Measures of Relative Fitness of Social Behaviors in Finite Structured Population Models

Corina E. Tarnita[1,*] and Peter D. Taylor[2]

1. Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544; 2. Department of Mathematics and Statistics, Queen's University, Kingston, Ontario K7L 3N6, Canada

Submitted August 23, 2013; Accepted May 29, 2014; Electronically published September 5, 2014

Online enhancement: appendixes.

abstract: How should we measure the relative selective advantage of different behavioral strategies? The various approaches to this question have fallen into one of the following categories: the fixation probability of a mutant allele in a wild type population, some measures of gene frequency and gene frequency change, and a formulation of the inclusive fitness effect. Countless theoretical studies have examined the relationship between these approaches, and it has generally been thought that, under standard simplifying assumptions, they yield equivalent results. Most of this theoretical work, however, has assumed homogeneity of the population interaction structure—that is, that all individuals are equivalent. We explore the question of selective advantage in a general (heterogeneous) population and show that, although appropriate measures of fixation probability and gene frequency change are equivalent, they are not, in general, equivalent to the inclusive fitness effect. The latter does not reflect effects of selection acting via mutation, which can arise on heterogeneous structures, even for low mutation. Our theoretical framework provides a transparent analysis of the different biological factors at work in the comparison of these fitness measures and suggests that their theoretical and empirical use needs to be revised and carefully grounded in a more general theory.

Keywords: evolutionary game theory, heterogeneous networks, fixation probability, gene frequency change, inclusive fitness.

Introduction

Ever since Darwin's work, evolutionary biologists have fastened their attention on understanding the factors that cause interacting genetic traits to increase or decrease in frequency. Starting in the middle of the twentieth century, the bulk of theoretical work in this area focused on social traits, those whose expression affects the fitness of neighbors, particularly those neighbors who might also be carrying the genes responsible for the trait. Two of the principle architects of this program are Bill Hamilton (1964) and John Maynard Smith (1982), and their concepts of inclusive fitness and evolutionary game theory give significant shape to the thinking of biologists today.

The right way to measure the relative fitness of a given social behavior has, however, remained a matter of contention among evolutionary biologists, and even recently, this question has sparked heated debate in the context of the evolution of eusociality (Doebeli 2010; Nowak et al. 2010; van Veelen et al. 2010; Abbott et al. 2011; Ferriere and Michod 2011).

This question needs to be answered in the context of a more or less precise model, and there are a variety of such models that have been developed. To make analytical progress, one needs a model that simplifies greatly the biological world, and a standard current approach is the one we adopt in this article—that the behavior is determined by a gene at a single locus at which mutation is rare enough that we need take account of only two alleles acting in the population at any particular time. We call this low mutation. In addition, because of the social nature of the trait, we need a precise population structure that allows us to keep track of interactions and offspring movements; for this, we use a graph structure (Killingback and Doebeli 1996; Hauert and Doebeli 2004; Ohtsuki et al. 2006; Taylor et al. 2007*a*). This is not as restrictive an assumption as one might suppose, as most of the standard population structures—for example, Wright's (1943) island model—can be modeled as finite or infinite graphs. In this article, we focus on finite populations, where the dynamics is governed by an interplay of selection, mutation, and drift. The populations are represented as finite graphs—a finite set of nodes, each occupied by a single asexual haploid breeder, together with a specification of two kinds of connections between nodes: the probability that two breeders will interact and the probability that an offspring of one breeder will displace another breeder. The graph is called homogeneous or transitive (Taylor et al. 2007*a*) if it looks

* Corresponding author; e-mail: ctarnita@princeton.edu.

Am. Nat. 2014. Vol. 184, pp. 477–488.
0003-0147/2014/18404-54919$15.00.
DOI: 10.1086/677924

the same from every node (i.e., all locations are equivalent). A special case of such a population structure is the well-mixed, panmictic population. Otherwise, the graph is called heterogeneous.

It is worth noting that in early behavioral models, the heterogeneity was typically generated by different roles (classes)—males and females, adults and juveniles, parents and helpers—and often the behavior studied and the production of offspring were restricted to certain classes, though they could affect the fitness of all classes. To some extent, our graph structure can accommodate such roles using nodes of different types. Not every type of node need play the game or have offspring. One thing we have not explicitly done is provide a mechanism for individuals to move from one node to another (e.g., to change classes), as happens in an age-structured model, though this is often modeled by having the individual be an offspring of itself. Another interpretation of heterogeneity would allow a neutral variation in fecundity among nodes (e.g., a population along a resource cline). We have not built this into our model, but we suspect that an analogue of our model for such a population would show behavior comparable to that found in the results of this work. The heterogeneity in our model stems from the relationship of a node to the population-wide specification of breeder interactions and offspring dispersal, and the class structure is determined by that. Heterogeneous populations, which constitute the focus of our article, are not only the norm in nature, but in the very special cases where they have been analytically tractable, they have shown interesting and unexpected behavior (Lieberman et al. 2005; Frean et al. 2013).

In each time interval (called an update step), individuals interact with partners determined by the population structure employing one of two strategies, *A* or *B*, determined genetically at a single locus by alleles *A* and *B*. A very slightly modified analysis can be employed for more than two strategies (e.g., Tarnita et al. 2011), but here, for simplicity of exposition, we focus on two. The result of the interactions between individuals is captured in the form of a general game with a given payoff matrix, and fitness depends on the payoff accumulated. As a consequence of the update step, the nodes change their strategies (through offspring replacement), but typically the underlying graph connections remain fixed (Lieberman et al. 2005; Ohtsuki et al. 2006; Taylor et al. 2007*a*), and in this case, we call the graph structure static. If the node connections can change as well, we call the structure dynamic. In the latter case, the definitions of homogeneity and heterogeneity discussed above do not immediately hold, since the structure evolves through different states. Some of our results can apply for such dynamical networks (Antal et al. 2009; Tarnita et al. 2009*a*; Taylor and Grafen 2010); however, our main results hold for fixed structures, and those will be the focus of our article.

In this context, our question becomes, how are we to measure the effect of selection on allele frequency? In the recent finite population literature, four different measures have been proposed. The first approach is based on the comparison of fixation probabilities, the second on assessing the effect of the mutation-selection process on the stationary equilibrium gene frequency, the third on the expected one-step selective change in gene frequency in the limit of low mutation, and the fourth, in the limit of low mutation and weak selection, on the one-step change in reproductive value (RV)-weighted gene frequency at the neutral stationary equilibrium. Under additional assumptions, this fourth measure turns out to be equivalent to the inclusive fitness effect. In "Model, Measures, and Overview of Results" below, we introduce and discuss each of these measures.

There are a number of studies that give precise conditions for the equivalence of these approaches (Rousset and Billiard 2000; Taylor et al. 2007*b*; app. A in Nowak et al. 2010), and two critical assumptions, made clear in these studies, are that the effects of selection are small ("weak" selection) and that the population structure is homogeneous (or transitive; Taylor et al. 2007*a*), which implies that individuals are more or less equivalent. In fact, in most of the literature studying these questions, this homogeneity is taken for granted, as evidenced by the use of terms such as "typical," "arbitrary," or "focal" to describe an individual. And a number of recent publications refer explicitly to the equivalence of these properties without any mention of this assumption (Grafen 2009; Wakano and Lehmann 2012; West and Gardner 2013).

It is not our purpose to be critical of this. In Grafen (2009), the author himself distinguishes between the gold and plastic standards of precision in modeling and points out that the literature needs both, one to be rigorous and the other to be useful. After all, there are no homogenous populations out there, and fitness effects are rarely weak, seldom additive. But the four measures we refer to above are the fundamental approaches that we all use to get hold of the effects of selection, and at some point, we need to have a precise, general, and rigorous understanding of their relationship with one another; in the absence of such an understanding, the plastic standard can lead to misleading results.

Our purpose in this article is to review the four fitness measures and to provide a transparent analysis of the different biological factors at work in their comparison. We first develop a general framework that makes heterogeneous networks analytically tractable ("Model, Measures, and Overview of Results"). We illustrate our results with a simple example of heterogeneous structures that shows

that the four measures are not, in general, equivalent ("Example: The Star Graph"). It emerges then that in heterogeneous populations, while there is a broad equivalence between the first three measures described (fixation probability, stationary gene frequency, and stationary one-step change in gene frequency), there is not, in general, an equivalence between all four, and the relationship between the first three and the last is quite sensitive to model assumptions.

Model, Measures, and Overview of Results

We consider a population of N individuals, each being one of two types, A or B. The genotype x_i of node i is defined to be 1 if individual i bears allele (has strategy) A and 0 if it bears allele B. A state S of the population is a specification of the genotypic values x_i for all nodes. Individuals interact according to the general matrix

$$\begin{array}{c} \\ A \\ B \end{array} \begin{array}{c} \begin{array}{cc} A & B \end{array} \\ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \end{array}. \tag{1}$$

As a result of these interactions, individual i receives a payoff ϵ_i, which contributes to its effective payoff $f_i = 1 + \delta\epsilon_i$. Here, δ is the selection coefficient. The limit of weak selection is $\delta \to 0$; when $\delta = 0$, the process is neutral. Individual birthrates b_i (fecundity) and death rates d_i (mortality) are functions of the set of effective payoffs.

During each time interval, after the game is played, an update occurs. Our analysis applies to any kind of update rule, as long as it is well behaved in the limit of weak selection (i.e., the transition probabilities between states are continuous and differentiable at $\delta = 0$) and the total birthrate is the same in each of the two pure states (all A and all B). Examples of such processes are the Moran process, with either birth-death or death-birth updating (Moran 1962; Ohtsuki et al. 2006), in which at the end of each time interval, exactly one individual dies and one individual reproduces, and the Wright-Fisher process (Imhof and Nowak 2006), in which, at the end of each time interval, all individuals die and the new generation is sampled from their offspring.

Reproduction occurs with errors, and we consider an asymmetric mutation rate. Thus, when an individual reproduces, it does so accurately with probability $1 - u$; with probability u, its offspring mutates and adopts strategy A with probability p and strategy B with probability $1 - p$. In this article, we are concerned with the limit of low mutation, $u \ll (1/N)$; in this limit, a new mutant either takes over the population or goes extinct before the next mutation event. Stable polymorphism is not possible in our finite population model, because as long as mutation is nonzero, no states are absorbing; moreover, when mutation is zero, the only absorbing states are the pure all-A and all-B states. Next, we discuss the four measures of selection employed to study strategy selection.

Measure 1: Comparison of Fixation Probabilities

The first measure is based on the comparison of fixation probabilities (Kimura 1962; Karlin and Taylor 1975). The fixation probability of an allele A is the probability that a single A player in an otherwise B population will eventually take over the population (with no mutation). To account for the fact that A players placed at different locations on a structure might have different fixation probabilities, we define $\rho_{A,i}$ to be the fixation probability of an A placed at node i in a population of B players. If a structure is homogeneous, then every position is effectively identical, and hence, a mutant has the same likelihood of appearing on any location. On a heterogeneous graph, however, locations are different, and calculating a total fixation probability is less straightforward. Mutants do not appear at random at every location, but instead there is a mutant appearance distribution, as discussed in Allen and Tarnita (2014). Such a distribution captures the probability that the first mutant appears at a given node and gives us the expression $\rho_A = \sum_i \Pr(\text{first mutant arises at location } i)\rho_{A,i}$. This is our general definition of fixation probability. Then a measure that strategy A is selected over strategy B is

$$\rho_A > \rho_B. \tag{2}$$

Measure 2: Stationary Equilibrium Gene Frequency

The second measure of selective success relies on assessing the effect of the mutation-selection process on the stationary equilibrium gene frequency. Let $x_{i,S}$ denote the genotype of node i in state S; then the frequency of gene A in state S is $x_S = (1/N)\sum_i x_{i,S}$. Since the frequency of A varies from state to state, we use as a measure the long-term average of the state frequencies (Rousset and Billiard 2000). For fixed mutation rate u and selection strength δ, the state frequency distribution will attain a long-term equilibrium with average allele frequency $\langle x \rangle = \sum_S x_S \pi_S(u, \delta)$, where $\pi_S(u, \delta)$ is the frequency of state S at this equilibrium.

The angle brackets $\langle \cdot \rangle$ represent the average taken over all population states, with state frequencies determined by the mutation-selection equilibrium. In what follows, we will drop the S to simplify the notation. In the absence of selection, we let $\langle x \rangle_0$ denote the average allele frequency at the neutral equilibrium and take our measure of the selective advantage of allele A to be the difference $\langle x \rangle - \langle x \rangle_0$. This measures the change in the long-term allele frequency as we move from the neutral equilibrium to the new equi-

librium, which balances mutation, drift, and selection. The neutral frequency $\langle x \rangle_0$ is easily seen to be p, determined by the effect of the asymmetric mutation (Taylor et. al 2007*b*). So the second measure of strategy success is

$$\langle x \rangle > p. \tag{3}$$

This condition is meant to apply to any structure and any intensity of selection and mutation.

Measure 3: Stationary Equilibrium One-Step Selective Change in Gene Frequency

The third measure of selective success is the classic measure from population genetics (Ewens 1969; Wright 1969; Crow and Kimura 1970), the expected one-step selective change in gene frequency. In a wide variety of population models, it has taken many forms; in our simple model, in a given state S and under the assumption that individuals breed true, the change due only to selection in the frequency of A individuals is $\Delta_S x_S^{\text{sel}} = (1/N) \sum_i x_{i,S}(b_{i,S} - d_{i,S})$. Again, in what follows, we will drop the S to simplify the notation. Since this quantity also changes between states, we employ the same long-term average, and the third measure of the selective advantage of A is that the stationary average one-step selective change is positive:

$$\langle \Delta x^{\text{sel}} \rangle > 0. \tag{4}$$

To relate this to the first two measures, we observe that at the mutation-selection equilibrium, the overall average one-step change in gene frequency will be zero. We write this total change as a sum of its two components, due to selection and to mutation

$$0 = \langle \Delta x^{\text{tot}} \rangle = \langle \Delta x^{\text{sel}} \rangle - \langle \Delta x^{\text{mut}} \rangle, \tag{5}$$

where we have taken the second term to be the negative change of A frequency due to mutation. This gives us the equilibrium equality $\langle \Delta x^{\text{sel}} \rangle = \langle \Delta x^{\text{mut}} \rangle$ representing the balance between selection and mutation. At the neutral equilibrium (i.e., in the absence of selection, $\delta = 0$), not only is the total change zero, but each of its two components, due to selection and due to mutation, is zero. However, in moving to the mutation-selection equilibrium, both components become nonzero. Thus, the mutation term should really be described as the effect of selection on the change due to mutation. Since we are concerned with the limit of low mutation and the frequency of mixed states is of order u, the rate at which mutation negatively affects the frequency of A is the difference between the rate at which it generates new copies of B in the all-A state (which is $u(1 - p)\mathbf{b}$) and the rate at which it generates new copies of A in the all-B state (which is $up\mathbf{b}$), where $\mathbf{b}$ is the average fecundity (assumed above to be the same in the two pure states). Thus, at low mutation,

$$\begin{aligned} \langle \Delta x^{\text{mut}} \rangle &= u\mathbf{b}[(1 - p)\pi_1 - p\pi_0] \\ &= u\mathbf{b}[\pi_1 - p] = u\mathbf{b}[\lim_{u \to 0} \langle x \rangle - p], \end{aligned} \tag{6}$$

where π_0 and π_1 are the frequencies of the pure states at the long-term equilibrium and the last two equalities are due to the fact that at low mutation, $\pi_0 + \pi_1 = 1$ and $\langle x \rangle = \pi_1$.

Using the relationship between the change due to selection and the change due to mutation given by (5), we are now able to formulate the equivalence of the first three measures of selective advantage, which, in the limit of low mutation, holds for any population structure, any intensity of selection, and any update rule for which the average birthrate is the same in each of the two pure states (as assumed above),

$$\rho_A > \rho_B \Leftrightarrow \lim_{u \to 0} \langle x \rangle > p \Leftrightarrow \langle \Delta x^{\text{sel}} \rangle^{(u)} > 0. \tag{7}$$

Here and henceforth, $h^{(y)}$ represents the coefficient of the linear term in the Taylor expansion of function h with respect to y. Both these equivalences have been shown previously—the former in Taylor et al. (2007*b*), equation (4.2), and Allen and Tarnita (2014) and the latter, in various forms, in Rousset and Billiard (2000) and appendix A in Nowak et al. (2010)—but here we extend those results to a larger class of processes and give a more straightforward proof. For details, see appendix A; appendixes A–D are available online. The significance of these equivalences is the possibility of using (4), the one-step selective change, to calculate the other two measures. Indeed, (4) can in principle be calculated from selective effects on the fitness of the different nodes, and this can be obtained from the effective payoffs and the pattern of offspring dispersal. In fact, the theory of inclusive fitness was designed to do exactly that. The difficulty, however, is that the calculation of the average change requires knowledge of the state distribution, and this is hard to get hold of when selection is acting, even in simple population models. It is this that has focused much of the theoretical work on the limit of weak selection, and the measure that has proven most receptive to this approach is the one-step selective change.

Measure 4 (Weak Selection, Homogeneous Structures): Neutral Stationary Equilibrium One-Step Selective Change in Gene Frequency

In the absence of selection, the term $\langle \Delta x^{\text{sel}} \rangle$ is zero. Moreover, selection acts on it in two ways: first, through its effect on allele fitness in each state and, second, through its effect on the distribution of states. We can thus write, to first order, the effect of selection on the one-step change in gene frequency as

$$\langle \Delta x^{\mathrm{sel}} \rangle = \delta(\langle \Delta^{(\delta)} x^{\mathrm{sel}} \rangle_0 + \langle \Delta^{(0)} x^{\mathrm{sel}} \rangle_\delta). \quad (8)$$

The first term above is the effect of individual fitness in each state averaged over the neutral distribution (denoted by $\langle \cdot \rangle_0$); the second term accounts for the effect of selection on the equilibrium distribution of states (denoted by $\langle \cdot \rangle_\delta$). Since the neutral distribution is generally accessible, what we can feasibly calculate is the first of these, $\langle \Delta^{(\delta)} x^{\mathrm{sel}} \rangle_0$. One way to think of this term is to imagine two competing alleles with identical phenotypic effects at frequency equilibrium under bidirectional mutation. Then an environmental change causes a phenotypic difference between them, and we measure the one-step change in the frequency of the focal allele, starting at its neutral equilibrium frequency distribution. But what about the second term above? It turns out that in a homogeneous population, the second term is zero, because in every state, the birthrates and death rates of every individual are the same at neutrality. Therefore, this has led to the widely used weak selection and low mutation measure of selective advantage

$$\langle \Delta^{(\delta)} x^{\mathrm{sel}} \rangle_0^{(u)} > 0. \quad (9)$$

But what about heterogeneous populations? It turns out that in a heterogeneous population, some nodes can have a higher fitness than others even at neutrality, and in a state in which those are the nodes that carry A, the one-step change in gene frequency will be positive: $\langle \Delta^{(0)} x^{\mathrm{sel}} \rangle > 0$. (In the star graph discussed in "Example: The Star Graph," the leaves have a higher fitness than the hub.) As a result of this, a perturbation of the state frequencies can affect the long-term average one-step change in allele frequency, and the last term in (8) can be nonzero. Therefore, in heterogeneous populations, (9) is not, in general, equivalent to the conditions in (7). The next measure we discuss makes ingenious use of the classic notion of reproductive value to handle the nonzero term in (8) for heterogeneous populations.

Measure 4 (Weak Selection, Heterogeneous Structures): Neutral Equilibrium One-Step Selective Change in RV-Weighted Gene Frequency

The problem we encountered above derived from a variation in fitness among the nodes of a heterogeneous network, even at neutrality. This can be fixed by assigning weights to the nodes so that, at neutrality, they all have the same weighted fitness—that is, at each node, the weighted birthrate and death rate will be equal. The weights that satisfy this have been described before (Taylor 1990, 2009; Leturque and Rousset 2002) and have been called reproductive values by analogy with Fisher (1930). Let v_i be the weight of node i. Then the weighted death rate and birthrate of node i are given by $\hat{d}_i = d_i v_i$ and $\hat{b}_i = b_i \sum_{j=1}^{N} p_{ij} v_j$, where p_{ij} is the probability that an offspring of individual i colonizes location j. Then the v_i are determined up to a multiplicative constant by the condition $\hat{d}_i = \hat{b}_i$ that the weighted birthrate and death rate are equal at neutrality for any node i. The classic notion of RV (Fisher 1930) is the long-term contribution of the node-i individual to the population; at equilibrium, the expected one-step increase in this contribution (through birth) must equal its expected decrease (through death), and that gives us exactly the condition $\hat{d}_i = \hat{b}_i$. Therefore, the two notions (the one used in this article and the classic notion of RV) are identical and generated by the same condition.

For homogeneous populations, all nodes have the same reproductive value (and we take $v_i = 1$ for all i). However, for heterogeneous populations, this will not typically be the case, and we work instead with the RV-weighted gene frequency. Therefore, for heterogeneous populations, the neutral average selective change in RV-weighted gene frequency will have the form $\langle \Delta^{(\delta)} \hat{x}^{\mathrm{sel}} \rangle_0$, where $\hat{x} = \sum_i x_i (\hat{b}_i - \hat{d}_i)$. We use the caret to differentiate between the unweighted and weighted measures, where there is a potential that they might be different. The detailed discussion of this RV-modified quantity and its relationship to the previous measures is provided in appendix B. The RV-weighted analogue of (8) becomes

$$\langle \Delta \hat{x}^{\mathrm{sel}} \rangle = \delta(\langle \Delta^{(\delta)} \hat{x}^{\mathrm{sel}} \rangle_0 + \langle \Delta^{(0)} \hat{x}^{\mathrm{sel}} \rangle_\delta), \quad (10)$$

and the effect of the RV weighting is, as under homogeneity, to make $\Delta^{(0)} \hat{x}^{\mathrm{sel}}$ equal to zero in every state so that $\langle \Delta \hat{x}^{\mathrm{sel}} \rangle^{(\delta)} = \langle \Delta^{(\delta)} \hat{x}^{\mathrm{sel}} \rangle_0$; we thus recover the analogue of (9) for heterogeneous populations,

$$\langle \Delta^{(\delta)} \hat{x}^{\mathrm{sel}} \rangle_0^{(u)} > 0. \quad (11)$$

For a homogeneous population, this measure becomes equivalent to the one in (9). In general, according to the existing literature, we take (11) to be the fourth measure of the selective advantage of A.

Whereas we (and others before; see Rousset and Billiard 2000; Nowak et al. 2010) have already established above the equivalence of the four measures for homogeneous populations, we claim that the fourth measure is, contrary to expectations, not equivalent to the first three for general heterogeneous structures and that, therefore, it does not provide, in general, a measure for the selective advantage of a strategy. To understand this, we attempt to reproduce the argument above with an RV weighting. First of all, the analogue of (5) holds and gives $\langle \Delta \hat{x}^{\mathrm{sel}} \rangle = \langle \Delta \hat{x}^{\mathrm{mut}} \rangle$. From there, the argument leading to (6) requires average fecundity to be the same in each of the two pure states. However, this need not be the case for RV-weighted fecundity, and if we use $\hat{\mathbf{b}}_0$ and $\hat{\mathbf{b}}_1$ to represent average fecundity in the

pure states (all B and all A) and $\hat{\mathbf{b}}$ to be the average fecundity at neutrality (which is independent of state), then the analogue of (6) is

$$\langle \Delta \hat{x}^{\text{mut}} \rangle = \delta u \Big[\underbrace{p(1-p)(\hat{\mathbf{b}}_1^{(\delta)} - \hat{\mathbf{b}}_0^{(\delta)})}_{\text{additional term}} + \hat{\mathbf{b}} \lim_{u \to 0} (\langle x \rangle - p)^{(\delta)} \Big]. \tag{12}$$

This new mutation-selection term contains an additional quantity that reflects the fact that, in a heterogeneous population, the probability of being displaced by an offspring can vary among nodes. For example, in the neutral star graph with birth-death (BD) updating described in "Example: The Star Graph" below, the hub receives new offspring at four times the rate of a leaf. Since we model mutation as occurring in new offspring, mutants will arrive at different nodes—and, therefore, with different reproductive values—at different rates. When we turn selection on, it can further perturb this effect and provide a new component of allele-frequency change, which we will refer to as the effect of selection through mutation. Therefore, the correct weak selection and low mutation measure for heterogeneous structures is

$$\langle \Delta^{(\delta)} \hat{x}^{\text{sel}} \rangle_0^{(u)} > p(1-p)(\hat{\mathbf{b}}_1^{(\delta)} - \hat{\mathbf{b}}_0^{(\delta)}), \tag{13}$$

and this is the main contribution of our article. For a homogeneous population, the additional term is always zero since the RVs are identical for all locations and, in that case, we recover the known results. In heterogeneous populations, however, the use of RV succeeds in simplifying the selection term and allows its calculation at weak selection; but, at the same time, it complicates the mutation-selection term and ultimately gives us an extra term that derives from the differential fecundity among nodes, even in the absence of selection. One important final remark is that the use of the RV transformation provides a correct measure only for weak selection. Reproductive values are calculated in the neutral population, and away from weak selection, the RV-weighted terms will generally not give the same results as their nonweighted counterparts. Therefore, away from weak selection, the correct result that holds for any population structure is the one given by (3), without any modification.

Measure 4′: Inclusive Fitness

Hamilton (1964) suggested that, under certain conditions, the neutral selective change in allele frequency can be written as a weighted sum of fitness effects, where the weights are the relatedness of the actor to the individuals whose fitness is affected by the behavior. To be more precise, take a focal actor and tabulate the fitness effects on all individuals of the change in the behavior resulting from a genetic change from B to A in the actor. Then the inclusive fitness (IF) effect of A is defined as the sum of these weighted effects $\hat{e}_i$ over all individuals i, each effect weighted by the relatedness $R_i = R_i^0$ of the actor to the affected individual (the recipient): $\widehat{W}_{\text{IF}} = \sum_i \hat{e}_i R_i^0$. We give R the superscript 0 to emphasize that it is generally understood that the relatedness coefficients, which are measures of genetic similarity, are calculated at the neutral equilibrium, it being generally infeasible to calculate them when selection is acting. Hamilton (1964) himself suggested that these neutral relatedness coefficients should give "a good approximation to the truth when selection is slow" (p. 4). Indeed, a number of subsequent studies with various modeling approaches (e.g., Charlesworth and Charnov 1981; Taylor 1990, 2009; Rousset and Billiard 2000; Taylor et al. 2007*b*; Nowak et al. 2010; Wakano et al. 2013) have shown that under certain conditions (e.g., weak selection, low mutation, additive pairwise games, additive fitness effects between actors), the inclusive fitness effect has the same sign as and is proportional to the neutral first-order one-step change in RV-weighted gene frequency:

$$\widehat{W}_{\text{IF}} \propto \langle \Delta^{(\delta)} \hat{x}^{\text{sel}} \rangle_0^{(u)}. \tag{14}$$

This is why we do not treat inclusive fitness as a separate measure but rather as a variant of the fourth measure presented above. With the use of neutral relatedness coefficients, $\widehat{W}_{\text{IF}}$ explicitly calculates the neutral RV-weighted average one-step selective change $\langle \Delta^{(\delta)} \hat{x}^{\text{sel}} \rangle_0$. It should be mentioned that Hamilton (1964) did not explicitly incorporate reproductive value in his fitness expressions, and the discussion there was (often implicitly) in the context of a homogeneous population. In his work with altruistic behavior in a haplodiploid genetic system, Hamilton (1972) emphasized the importance of using these RV weights and proposed incorporating them in the relatedness coefficients, producing what he called "life-for-life" coefficients. In Hamilton (1972) and subsequent work with genetic asymmetries and with age structure (e.g., Charlesworth and Charnov 1981; Taylor 1990), the use of reproductive value as the correct weighting for fitness was justified with the idea that it is the long-term fitness of the allele that matters, and, in fact, that was how reproductive value was (and still is) defined and regarded. But our treatment of RV above gives an alternative way to regard its role.

Since we show that $\langle \Delta^{(\delta)} \hat{x}^{\text{sel}} \rangle_0^{(u)}$ does not provide a correct (or rather a complete) measure of selective advantage for general heterogeneous populations, it is implied that the same holds true for inclusive fitness. Thus, instead of the familiar measure of selective advantage

$$\widehat{W}_{\text{IF}} > 0, \tag{15}$$

the correct measure contains the extra term reflecting the effect of mutation via selection:

$$\widehat{W}_{\text{IF}} > (\hat{\mathbf{b}}_1^{(\delta)} - \hat{\mathbf{b}}_0^{(\delta)}). \tag{16}$$

These results are summarized in table 1, and the notations used are recorded in table 2.

Example: The Star Graph

We start by analyzing one of the simplest examples of heterogeneous structure, the *N*-star graph, which will be sufficient to exemplify the discrepancies between the four measures. The *N*-star graph has one central node called the "hub" and $N - 1$ peripheral nodes called the "leaves." We identify two classes of nodes denoted by H (hub) and L (leaf). The hub has degree $N - 1$ since it is connected to all the leaves, while the leaves have degree 1 since they are only connected to the hub (fig. 1). We adopt the convention common in evolutionary graph theory (Lieberman et al. 2005; Ohtsuki et al. 2006; Taylor et al. 2007*a*) that the connection edges indicate interactions as well as offspring dispersal, and we consider all edges to be weighted equally. The payoff of an individual is the total accumulated from all its interactions (the more neighbors a node has, the more interactions it will have; fig. 1*B*). In this section, we focus on the Moran process with BD updating and fecundity payoffs—an individual is picked to reproduce proportional to its effective payoff, and the offspring replaces one of the parent's neighbors at random. Below, we study strategy dominance on the star graph using the fixation probability and the inclusive fitness approaches. For simplicity, we show only the calculation for a three-star graph. This case captures all the aspects of heterogeneity that we are concerned with in this work.

Fixation Probabilities

On a heterogeneous network, different locations can have different probabilities for the first mutant to arise at that location. On a three-star graph with BD updating and an all-B population, the probability that the first A mutant arises at the hub is the probability that one of the neighbors of the hub is chosen to reproduce times the probability that it reproduces into the hub. The probability that the first A mutant arises at a given leaf is similarly the probability that the hub reproduces times the probability that its offspring replaces that leaf. We find

$$\begin{aligned} \Pr(\text{first mutant at } H) &= 2\frac{1+\delta d}{3+4\delta d}, \\ \Pr(\text{first mutant at } L) &= \frac{1}{2}\frac{1+2\delta d}{3+4\delta d}. \end{aligned} \tag{17}$$

Note that not only is this not a uniform distribution but that, moreover, the appearance probabilities depend on the payoff d of the resident population. This has interesting implications. For example, the probability that the mutant arises at the hub is a decreasing function of the payoff d. This means that the higher the payoff of the resident population, the less likely it will be for the first mutant to arise on the hub. Hence, for heterogeneous populations, the mutant appearance distribution is not independent of payoffs (as is the case for homogeneous populations), and this will influence strategy selection. This, together with the fact that, even at neutrality, a mutant is four times more likely to arise at the hub than at the leaves, will be the crucial element that allows selection to operate via

Table 1: Table of equivalences between the five measures

Measure	Low mutation, any selection, any structure	Low mutation, weak selection, homogeneous	Low mutation, weak selection, any structure
1. Fixation probabilities	$\rho_A > \rho_B$	$\rho_A^{(\delta)} > \rho_B^{(\delta)}$	$\rho_A^{(\delta)} > \rho_B^{(\delta)}$
2. Stationary gene frequency	$\lim_{u\to 0}\langle x\rangle > p$	$\lim_{u\to 0}(\langle x\rangle - p)^{(\delta)} > 0$	$\lim_{u\to 0}(\langle x\rangle - p)^{(\delta)} > 0$
3. Stationary gene frequency change	$\langle \Delta x^{\text{sel}}\rangle^{(u)} > 0$	$\langle \Delta x^{\text{sel}}\rangle^{(\delta u)} > 0$	$\langle \Delta x^{\text{sel}}\rangle^{(\delta u)} > 0$
4. Neutral stationary frequency change	...	$\langle \Delta^{(\delta)} x^{\text{sel}}\rangle_0^{(u)} > 0$	Incomplete: $\langle \Delta^{(\delta)} \hat{x}^{\text{sel}}\rangle_0^{(u)} > 0$ Complete: $\langle \Delta^{(\delta)} \hat{x}^{\text{sel}}\rangle_0^{(u)} > p(1-p)(\hat{\mathbf{b}}_1^{(\delta)} - \hat{\mathbf{b}}_0^{(\delta)})$
4′. Inclusive fitness	...	$W_{\text{IF}} > 0$	Incomplete: $\widehat{W}_{\text{IF}} > 0$ Complete: $\widehat{W}_{\text{IF}} > (\hat{\mathbf{b}}_1^{(\delta)} - \hat{\mathbf{b}}_0^{(\delta)})$

Note: Rows correspond to the five measures discussed, where we consider the last two to be identical (under the assumptions specified in the text). Columns give the equivalences of the four measures under the conditions specified at the top of each column. We use "Incomplete" to refer to the existing measures employed in the literature that we show to be not generally equivalent to the other measures. We use "Complete" to refer to the correct measures, as derived in this article. We do not specify what measures 4 and 4′ look like when selection need not be weak; definitions of the inclusive fitness effect for any intensity of selection have been proposed, but they are not standard in the literature and they do not represent the focus of our article.

mutation even when mutation is weak and leads to the discrepancies between the measures of selection.

Next, we calculate the total fixation probability. The fixation probabilities from any location on the star have been calculated already (Tarnita et al. 2009*b*; app. A) and, together with the mutant appearance distribution above, in the limit of weak selection yield

$$\rho_A = \Pr(\text{first mutant at } H)\rho_{A,H} + 2\Pr(\text{first mutant at } L)\rho_{A,L} \tag{18}$$
$$= \frac{2}{15}\left[2 + \delta\left(\frac{4}{5}a + 3b - \frac{7}{3}c - \frac{8}{15}d\right)\right].$$

The probability ρ_B that a B mutant fixates into a population of otherwise A players can be obtained by symmetry from the above. Then the condition for strategy A to be dominant over strategy B is given by $\rho_A^{(\delta)} > \rho_B^{(\delta)}$, which is equivalent to $a + 4b > 4c + d$. For the simplified prisoner's dilemma, with benefit B and cost C, the condition that cooperation is favored over defection becomes

$$-3B - 5C > 0. \tag{19}$$

Since $B > C > 0$, it is implied that cooperation is never favored on a three-star graph with BD updating.

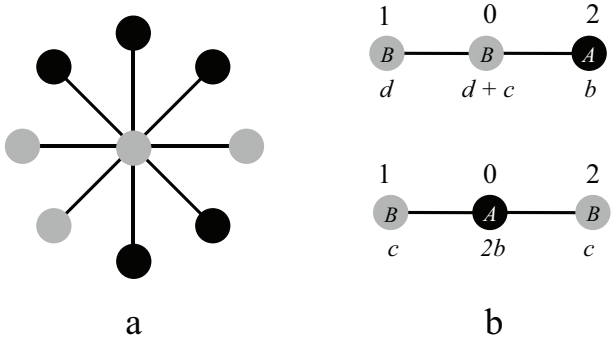

Figure 1: *a*, Nine-star graph; black = strategy A; gray = strategy B. *b*, Examples of different states of a three-star graph showing the payoffs ϵ_i. The payoff of an individual is the total accumulated from all its interactions. Upper state is *BBA*; lower state is *BAB*. The numbers represent the locations (indexes) of the individuals.

Inclusive Fitness

When the population has different classes of actors, we take a focal individual in each class and calculate the inclusive fitness effect of its behavior, weighting fitness effects

Table 2: Notations used in the text

Notation	Definition
N	Fixed population size
δ	Selection coefficient; $\delta = 0$ is neutrality; $\delta \to 0$ is weak selection
u	Mutation rate
p	Probability that mutation results in A offspring; $1 - p$ is probability that mutation results in B offspring
S	State of the population
x_i	Genotype of individual i; $x_i = 1$ or 0 (if strategy of node i is A or B)
x	Gene frequency, or average genotypic value of the nodes, in a state
v_i	Reproductive value (RV) of node i
ϵ_i	Payoff of individual on node i in a state
f_i	$1 + \delta\epsilon_i$; effective payoff of individual on node i in a state
ϵ	Total payoff of the population in a state
F	Total effective payoff of the population in a state
ρ_A	Fixation probability of an A mutant in a population of B
$\rho_{A,i}$	Fixation probability of an A mutant placed on node i in a population of B
b_i, $(\hat{b}_i)$	(RV-weighted) birthrate of individual on node i in a state
d_i, $(\hat{d}_i)$	(RV-weighted) death rate of individual on node i in state S
p_{ij}	Probability that an offspring of individual i colonizes location j
$\mathbf{b}_S$ $(\hat{\mathbf{b}}_S)$	Average (RV-weighted) birthrate in state S
$\pi_S(u, \delta)$	Probability that the system is in state S at equilibrium; depends on game and mutation rate
$\Delta_S x^{\text{sel}}$, $(\Delta_S \hat{x}^{\text{sel}})$	Change in (RV-weighted) gene frequency due to selection in state S
$\Delta_S x^{\text{mut}}$, $(\Delta_S \hat{x}^{\text{mut}})$	Change in (RV-weighted) gene frequency due to mutation in state S
$\Delta_S x^{\text{tot}}$, $(\Delta_S \hat{x}^{\text{tot}})$	Total change in (RV-weighted) gene frequency in state S
$h^{(y)}$	Coefficient of the linear term of the Taylor expansion of function h with respect to y
$\langle\cdot\rangle$	Average over the stationary distribution
$\langle\cdot\rangle_0$	Neutral stationary average (i.e., in the absence of selection, $\delta = 0$)
G_{ij}	Probability i and j are identical by descent
W_{IF}	Inclusive fitness effect
$\widehat{W}_{\text{IF}}$	RV-modified average inclusive fitness effect

by reproductive value (Taylor 2009). Thus, a breeder's death is weighted by the RV of the node it leaves, and a new offspring is weighted by the RV of the node it arrives at. Having calculated these class-specific effects, one for each actor class, the overall inclusive fitness effect is their weighted sum, with each class weighted by the class size (Taylor and Frank 1996). Thus, in the case of the star, having calculated $\widehat{W}_{\mathrm{IF}}(H)$ and $\widehat{W}_{\mathrm{IF}}(L)$ for the hub actor (interacting with both leaves) and for a leaf actor (interacting with the hub), the overall inclusive-fitness effect is the weighted sum of these class-specific effects, with class sizes as weights, $\widehat{W}_{\mathrm{IF}} = \widehat{W}_{\mathrm{IF}}(H) + 2\widehat{W}_{\mathrm{IF}}(L)$.

It is worth noting that each fitness effect is the sum of terms with a simple structure (Taylor 2009). A typical fitness component of a focal actor i will change the probability that an offspring from node j will displace the breeder on node k. The corresponding term will have the form $ev_k(G_{ij} - G_{ik})$, where e is the fitness effect (the change in probability), v_k is the RV of node k, and G_{ij} is the probability that nodes i and j are identical by descent at neutrality. We will use these identity-by-descent probabilities as relatedness coefficients (see apps. C and D). Following this structure, the inclusive fitness effect of the hub 0 (giving benefit B to each leaf at cost C) and of the left leaf actor (giving benefit B to the hub at cost C) are expressed as

$$\widehat{W}_{\mathrm{IF}}(H) = 2\frac{1}{9}\left[Bv_0(G_{01} - G_{00}) - \frac{C}{2}v_1(G_{00} - G_{01}) - \frac{C}{2}v_2(G_{00} - G_{02})\right],$$

$$\widehat{W}_{\mathrm{IF}}(L) = \frac{1}{9}\left[\frac{B}{2}v_1(G_{10} - G_{11}) + \frac{B}{2}v_2(G_{10} - G_{12}) - Cv_0(G_{11} - G_{10})\right]. \tag{20}$$

The 2 in the calculation of $\widehat{W}_{\mathrm{IF}}(H)$ comes from the fact that the hub has two symmetric interactions, one with each leaf. Notice that we have partitioned any fecundity effect on the hub into two components, as the resulting offspring can go to either leaf. These can readily be combined into one, but we keep them separate here for clarity. Note also the factor 1/9 (which is $(1/N)^2$) in each term. This is a normalizing factor that most inclusive fitness analyses would omit, but it is required if we want our expression to provide the right measure of gene frequency change. First, the fecundity increments B and C are effective only when the recipient is chosen to reproduce and the probability of that is $1/N$. Second, the increments B and C, if effective, count extra A alleles. To translate this into extra A frequency, a second division by N is required.

In appendix D, the G coefficients are calculated for low mutation to be $G_{01} = G_{02} = 1 - 7u/3$, $G_{12} = 1 - 10u/3$, and, of course, $G_{00} = G_{11} = 1$. From the definition of RV, we conclude that $v_2 = v_1 = 2v_0$, so we pick $v_0 = 1$ and $v_1 = v_2 = 2$. Using all these values together, we get the overall inclusive fitness effects to be $\widehat{W}_{\mathrm{IF}} = \widehat{W}_{\mathrm{IF}}(H) + 2\widehat{W}_{\mathrm{IF}}(L) = 2(-11B - 21C)/27$, and thus the condition that the inclusive fitness effect is positive becomes

$$-11B - 21C > 0. \tag{21}$$

Understanding the Star Example

The inclusive fitness condition (21) is not identical to the fixation probabilities condition (19), which shows that, for heterogeneous populations, the last measure of selective advantage is not necessarily equivalent to the first three. What does the inclusive fitness effect calculate then, and where does the discrepancy come from? Take a fixed population state—for example, state *BBA* (upper state in fig. 1*B*). In this state, the total reproductive value of A individuals (which we will denote by A-RV) is v_L (the RV of a leaf node). Now let selection (in the absence of mutation) act, and then go one time step and calculate the new average A-RV. The possibilities for the new state after one step are *BBA*, *BBB*, and *BAA*, and the new values of A-RV are v_L, 0, and $v_H + v_L$, respectively. The average one-step increase in A-RV is the average of the differences 0, $-v_L$, and v_H, respectively, weighted by the three transition probabilities from the old state to each of the possible new states. That average gives us $\Delta^{(\delta)}\hat{x}^{\mathrm{sel}}$ for the state *BAA*. Now, make the same calculation for every starting state—there are six of these, four mixed and two pure, but as we are ignoring mutation, only the mixed states will yield a nonzero change—and take the average of these six (in fact four) changes in A-RV using the neutral mutation-selection equilibrium state frequencies as weights. What we get is what we have called $\langle\Delta^{(\delta)}\hat{x}^{\mathrm{sel}}\rangle_0^{(u)}$, and it is that which $\widehat{W}_{\mathrm{IF}}$ measures. For the star, this is shown in appendix C.

But there are two effects of selection, even when weak, that $\widehat{W}_{\mathrm{IF}}$ does not measure. The first is the effect of selection on the distributions of states (what we have called $\langle\Delta^{(0)}\hat{x}^{\mathrm{sel}}\rangle_\delta$), but as we have seen, the use of RV-weighted fitness effects makes this zero. The second is the effect of selection on the distribution among nodes of a new mutant, and in a heterogeneous population, this can be different for new A and B alleles, leading to a different RV for new A and B mutants. This missing term is given by the right-hand side of (13); in the case of the three star, it is $2(B - C)/27$. Subtracting this from $\widehat{W}_{\mathrm{IF}}$ gives the equivalent result to that obtained via the comparison of the fixation probabilities in (19).

Thus, for the three-star graph, the inclusive fitness effect provides the correct one-step selective change in RV-weighted allele frequency if the effects of selection acting through mutation are ignored, but it predicts neither the relative fixation probabilities nor the relationship between

the neutral and mutation-selection equilibria and, therefore, does not tell the whole story.

Finally, it is worth noting that the additional mutation term is not always problematic. In this section, we analyzed the star with a Moran process with BD updating and fecundity payoffs; for a Moran process with death-birth (DB) updating with fecundity payoffs, the additional mutation term will always be zero, and we will in this case obtain the known condition $\widehat{W}_{\mathrm{IF}} > 0$. The argument for this is simple. Under DB updating with uniform survival, the breeder replaced in a reproductive event is always chosen at random, so the average RV of the new offspring is independent of the effects of selection. Thus, $\hat{\mathbf{b}}_1^{(\delta)} = \hat{\mathbf{b}}_0^{(\delta)}$. However, one can only know whether the additional mutation term is problematic after performing the calculation; therefore, one always has to start the analysis with the generally correct formulation in (16).

Discussion

In this article, we explore several different measures of strategy success in an evolutionary process on static, finite graphs—comparison of fixation probabilities, long-term average gene frequency, long-term average change in gene frequency, and a weak selection variant of the latter that is closely connected to the inclusive fitness effect via the Price (1970, 1972) equation. While these conditions are broadly equivalent for homogeneous populations, we showed that the same is not always true for heterogeneous structures. In our analysis, the notion of reproductive value gave us a simple and powerful modification of quantities used to describe homogeneous populations that permitted calculations to be made. The idea that, in a heterogeneous population, measures of fitness and gene frequency change need to include an RV weighting goes back at least to Fisher (1930) working with age-structured populations, and it was perhaps first formally treated in this context by Charlesworth and Charnov (1981). That same RV notion is used by us here for a different purpose, and that is to handle the effect of selection on the distribution of alleles in the limit of weak selection; general treatments of its use for this purpose go back to Taylor (1990) in infinite populations and Leturque and Rousset (2002) in finite population models.

This generalized approach showed that while the first three measures remain equivalent for any population structure, the modified neutral selective change in gene frequency and, implicitly, the inclusive fitness effect are not sufficient to provide an equivalent measure to the first three, as is the case in homogeneous (transitive) structures; an extra term that encompasses effects of selection that are playing out through mutation is necessary for completeness. The main result of our article provides the following equivalences:

$$\begin{aligned} \langle x\rangle > p \overset{u\to 0}{\Leftrightarrow} \ \rho_A > \rho_B \overset{u\to 0}{\Leftrightarrow} \langle \Delta x^{\mathrm{sel}}\rangle > 0 \overset{u,\delta\to 0}{\Leftrightarrow} \\ \langle \Delta^{(\delta)}\hat{x}^{\mathrm{sel}}\rangle_0^{(u)} > p(1-p)(\hat{\mathbf{b}}_1^{(\delta)} - \hat{\mathbf{b}}_0^{(\delta)}) \qquad (22) \\ \Leftrightarrow \widehat{W}_{\mathrm{IF}} > \hat{\mathbf{b}}_1^{(\delta)} - \hat{\mathbf{b}}_0^{(\delta)}. \end{aligned}$$

The first measure is the most general one, holding for any mutation strength and selection intensity and for any population structure. The first two equivalences hold for low mutation. The remaining equivalences hold in the limit of weak selection; the last of these requires additional assumptions that allow for the formulation of the inclusive fitness effect. The inclusive fitness effect can be rigorously formulated for additive two-person games in static population structures (for weak selection and low mutation), and we have made reference to considerable recent work (most of which is ongoing) that has focused on extending its formulation to more complex structures. Indeed, our result here also represents a step forward in generalizing the concept of inclusive fitness to static heterogeneous populations.

Having said that, the last equivalence is certainly of a different kind than the first three. Inclusive fitness is an accounting method based on a powerful heuristic concept that, right from its initial formulation (Hamilton 1964), has been continuously refined and upgraded to apply (and deliver its insights) to an increasing set of behavioral models. But in each new case and each new modeling approach, it was never clear whether or to what extent the inclusive fitness formulation would produce the desired gene frequency/fixation outcome and whether it would offer a computationally feasible approach. For example, we now understand the modifications of the IF formulation that will allow it to apply to synergistic games (Queller 1985; Ohtsuki 2010; Taylor and Maciejewski 2012; Taylor 2013), and in this article, we have shown how the RV-weighted IF effect relates in heterogeneous populations to other standard measures of the selective advantage of an allele. We were indeed surprised to discover that, while inclusive fitness still measures the selective change in gene frequency experienced in one behavioral time step, as Hamilton (1964) intended it to do, it does not, in general, give us the standard gene frequency/fixation conditions that determine the success of the behavior under study; instead, it requires the calculation of an additional term accounting for the effect of selection through mutation.

Here, in order to highlight some of the differences between the four measures, we had to choose a simplified and somewhat abstract, but nevertheless general, modeling framework that covers a vast number of theoretical studies in finite populations from the last decade. As with any mod-

eling framework, however, changes in the assumptions are likely to lead to different results. For example, we assume a fixed population size where the birthrates and death rates of individuals depend on the payoffs but not explicitly on the location on the graph and where, moreover, the average birthrate is the same whenever all individuals are of only one type; in addition, we assume that mutations occur only at birth (i.e., only in offspring). Extending this to allow for varying population size, location-dependent fitness, different average birthrates in the pure states, and mutations occurring in adults is likely to yield different results. The point that we are trying to make is not that this particular modeling framework is the most biologically appropriate or even the most appropriate theoretically to reveal all possible discrepancies, but that, depending on the modeling framework, discrepancies can and do exist between these four very general and very generally employed measures. The important conclusion of our article is, therefore, that a rigorous and general formulation of these various measures was necessary to allow us to see that they are not, in general, equivalent and, furthermore, to approach the questions of how and when they can be equivalent. Within our framework, we find that when rigorously defined and appropriately modified, the different measures provide equivalent answers but different angles from which one can view the effects of population structure on the selection of behavioral strategies, and this can be of great value.

Finally, the results presented here are derived for populations of finite size. However, the literature abounds with two types of models—infinite population models and finite population models, with the latter having garnered interest much more recently. One significant difference is that in the first type, the effects of selection on allele frequency are generally of a larger order of magnitude than mutation effects, while in the second type, these are of the same order of magnitude, essentially because the proportion of time that the population is in a mixed state is of order u. Classical inclusive fitness models did not generally take mutation into account, as they worked implicitly with what we are calling infinite population models, where the effects of mutation were of a smaller magnitude and thus could be ignored. Of course, one still needs a source of genetic novelty, and in the infinite population model, this is generally provided by distant migration.

A question that then arises is whether there could be an effect of selection through migration in the case of infinite populations that is analogous to the effect of selection through mutation identified in this article in the case of finite populations. Suppose we have an island model with rare migration and heterogeneous islands. Then the frequency of mixed-state islands will be of the same order of magnitude as the migration rate, and selective effects will therefore also be of the same order of magnitude as the migration rate. Then almost all migrants will emanate from an all-A or an all-B island, and the selective regimes might be different on these two types of island. But will that difference result in a difference in the RV of the nodes colonized by the A migrants and the B migrants? This would require immigrants to carry some "memory" of the node they came from on their native island and for that to condition the choice of node they colonized. In terms of the star graph, for example, an immigrant born at a hub node might be able to choose to colonize a leaf node. A full analysis of the analog infinite population model is yet to be performed, and together the results for finite and infinite heterogeneous population models will provide the strongest connection to the biological world. In this sense, these finite population results are a first step.

Acknowledgments

We would like to thank B. Allen, W. Maciejewski, and R. Pringle for useful discussions and A. Grafen, F. Weissing, and an anonymous reviewer for significant commentary on an earlier version of the manuscript. C.E.T. acknowledges support from the Templeton Foundation (Foundational Questions in Evolutionary Biology grant RFP-12-14). P.D.T. is supported by the Natural Sciences and Engineering Research Council of Canada.

Literature Cited

Abbot, P., J. Abe, J. Alcock, S. Alizon, J. A. C. Alpedrinha, M. Andersson, J.-B. Andre, et al. 2011. Inclusive fitness theory and eusociality. Nature 471. doi:10.1038/nature09831.

Allen, B., and C. E. Tarnita. 2014. Measures of success in a class of evolutionary models with fixed population size and structure. Journal of Mathematical Biology 68:109–143.

Antal, T., H. Ohtsuki, J. Wakeley, P. D. Taylor, and M. A. Nowak. 2009. Evolutionary game dynamics in phenotype space. Proceedings of the National Academy of Sciences of the USA 106:8597–8601.

Charlesworth, D., and E. L. Charnov. 1981. Kin selection in age-structured populations. Journal of Theoretical Biology 88:103–119.

Crow, J. F., and M. Kimura. 1970. An introduction to population genetics theory. Blackburn, Caldwell, NJ.

Doebeli, M. 2010. Inclusive fitness is just bookkeeping. Nature 467. doi:10.1038/467661b.

Ewens, W. J. 1969. Population genetics. Methuen's Monographs on Applied Probability and Statistics. Methuen, London.

Ferriere, R., and R. E. Michod. 2011. Inclusive fitness in evolution. Nature 471. doi:10.1038/nature09834.

Fisher, R. A. 1930. The genetical theory of natural selection. Clarendon, Oxford.

Frean, M., P. B. Rainey, and A. Traulsen. 2013. The effect of population

structure on the rate of evolution. Proceedings of the Royal Society B: Biological Sciences 280:20130211. doi:10.1098/rspb.2013.0211.

Grafen, A. 2009. Formalizing Darwinism and inclusive fitness theory. Philosophical Transactions of the Royal Society B: Biological Sciences 364:3135–3141.

Hamilton, W. D. 1964. The genetical evolution of social behaviour. I, II. Journal of Theoretical Biology 7:1–16.

———. 1972. Altruism and related phenomena, mainly in social insects. Annual Review of Ecology and Systematics 3:192–232.

Hauert, C., and M. Doebeli. 2004. Spatial structure often inhibits the evolution of cooperation in the snowdrift game. Nature 428: 643–646.

Imhof, L. A., and M. A. Nowak. 2006. Evolutionary game dynamics in a Wright-Fisher process. Journal of Mathematical Biology 52: 667–681.

Karlin, S., and H. M. Taylor. 1975. A first course in stochastic processes. 2nd ed. Academic Press, London.

Killingback, T., and M. Doebeli. 1996. Spatial evolutionary games: hawks and doves revisited. Proceedings of the Royal Society B: Biological Sciences 263:1135–1144.

Kimura, M. 1962. On the probability of fixation of mutant genes in a population. Genetics 47:713–719.

Leturque, H., and F. Rousset. 2002. Dispersal, kin competition, and the ideal free distribution in a spatially heterogeneous population. Theoretical Population Biology 62:169–180.

Lieberman, E., C. Hauert, and M. A. Nowak. 2005. Evolutionary dynamics on graphs. Nature 433:312–316.

Maynard Smith, J. 1982. Evolution and the theory of games. Cambridge University Press, Cambridge.

Moran, P. A. P. 1962. The statistical processes of evolutionary theory. Clarendon, Oxford.

Nowak, M. A., C. E. Tarnita, and E. O. Wilson. 2010. The evolution of eusociality. Nature 466:1057–1062.

Ohtsuki, H. 2010. Evolutionary games in Wright's island model: kin selection meets evolutionary game theory. Evolution 64:3344–3353.

Ohtsuki, H., C. Hauert, E. Lieberman, and M. A. Nowak. 2006. A simple rule for the evolution of cooperation on graphs and social networks. Nature 441:502–505.

Price, G. R. 1970. Selection and covariance. Nature 227:520–521.

———. 1972. Extension of covariance selection mathematics. Annals of Human Genetics 35:485–490.

Queller, D. C. 1985. Kinship, reciprocity and synergism in the evolution of social behaviour. Nature 318:366–367.

Rousset, F., and S. Billiard. 2000. A theoretical basis for measures of kin selection in subdivided populations: finite populations and localized dispersal. Journal of Evolutionary Biology 13:814–826.

Tarnita, C. E., T. Antal, H. Ohtsuki, F. Fu, and M. A. Nowak. 2009*a*. Evolutionary dynamics in set structured populations. Proceedings of the National Academy of Sciences of the USA 106:8601–8605.

———. 2009*b*. Strategy selection in structured populations. Journal of Theoretical Biology 259:570–581.

Tarnita, C. E., N. Wage, and M. A. Nowak. 2011. Multiple strategies in structured populations. Proceedings of the National Academy of Sciences of the USA 108:2334–2337.

Taylor, P. D. 1990. Allele-frequency change in a class-structured population. American Naturalist 135:95–106.

———. 2009. Decompositions of Price's formula in an inhomogeneous population structure. Journal of Evolutionary Biology 22: 201–213.

———. 2013. Inclusive and personal fitness in synergistic evolutionary games on graphs. Journal of Theoretical Biology 325:76–82.

Taylor, P. D., T. Day, and G. Wild. 2007*a*. Evolution of cooperation in a finite homogeneous graph. Nature 447:469–472.

———. 2007*b*. From inclusive fitness to fixation probability in homogeneous structured populations. Journal of Theoretical Biology 249:101–110.

Taylor, P. D., and S. Frank. 1996. How to make a kin selection argument. Journal of Theoretical Biology 180:27–37.

Taylor, P. D., and A. Grafen. 2010. Relatedness with different interaction configurations. Journal of Theoretical Biology 262:391–397.

Taylor, P. D., and W. Maciejewski. 2012. An inclusive fitness analysis of synergistic interactions in structured populations. Proceedings of the Royal Society B: Biological Sciences 279:4596–4603.

van Veelen, M., J. Garcia, M. W. Sabelis, and M. Egas. 2010. Call for a return to rigour in models. Nature 467:661.

Wakano, J., and L. Lehmann. 2012. Evolutionary and convergence stability for continuous phenotypes in finite populations derived from two-allele models. Journal of Theoretical Biology 310:206–215.

Wakano, J. Y., H. Ohtsuki, and Y. Kobayashi. 2013. A mathematical description of the inclusive fitness theory. Theoretical Population Biology 84:46–55.

West, S., and A. Gardner. 2013. Adaptation and inclusive fitness. Current Biology 23:R577–R584.

Wright, S. 1943. Isolation by distance. Genetics 28:114–138.

———. 1969. Evolution and the genetics of populations. Vol. II. The theory of gene frequencies. University of Chicago Press, Chicago.

Associate Editor: Franz J. Weissing
Editor: Susan Kalisz

 DOI: 10.1086/677924

Appendix A from C. E. Tarnita and P. D. Taylor, "Measures of Relative Fitness of Social Behaviors in Finite Structured Population Models"

(Am. Nat., vol. 184, no. 4, p. 477)

Equivalence of First Three Measures and Discussion of Weak Selection

In this appendix, we perform a general analysis on a heterogeneous network to relate the selective long-term equilibrium frequency $\langle x\rangle$ to the long-term equilibrium of the one-step frequency change for any intensity of selection. Then, taking the weak selection limit of the latter, we will make the connection with the neutral average change. The first part of the argument below follows closely but simplifies and improves the proof in Nowak et al. (2010), appendix A. At first we allow nonweak mutation rates and take the low mutation limit later. The frequency of A will increase when an A individual gives birth accurately or when a B individual gives birth to an A mutant, and it will decrease when an A individual dies. Thus, keeping in mind that mutation is asymmetric, we can write the total one-step change in gene frequency due to both selection and mutation as

$$\Delta_S x^{\text{tot}} = \frac{1}{N}\Big\{[1-u(1-p)]\sum_i x_i b_i + up\sum_i (1-x_i)b_i - \sum_i x_i d_i\Big\} = \Delta_S x^{\text{sel}} - \Delta_S x^{\text{mut}}, \tag{A1}$$

where we have denoted $\Delta_S x^{\text{sel}} = (1/N)\sum_i x_i(b_i - d_i)$ to be the one-step change in gene frequency due to selection and $\Delta_S x^{\text{mut}} = (u/N)\sum_i b_i(x_i - p)$ to be the one-step negative change in gene frequency due to the interaction between mutation and selection in a state S. At the stationary equilibrium, the total change is zero, and hence,

$$\langle \Delta_S x^{\text{sel}}\rangle = \langle \Delta_S x^{\text{mut}}\rangle. \tag{A2}$$

This holds for any mutation rate and any intensity of selection and simply says that in a process with both mutation and selection, on average, these two forces must balance each other out. Now let us explore what happens to the mutation term above in the limit of low mutation. In the mixed states, the stationary probabilities will be of order u since they must approach zero, as u approaches zero. Thus, since the mutation term is already linear in u, the calculation of its long-term average will only involve the two pure states. Hence, the first-order term in u of this long-term average can be written as $u[\mathbf{b}_1(1-p)\pi_1(0,\delta) + \mathbf{b}_0(-p)\pi_0(0,\delta)]$. Here, $\mathbf{b_S} = (1/N)\sum_i b_{i,S}$ is the average birthrate (average fecundity) in state S. For processes for which in the pure states the average birthrate is identical (i.e., $\mathbf{b}_1 = \mathbf{b}_0 = \mathbf{b}$), the above can be simplified further. This assumption is in fact very weak, and many processes of interest satisfy it—for example, the Moran process with birth-death or death-birth updating and the Wright-Fisher process. Under this assumption, and using the fact that for low mutation $\pi_1 + \pi_0 = 1$ and $\langle x\rangle = \pi_1(0,\delta)$, the above mutation term becomes equal to $u\mathbf{b}[\lim_{u\to 0}\langle x\rangle - p]$. Then, using (A2), we conclude that in the limit of low mutation, for processes for which the total birthrate is constant in the pure states,

$$\langle \Delta_S x^{\text{sel}}\rangle^{(u)} > 0 \Leftrightarrow \lim_{u\to 0}\langle x\rangle > p \Leftrightarrow \rho_A > \rho_B, \tag{A3}$$

where $h^{(y)}$ represents the coefficient of the linear term of h with respect to y. This condition asserts the equivalence of the first three measures for any type of structure and any intensity of selection. As we have pointed out, none of these three conditions can be easily calculated, and that is what turns our attention to a weak-selection condition.

Low Mutation and Weak Selection

In this section, we will use Taylor expansions with respect to two variables, u and δ. Since our functions are continuous and differentiable with respect to either one of the variables when the other is kept fixed, we can take the Taylor expansions sequentially and interchangeably. The first-order Taylor expansion with respect to δ of the average change due to selection is

$$\langle \Delta x^{\text{sel}}\rangle = \delta\left[\sum_S (\Delta_S^{(\delta)} x^{\text{sel}})\pi_S(u,0) + \sum_S (\Delta_S x^{\text{sel}})|_{\delta=0}\pi_S^{(\delta)}(u,0)\right], \tag{A4}$$

where we used the fact that the constant term of $\langle \Delta x^{\text{sel}} \rangle$ is 0 because at neutrality there is no change due to selection.[1] The first term on the right-hand side is the kind of term that we are able to calculate because it contains only the stationary distribution at neutrality ($\delta = 0$). For a homogeneous population, at neutrality, the births and deaths at every node are equal (because nodes are effectively identical). Hence, $(\Delta_S x^{\text{sel}})\big|_{\delta = 0}$ is zero in every state, and therefore the second term on the right-hand side of the above is zero. If we further consider the first-order Taylor expansion with respect to u of the remaining nonzero term (the first term on the right-hand side of [A4]), we find that

$$\langle \Delta x^{\text{sel}} \rangle = \delta u \langle \Delta^{(\delta)} x^{\text{sel}} \rangle_0^{(u)}, \tag{A5}$$

where $\langle \cdot \rangle_0$ denotes the stationary average in the absence of selection ($\delta = 0$) and $h^{(u)}$ represents the coefficient of the linear term of the Taylor expansion of the function h with respect to u. Thus, for homogeneous populations, the weak selection and low mutation limit of the stationary average change in gene frequency is simply (and luckily) the first-order term with respect to u of the neutral stationary average of the first-order term with respect to δ of the one-step gene frequency change. So in the limit of weak selection in homogeneous populations, we recover the known result

$$\langle x \rangle > p \overset{u \to 0}{\Leftrightarrow} \rho_A > \rho_B \overset{u \to 0}{\Leftrightarrow} \langle \Delta_S x^{\text{sel}} \rangle^{(u)} > 0 \overset{u, \delta \to 0}{\Leftrightarrow} \langle \Delta_S^{(\delta)} x^{\text{sel}} \rangle_0^{(u)} > 0 \Leftrightarrow W_{\text{IF}} > 0. \tag{A6}$$

The conditions under which the last equivalence holds have previously been shown (Rousset and Billiard 2000; Nowak et al. 2010; Wakano et al. 2013). For a heterogeneous population, however, the births and deaths at every node do not necessarily balance each other out, even at neutrality. Hence, $(\Delta_S x^{\text{sel}})\big|_{\delta = 0}$ is not necessarily zero, and therefore the second term on the right-hand side of (A4) can be nonzero. That term, however, still contains the nonneutral stationary probabilities that prevented the calculation of the change due to selection in the previous section, a problem that we hoped to solve by taking the weak selection limit. Thus, because of heterogeneity, this approach remains hard to calculate. The effective problem is that selection, in affecting behavior, also changes fitness; this, in turn, affects genotypic distributions, and in heterogeneous populations, this makes a contribution to gene frequency change. This problem, in various settings, has been known for some time and the resolution, developed in infinite population models, has been to make use of reproductive value (Taylor 1990; Leturque and Rousset 2002).

[1] Here we make a note that when we refer to the coefficient of the linear term in y of a function h, we use the notation $h^{(y)}$. However, when we want to write the actual expansion of the function h to first order in y, we will write $h = h(0) + yh^{(y)}$. So in the text, when we write $h = yh^{(y)}$, it simply means that $h(0) = 0$. In other words, we distinguish between the coefficient of the linear term and the actual linearization of the function.

 DOI: 10.1086/677924

Appendix B from C. E. Tarnita and P. D. Taylor, "Measures of Relative Fitness of Social Behaviors in Finite Structured Population Models"

(Am. Nat., vol. 184, no. 4, p. 477)

RV-Weighted Gene Frequency

In the main text, we claimed that, at the stationary equilibrium, selection balances mutation:

$$\langle \Delta_S \hat{x}^{\text{sel}} \rangle = \langle \Delta_S \hat{x}^{\text{mut}} \rangle, \tag{B1}$$

where $\Delta_S \hat{x}^{\text{sel}} = (1/N)\sum_i x_i(\hat{b}_i - \hat{d}_i)$ and $\Delta_S \hat{x}^{\text{mut}} = (u/N)\sum_i \hat{b}_i(x_i - p)$. Below, we look at both these averages in the limit of low mutation and weak selection. In what follows, we will take Taylor expansions with respect to both u and δ; the order in which we do this does not matter since, under our assumptions, all functions we are dealing with are continuous and differentiable with respect to each (and both) of these variables when the other variable is fixed. So the order we choose is simply the one that is most convenient for our analysis.

Step 1. $\langle \Delta \hat{x}^{\text{sel}} \rangle$ Has the Form $u\delta K$ to First Order in u and δ

Using the same arguments, as we did before for the change in unweighted gene frequency, we can write the linear term of the average first as a function of δ:

$$\langle \Delta \hat{x}^{\text{sel}} \rangle = \delta \left[\sum_S \underbrace{\frac{\partial}{\partial \delta} (\Delta_S \hat{x}^{\text{sel}})|_{\delta=0}}_{\Delta_S^{(\delta)} \hat{x}^{\text{sel}}} \pi_S(u,0) + \sum_S (\Delta_S \hat{x}^{\text{sel}})|_{\delta=0} \underbrace{\frac{\partial}{\partial \delta} \pi_S(u,0)|_{\delta=0}}_{\pi_S^{(\delta)}(u,0)} \right]. \tag{B2}$$

From the definition of RV in the main text, it follows that at neutrality, $\hat{b}_i = \hat{d}_i$, which yields that the change due to selection is zero in all states: $\Delta \hat{x}^{\text{sel}}|_{\delta=0} = 0$. Hence, the second term in the above is zero, and we obtain that in the limit of weak selection:

$$\langle \Delta \hat{x}^{\text{sel}} \rangle = \delta \sum_S \Delta_S^{(\delta)} \hat{x}^{\text{sel}} \pi_S(u,0) = \delta \langle \Delta^{(\delta)} \hat{x}^{\text{sel}} \rangle_0. \tag{B3}$$

Next, we write the linear term of the above as a function of u

$$= \delta \left(\sum_S (\Delta_S^{(\delta)} \hat{x}^{\text{sel}}) \pi_S(0,0) + u \sum_S (\Delta_S^{(\delta)} \hat{x}^{\text{sel}}) \pi_S^{(u)}(u,0) \right). \tag{B4}$$

The first term in the above sum is zero because in the absence of mutation, $\pi_S(0,\delta) = 0$ for all mixed states S, and in the pure states, the change due to selection is zero since in the absence of mutation, selection alone cannot push the system out of pure states. Thus,

$$\langle \Delta \hat{x}^{\text{sel}} \rangle = \delta u \sum_S (\Delta_S^{(\delta)} \hat{x}^{\text{sel}}) \pi_S^{(u)}(u,0) = \delta u \langle \Delta^{(\delta)} \hat{x}^{\text{sel}} \rangle_0^{(u)}. \tag{B5}$$

Here, the notation $\langle \cdot \rangle_0$ means that the average is taken over the neutral stationary distribution ($\delta = 0$). The notation $h^{(x)}$ signifies the coefficient of the linear term of the Taylor expansion of function h with respect to x.

Step 2. $(u/N)\langle \sum_i \hat{b}_i(x_i - p) \rangle$ Also Has the Form $u\delta K$ to First Order in u and δ

For the second term, we take the limit of low mutation and ignore all but the pure states (as explained in the main text). Then the linear term in u is

$$\frac{u}{N} \left[(1-p)\pi_1(0,\delta) \sum_i \hat{b}_{i,1} - p\pi_0(0,\delta) \sum_i \hat{b}_{i,0} \right]. \tag{B6}$$

Here $\hat{b}_{i,S}$ is $\hat{b}_i$ in state S. Next, we write the linear term in δ, keeping in mind that at neutrality, $\pi_0(0,0) = 1-p$, $\pi_1(0,0) = p$, and fecundity is independent of state. We find that the above has the form

$$\frac{\delta u}{N}\left\{p(1-p)\frac{\partial}{\partial\delta}\left(\sum_i \hat{b}_{i,1} - \sum_i \hat{b}_{i,0}\right)\Bigg|_{\delta=0} + \sum_i \hat{b}_i^{(0)}\frac{\partial}{\partial\delta}[(1-p)\pi_1(0,\delta) - p\pi_0(0,\delta)]|_{\delta=0}\right\}. \tag{B7}$$

This concludes the proof of step 2. Combining steps 1 and 2 and noting that the difference $(1-p)\pi_1(0,\delta) - p\pi_0(0,\delta)$ is proportional to $\rho_A - \rho_B$ (as discussed in the main text), we conclude that

$$\rho_A^{(\delta)} - \rho_B^{(\delta)} = \frac{\partial}{\partial\delta}(\rho_A - \rho_B)|_{\delta=0} \propto \langle \Delta^{(\delta)}\hat{x}^{\mathrm{sel}}\rangle_0^{(u)} - p(1-p)(\hat{\mathbf{b}}_1^{(\delta)} - \hat{\mathbf{b}}_0^{(\delta)}). \tag{B8}$$

Here, $\hat{\mathbf{b}}_{\mathbf{S}}^{(\delta)} = (\partial/\partial\delta)[(1/N)\sum_i \hat{b}_{i,\mathbf{S}}]|_{\delta=0}$ is the first-order term in δ of the average birthrate in state S. In (B8), the first term is simply the change due to selection in the RV-weighted allele frequencies; the second term is the part due to mutation and birth that is not necessarily zero when the population is heterogeneous.

 DOI: 10.1086/677924

Appendix C from C. E. Tarnita and P. D. Taylor, "Measures of Relative Fitness of Social Behaviors in Finite Structured Population Models"

(Am. Nat., vol. 184, no. 4, p. 477)

Calculation for the Star Using RV-Weighted Change in Gene Frequency

Here, we perform a calculation for the star using the average change in gene frequency due to selection, together with the RV transformation. But first we show how the quantities of interest look for any heterogeneous structure with birth-death updating. First, we relate the effective payoff to the birthrate and death rate for a process with birth-date (BD) updating and fecundity payoffs. With this, birth happens first with a probability proportional to breeder fecundity; the offspring of the reproducing individual then replaces one of the neighbors of the parent at random. So we can write $b_i = f_i/F$ and $d_i = \sum_{j=1}^{N} b_j p_{ji}$, where p_{ji} is the probability that an offspring of individual j will colonize location i. Since an offspring will replace any one of the parent's neighbors at random, then $p_{ji} = 0$ if i is not a neighbor of j, and otherwise, $p_{ji} = 1/n_j$, where n_j is the number of neighbors of individual j.

For BD updating, the RV-weighted birthrate and death rate become $\hat{b}_i = b_i \sum_j p_{ij} v_j = (f_i/F) \sum_j p_{ij} v_j$ and $\hat{d}_i = d_i v_i = v_i \sum_j b_j p_{ji}$. Then, in the limit of weak selection, the coefficient of the linear term with respect to δ of the difference between the birthrate and death rate is

$$\hat{b}_i^{(\delta)} - \hat{d}_i^{(\delta)} = \frac{1}{N}\left(\epsilon_i \sum_j p_{ij} v_j - v_i \sum_j \epsilon_j p_{ji}\right), \tag{C1}$$

where ϵ_i is the accumulated payoff of individual i.

Now we go back to the explicit calculation for the three star. We denote the hub by 0 and the left and right leaves by 1 and 2, respectively (see fig. 1*B* in the main text). Using the above definition for the star, it is easy to check that, at neutrality, $b_0 = 1/3 \neq 2/3 = d_0$, $b_1 = 1/3 \neq 1/6 = d_1$, $b_2 = 1/3 \neq 1/6 = d_2$, so the star is the type of network for which, even at neutrality, the death rate and birthrate of individuals are not necessarily the same. This means that we are in the case where, to be able to calculate in the limit of weak selection, we need to use reproductive values. It is easy to see that for the three star, $v_1 = v_2 = 2v_0$. This, together with the normalizing condition $\sum_i v_i = 1$, determines the v_i; however, for convenience, we will use the unnormalized values $v_1 = v_2 = 2$ and $v_0 = 1$. Let f_0, f_1, and f_2 be the effective payoffs of individuals at the three locations, and let $F = f_0 + f_1 + f_2$ be the total effective payoff in the population. Using the above births and deaths, we write

$$\begin{array}{lll} \hat{b}_0 = \dfrac{f_0}{F}\left(\dfrac{1}{2}v_1 + \dfrac{1}{2}v_2\right) & \hat{b}_1 = \dfrac{f_1}{F}v_0 & \hat{b}_2 = \dfrac{f_2}{F}v_0 \\ \hat{d}_0 = \dfrac{f_1 + f_2}{F}v_0 & \hat{d}_1 = \dfrac{f_0}{F}\dfrac{1}{2}v_1 & \hat{d}_2 = \dfrac{f_0}{F}\dfrac{1}{2}v_2 \end{array}. \tag{C2}$$

It is now easy to see that at neutrality, for each node, the weighted birthrate is equal to the weighted death rate. Let us now write the effective payoffs of the three individuals so we can proceed to calculate the condition for strategy A to be favored over strategy B. Our calculation works for any game, given by a general matrix. However, due to the Σ result in Tarnita et al. (2009*b*; see also Taylor and Maciejewski 2012), we know that in the limit of weak selection, it suffices to study one-parameter games such as the simplified Prisoner's Dilemma with benefit B and cost C. Then the effective payoffs of the individuals at the three locations are given by $f_0 = 1 + \delta\epsilon_0 = 1 + \delta[-2Cx_0 + B(x_1 + x_2)]$, $f_1 = 1 + \delta\epsilon_1 = 1 + \delta(-Cx_1 + Bx_0)$, and $f_2 = 1 + \delta\epsilon_2 = 1 + \delta(-Cx_2 + Bx_0)$. We have now all the elements necessary to calculate the condition given by (B8).

First, we calculate the simpler of the two terms on the right in (B8), the mutation term (the second term), using the expressions for the birthrates in (C2):

$$\begin{aligned}\hat{\mathbf{b}}_1^{(\delta)} &= \frac{1}{3}(\hat{b}_{H,1}^{(\delta)} + \hat{b}_{L,1}^{(\delta)} + \hat{b}_{R,1}^{(\delta)}) = \frac{1}{27}(2B - 2C)(v_1 - v_0) = \frac{1}{27}v_0(2B - 2C),\\ \hat{\mathbf{b}}_0^{(\delta)} &= \frac{1}{3}(\hat{b}_{H,0}^{(\delta)} + \hat{b}_{L,0}^{(\delta)} + \hat{b}_{R,0}^{(\delta)}) = 0.\end{aligned} \tag{C3}$$

These represent the first-order term in δ of the average birthrate in the state where all players are cooperators, respectively, defectors. Then the mutation term is

$$p(1-p)\left(\hat{\mathbf{b}}_1^{(\delta)} - \hat{\mathbf{b}}_0^{(\delta)}\right) = \frac{1}{27}p(1-p)v_0(2B - 2C). \tag{C4}$$

Next, we calculate the selection term (the first term on the right) in (B8), and using (C2) and the expression for the effective payoffs into (C1), we find that it is equal to

$$\begin{aligned}\frac{1}{3}\langle x_0(\hat{b}_H^{(\delta)} - \hat{d}_H^{(\delta)}) + x_1(\hat{b}_L^{(\delta)} - \hat{d}_L^{(\delta)}) + x_2(\hat{b}_R^{(\delta)} - \hat{d}_R^{(\delta)})\rangle_0^{(u)} = \frac{1}{9}\{v_0\big[-2B\langle x_0^2\rangle_0^{(u)} - 2C\langle x_1^2\rangle_0^{(u)} + (2B + 2C)\langle x_1 x_0\rangle_0^{(u)}\big]\\ + v_1\big[-B\langle x_1^2\rangle_0^{(u)} - 2C\langle x_0^2\rangle_0^{(u)} + (2B + 2C)\langle x_1 x_0\rangle_0^{(u)} - B\langle x_1 x_2\rangle_0^{(u)}\big]\}.\end{aligned} \tag{C5}$$

To simplify the above expression and show only the effects of different types of nodes (instead of the effects of each node), we used the symmetry of the left and right leaves at neutrality, which yields $\langle x_1 x_0\rangle_0 = \langle x_2 x_0\rangle_0$, together with the fact that $v_1 = v_2$. Now we need to interpret and calculate the neutral averages: $\langle x_1 x_0\rangle_0$ is the probability that a leaf and the hub are both cooperators; $\langle x_1 x_2\rangle_0$ is the probability that the two leaves are both cooperators. Such quantities have been related to the probability G that two individuals are identical by descent at neutrality, as $\langle x_i x_j\rangle_0 = pG_{ij} + p^2(1 - G_{ij}) = p(1-p)G_{ij} + p^2$. The probability that individuals are identical by descent can be calculated using recursions as shown in appendix D. Let G_{10} and G_{12} be the probabilities that a leaf and the hub, respectively, the two leaves, are identical by descent. We find (app. D) that in the limit of low mutation, $G_{10} = 1 - 7u/3$ and $G_{12} = 1 - 10u/3$. Using these in (C5), we find the selection term to be

$$\langle \Delta^{(\delta)}\hat{x}^{\text{sel}}\rangle_0^{(u)} = \frac{1}{27}p(1-p)v_0(-42C - 22B). \tag{C6}$$

Combining the mutation and selection terms into (B8), we obtain the condition for cooperators to be favored over defectors to be

$$\frac{8}{27}p(1-p)(-3B - 5C) > 0, \tag{C7}$$

which is the same $-3B - 5C > 0$ condition found using the fixation probabilities approach.

 DOI: 10.1086/677924

Appendix D from C. E. Tarnita and P. D. Taylor, "Measures of Relative Fitness of Social Behaviors in Finite Structured Population Models"

(Am. Nat., vol. 184, no. 4, p. 477)

Identity by Descent

Let G_{10}^t and G_{12}^t be the probabilities that a leaf and the hub and the two leaves, respectively, are identical by descent at time t. Then, in order to write what happens at time $t + 1$, we first point out that the only possible events from time t to time $t + 1$ are as follows. (*a*) The hub reproduces into the left leaf with probability 1/6, and then the only possibility for the left leaf and the hub to be identical by descent is if the hub reproduces correctly. (*b*) The hub reproduces into the right leaf with probability 1/6; in this case, the hub and the left leaf are identical by descent only if they were identical by descent in the previous step. (*c*) The left leaf reproduces into the hub with probability 1/3; then, as in case (*a*), the left leaf and the hub can be identical by descent only if the leaf reproduced correctly. Or (*d*), the right leaf reproduces into the hub with probability 1/3; then the left leaf and the hub are identical by descent only if the left leaf and the right leaf were identical by descent in the previous time step and the right leaf reproduced correctly. Similarly, one can analyze what happens in these four cases to the probability that the two leaves are identical by descent. We can then write

$$\begin{aligned} G_{10}^{t+1} &= \frac{1}{6}(1-u) + \frac{1}{6}G_{10}^t + \frac{1}{3}(1-u) + \frac{1}{3}G_{12}^t, \\ G_{12}^{t+1} &= \frac{1}{6}G_{10}^t(1-u) + \frac{1}{6}G_{10}^t(1-u) + \frac{1}{3}G_{12}^t + \frac{1}{3}G_{12}^t. \end{aligned} \tag{D1}$$

Since we are concerned with the long-term limit, we have $G_{10}^{t+1} = G_{10}^t = G_{10}$ and $G_{12}^{t+1} = G_{12}^t = G_{12}$, which allows us to solve the above system of equations and furthermore take the limit of low mutation of the results to find that

$$\begin{aligned} G_{10} &= 1 - \frac{7}{3}u, \\ G_{12} &= 1 - \frac{10}{3}u. \end{aligned} \tag{D2}$$