$

which gives

$$\frac{\sum n_i (q_i - Q)^2}{A(1-Q)^2 + S(0-Q)^2} > \frac{c}{b} \tag{A.9}$$

or

$$\frac{\text{var}_B(q_i)}{\text{var}_T(Q)} > \frac{c}{b}. \tag{A.10}$$

This is Hamilton’s rule for a whole-group trait where  $r$  is the between-group over total variance:  $rb > c$ .

**Appendix B. Price equation derivations**

Here we show how the Price covariance equation can be interpreted in terms of an idealized  $Q'$  and how this relates to our NPD fitness functions.

Starting with the Price between-group term,

$$\Delta Q_B = \frac{\text{cov}(w_i, q)}{E(w_i)}, \tag{B.1}$$

where  $w_i$  is the growth rate of a group,  $n'_i/n_i$ . By definition we get

$$\Delta Q_B = \frac{E(w_i q_i)}{E(w_i)} - \frac{E(w_i)E(q_i)}{E(w_i)}. \tag{B.2}$$

This can be written as

$$\Delta Q_B = \frac{\frac{1}{N} \sum n_i w_i q_i}{N'/N} - Q. \tag{B.3}$$

Canceling  $1/N$  and using the definition of  $w_i$  we get

$$\Delta Q_B = \frac{\sum n'_i q_i}{N'} - Q \tag{B.4}$$

which gives

$$\Delta Q_B = Q^* - Q, \tag{B.5}$$

where

$$Q^* = \frac{\sum n'_i q_i}{N'}. \tag{B.6}$$

Now given  $\Delta Q = Q' - Q$  and  $\Delta Q = \Delta Q_B + \Delta Q_W$ , it follows that  $\Delta Q_W = Q' - Q^*$ .

To show the connection to the NPD fitness functions, we start with

$$Q^* = \frac{\sum n'_i q_i}{N'}, \tag{B.7}$$

then from above we substitute for  $n'_i$  using average fitness to get

$$Q^* = \frac{\sum [q_i n_i (1 + w_{av} q_i)]}{N'} \tag{B.8}$$

or

$$Q^* = \frac{\sum [a_i (1 + w_{av} q_i)]}{N'} = \frac{A^*}{N'}, \tag{B.9}$$

where  $A^*$  is the idealized  $A'$  given by the Price between-group component of selection. That is, the Price equation’s idealization about the new number of altruists in the population is given by adding the existing number in each group,  $a_i$ , to the amount each receives based on the average fitness line in the NPD model (Fig. 1).

**References**

Aktipis, C.A., 2004. Know when to walk away: contingent movement and the evolution of cooperation. *J. Theor. Biol.* 231, 249–260.  
 Avilés, L., 2002. Solving the freeloaders paradox: genetic associations and frequency-dependent selection in the evolution of cooperation among nonrelatives. *Proc. Natl Acad. Sci. USA* 99, 14268–14273.

- Avilés, L., Fletcher, J.A., Cutter, A., 2004. The kin composition of groups: trading group size for degree of altruism. *Am. Nat.* 164, 132–144.
- Axelrod, R., 1984. *The Evolution of Cooperation*. Basic Books, Inc., New York.
- Axelrod, R., Hamilton, W.D., 1981. The evolution of cooperation. *Science* 211, 1390–1396.
- Bowles, S., Choi, J.-K., Hopfensitz, A., 2003. The co-evolution of individual behaviors and social institutions. *J. Theor. Biol.* 223, 135–147.
- Boyd, R., 1988. Is the repeated prisoner's dilemma a good model of reciprocal altruism? *Ethol. Sociobiol.* 9, 211–222.
- Boyd, R., Richerson, P.J., 2002. Group beneficial norms can spread rapidly in a structured population. *J. Theor. Biol.* 215, 287–296.
- Boyd, R., Gintis, H., Bowles, S., Richerson, P.J., 2003. The evolution of altruistic punishment. *Proc. Natl Acad. Sci. USA* 100, 3531–3535.
- Breden, F., 1990. Partitioning of covariance as a method for studying kin selection. *Trends Ecol. Evol.* 5, 224–228.
- Brosnan, S.F., de Waal, F.B.M., 2003. Monkeys reject unequal pay. *Nature* 425, 297–299.
- Dugatkin, L.A., 1997. *Cooperation Among Animals, An Evolutionary Perspective*. Oxford University Press, New York.
- Dugatkin, L.A., 2002. Cooperation in animals: an evolutionary overview. *Biol. Philos.* 17, 459–476.
- Fehr, E., Gächter, S., 2002. Altruistic punishment in humans. *Nature* 415, 137–140.
- Fletcher, J.A., Doebeli, M., 2006. How altruism evolves: assortment and synergy. *J. Evol. Biol.* 19, 1389–1393.
- Fletcher, J.A., Zwick, M., 2004. Strong altruism can evolve in randomly formed groups. *J. Theor. Biol.* 228, 303–313.
- Fletcher, J.A., Zwick, M., 2006. Unifying the theories of inclusive fitness and reciprocal altruism. *Am. Nat.* 168, 252–262.
- Foster, K.R., Wenseleers, T., Ratnieks, F.L.W., 2006. Kin selection is the key to altruism. *Trends Ecol. Evol.* 21, 57–60.
- Frank, S.A., 1995. George Price's contributions to evolutionary genetics. *J. Theor. Biol.* 175, 373–388.
- Frank, S.A., 1998. *Foundations of Social Evolution*. Princeton University Press, Princeton.
- Hamburger, H., 1979. *Games as Models of Social Phenomena*. W.H. Freeman and Co., New York.
- Hamilton, W.D., 1964. The genetical evolution of social behavior I and II. *J. Theor. Biol.* 7, 1–52.
- Hamilton, W.D., 1970. Selfish and spiteful behavior in an evolutionary model. *Nature* 228, 1218–1220.
- Hamilton, W.D., 1971. Selection of Selfish and Altruistic Behavior in some Extreme Models, Man and Beast: Comparative Social Behavior. Smithsonian Institution Press, Washington, DC, pp. 59–91.
- Hamilton, W.D., 1972. Altruism and related phenomena, mainly in social insects. *Annu. Rev. Ecol. Syst.* 3, 193–232.
- Hamilton, W.D., 1975. Innate social aptitudes of man: an approach from evolutionary genetics. In: Fox, R. (Ed.), *Biosocial Anthropology*. Wiley, New York, pp. 133–155.
- Hardin, G., 1968. The tragedy of the commons. *Science* 162, 1243–1248.
- Hardin, R., 1971. Collective action as an agreeable N-Prisoner's dilemma. *Behav. Sci.* 16, 472–481.
- Hauert, C., De Monte, S., Hofbauer, J., Sigmund, K., 2002. Volunteering as red queen mechanism for cooperators in public goods games. *Science* 296, 1129–1132.
- Kennedy, D. (Ed.), 2003. *Tragedy of the Commons?* Science, vol. 302, pp. 1861–1928.
- Kerr, B., Godfrey-Smith, P., 2002. Individualist and multi-level perspectives on selection in structured populations. *Biol. Philos.* 17, 477–517.
- Lehmann, L., Keller, L., 2006. The evolution of cooperation. A general framework and a classification of models. *J. Evol. Biol.* 19, 1365–1376.
- Leigh, E.G., 1999. Levels of selection, potential conflicts, and their resolution: the role of the "Common Good". In: Keller, L. (Ed.), *Levels of Selection in Evolution*. Princeton University Press, Princeton, pp. 15–30.
- Margulis, L., 1993. *Symbiosis in Cell Evolution: Microbial Communities in the Archean and Proterozoic Eons*. W.H. Freeman and Co., New York.
- Maynard Smith, J., 1964. Groups selection and kin selection. *Nature* 201, 1145–1147.
- Maynard Smith, J., 1982. *Evolution and the Theory of Games*. Cambridge University Press, Cambridge.
- Maynard Smith, J., 1998. The origin of altruism. *Nature* 393, 639–640.
- Maynard Smith, J., Price, G.R., 1973. The logic of animal conflict. *Nature* 246, 15–18.
- Maynard Smith, J., Szathmáry, E., 1995. *The Major Transitions in Evolution*. W.H. Freeman and Co., Oxford, England.
- McMillan, J., 1979. The free-rider problem: a survey. *Econ. Rec.* 55, 95–107.
- Michod, R.E., 1997. Evolution of the individual. *Am. Nat.* 150, S5–S21.
- Michod, R.E., 1999. Individuality, immortality, and sex. In: Keller, L. (Ed.), *Levels of Selection in Evolution*. Princeton University Press, Princeton, pp. 35–74.
- Nowak, M.A., Sasaki, A., Taylor, C., Fudenberg, D., 2004. Emergence of cooperation and evolutionary stability in finite populations. *Nature* 428, 646–650.
- Nunney, L., 1985. Group selection, altruism, and structured-deme models. *Am. Nat.* 126, 212–230.
- Nunney, L., 2000. Altruism, benevolence and culture: commentary discussion of Sober and Wilson's 'Unto others'. *J. Consciousness Stud.* 7, 231–236.
- Pepper, J.W., 2000. Relatedness in trait group models of social evolution. *J. Theor. Biol.* 206, 355–368.
- Pepper, J.W., Smuts, B.B., 2002. Assortment through environmental feedback. *Am. Nat.* 160, 205–213.
- Price, G.R., 1970. Selection and covariance. *Nature* 227, 520–521.
- Queller, D.C., 1985. Kinship, reciprocity and synergism in the evolution of social behaviour. *Nature* 318, 366–367.
- Queller, D.C., 1992. Quantitative genetics, inclusive fitness, and group selection. *Am. Nat.* 139, 540–558.
- Rapoport, A., Guyer, M., 1978. A taxonomy of  $2 \times 2$  games. *Gen. Syst.* XXIII, 125–136.
- Riolo, R.L., Cohen, M.D., Axelrod, R., 2001. Evolution of cooperation without reciprocity. *Nature* 414, 441–443.
- Simpson, E.H., 1951. The interpretation of interaction in contingency tables. *J. Roy. Stat. Soc. B* 13, 238–241.
- Sober, E., Wilson, D.S., 1998. *Unto Others, The Evolution and Psychology of Unselfish Behavior*. Harvard University Press, Cambridge, MA.
- Sober, E., Wilson, D.S., 2000. Morality and 'Unto others': response to commentary discussions. *J. Consciousness Stud.* 7, 257–268.
- Trivers, R.L., 1971. The evolution of reciprocal altruism. *Q. Rev. Biol.* 46, 35–57.
- Wade, M.J., 1978. A critical review of the models of group selection. *Q. Rev. Biol.* 53, 101–114.
- Wade, M.J., 1980. Kin selection: its components. *Science* 210, 665–667.
- Wilson, D.S., 1975. A theory of group selection. *Proc. Natl Acad. Sci. USA* 72, 143–146.
- Wilson, D.S., 1977. Structured demes and the evolution of group-advantageous traits. *Am. Nat.* 111, 157–185.
- Wilson, D.S., 1979. Structured demes and trait-group variation. *Am. Nat.* 113, 606–610.
- Wilson, D.S., 1987. Altruism in Mendelian populations derived from sibling groups: the haystack model revisited. *Evolution* 41, 1059–1070.
- Wilson, D.S., 1990. Weak altruism, strong group selection. *Oikos* 59, 135–140.
- Wilson, D.S., 1997. Altruism and organism: disentangling the themes of multilevel selection theory. *Am. Nat.* 150, S122–S134.
- Wilson, D.S., 2004. What is wrong with absolute individual fitness? *Trends Ecol. Evol.* 19, 245–248.