Behavioral Responses in Structured Populations Pave the Way to Group Optimality*

Erol Akçay[1,†] and Jeremy Van Cleve[2]

1. National Institute for Mathematical and Biological Synthesis (NIMBioS), University of Tennessee, Knoxville, Tennessee, 37996; 2. Santa Fe Institute, Santa Fe, New Mexico 87501

Submitted July 18, 2011; Accepted October 25, 2011; Electronically published December 21, 2011

Online enhancement: appendix.

abstract: An unresolved controversy regarding social behaviors is exemplified when natural selection might lead to behaviors that maximize fitness at the social-group level but are costly at the individual level. Except for the special case of groups of clones, we do not have a general understanding of how and when group-optimal behaviors evolve, especially when the behaviors in question are flexible. To address this question, we develop a general model that integrates behavioral plasticity in social interactions with the action of natural selection in structured populations. We find that group-optimal behaviors can evolve, even without clonal groups, if individuals exhibit appropriate behavioral responses to each other's actions. The evolution of such behavioral responses, in turn, is predicated on the nature of the proximate behavioral mechanisms. We model a particular class of proximate mechanisms, prosocial preferences, and find that such preferences evolve to sustain maximum group benefit under certain levels of relatedness and certain ecological conditions. Thus, our model demonstrates the fundamental interplay between behavioral responses and relatedness in determining the course of social evolution. We also highlight the crucial role of proximate mechanisms such as prosocial preferences in the evolution of behavioral responses and in facilitating evolutionary transitions in individuality.

Keywords: goal-oriented behavior, Price equation, two-tiered model, kin selection, multilevel selection, community reciprocity.

Introduction

Among the many debates on the evolution of social behaviors, perhaps none is older or more controversial than the one surrounding the role and importance of group-level selection in the evolutionary process (Wynne-Edwards 1962; Williams 1966; Hamilton 1975; West et al. 2007; Wilson and Wilson 2007; Leigh 2010). There is now at least one important consensus that multilevel-selection processes (which are motivated historically from models of group-level selection) are mathematically equivalent to kin-selection processes (which historically derive from analyses of individual-level selection; Hamilton 1975; Queller 1992). Even though few now dispute that some selection at the group level (between-group selection, in current terminology) occurs (West et al. 2007; Wilson and Wilson 2007; Gardner and Grafen 2009), other issues remain controversial. Among these is the original question in the debate, namely, how likely natural selection is to lead to behaviors that maximize fitness at the group level. It is well known that when selection within a group is completely abolished, between-group selection can lead to maximization of group fitness when appropriate genetic variation exists and other evolutionary forces are weak. Within-group selection disappears when the expected fitness of all individuals within a group is equalized either as a result of clonality within groups or through mechanisms such as policing (in social insects) or fair meiosis (genomes of sexual organisms; Leigh 1977; Alexander and Borgia 1978; Frank 2003). Some taxa, such as eusocial aphids, do live in clonal groups, but this is generally accepted to be a rare condition in nature outside of some prokaryotic lineages (Levin et al. 1999). Repression of competition is much more widespread and is found in social insects, vertebrates, and human societies (Frank 2003 and references therein).

Another route for diminishing within-group differences in fitness is behavioral coordination through individuals' responses to their groupmates' actions. Virtually all organisms exhibit social behaviors that are flexible or conditional on the behaviors of others, from bacteria (e.g., quorum sensing; Miller and Bassler 2001) to insects (e.g., reproductive strategies depending on social context; West-Eberhard 1987), birds (e.g., responses of parents to each other; Wright and Cuthill 1989), primates (e.g., reciprocal

* The two authors contributed equally to the design and analysis of the model and to the writing of this article.

† Corresponding author. Present address: Department of Ecology and Evolutionary Biology, Princeton University, Guyot Hall, Princeton, New Jersey 08544; e-mail: eakcay@princeton.edu.

Am. Nat. 2012. Vol. 179, pp. 257–269. 0003-0147/2012/17902-53193$15.00.
DOI: 10.1086/663691

altruism; Brosnan and de Waal 2002), and obviously, humans. Not surprisingly, a large theoretical literature focuses on the evolutionary consequences of particular kinds of flexible (or conditional) behavior (e.g., Axelrod and Hamilton 1981; McNamara et al. 1999; Lehmann and Keller 2006; Akçay et al. 2009; Boyd et al. 2010). However, the general question of whether and when flexible behaviors can allow group-optimal outcomes to be evolutionarily stable has not been answered. Another related body of work that deals with "indirect genetic effects" (IGEs; Griffing 1967, 1981*a*, 1981*b*; Moore et al. 1997; Wolf et al. 1999; Bijma et al. 2007; Bijma and Wade 2008; McGlothlin et al. 2010) uses quantitative genetics to measure how the response of an individual's phenotype to the phenotypes of others affects selection pressures in social interactions. However, these models do not consider how the IGEs themselves evolve (Bijma and Wade 2008; McGlothlin et al. 2010). An element missing from both strands of theory is the role of proximate mechanisms: behavioral responses (or IGEs) are produced by proximate mechanisms, which can either constrain or facilitate the evolution of different types of responses. Despite their significance, proximate mechanisms are only infrequently integrated into models of social evolution in structured populations. Thus, we currently lack a general framework for understanding both the selective effect and the evolution of behavioral responses in structured populations based on the proximate mechanisms that generate such responses.

In this article, we demonstrate how such a framework can be built from the Price equation (Price 1970, 1972). Our framework shows that behavioral responses and relatedness due to population structure play exactly symmetric roles in determining the direction of selection on a social trait. When using the multilevel-selection perspective, our framework also shows how cooperative behaviors are always selected against at the within-group level while being selected for at the between-group level, regardless of behavioral responsiveness or relatedness. We then apply our framework to the problem of when natural selection can lead to maximization of fitness at the group level, or group optimality. Our focus on group optimality is motivated in part by the fact that group optimality is intimately related to evolutionary transitions in individuality (ETIs), that is, groups becoming new, higher-level individuals (Maynard Smith and Szathmáry 1995; Michod 2005, 2006). We show that behavioral responses significantly increase the conditions under which group optimality is evolutionarily stable. Using a specific proximate mechanism for the production of behavioral responses, we also show how behavioral responses that lead to group optimality might evolve. The proximate mechanism we employ is the social-preferences, or motivations, model of Akçay et al. (2009), which we extend to the case of *N*-player public-goods games. Group-optimal outcomes can evolve in this model through the evolution of how much individuals value the public good versus their private costs.

Model

Integrating Behavioral Responses and Kin Selection

We use a central framework of social-evolution theory, the Price equation (Price 1970, 1972), to partition the effect of selection on a heritable trait into components due to the effect of the trait of the focal individual and those due to effects on the traits of others in the population (the so-called kin-selection partition; Price 1970; Queller 1992; Bijma and Wade 2008). The change in the population-average breeding value (additive genetic component) due to the effect of selection can be written as

$$\Delta\overline{G} \propto \mathrm{Var}\,(G_i)[\beta_{F_i p_i} + (N-1)r\beta_{F_i p_j}], \tag{1}$$

where G_i is the breeding value of the focal individual, N is the size of the social-interaction group, r is scaled genetic relatedness, and $\beta_{F_i p_i}$ and $\beta_{F_i p_j}$ are partial-regression coefficients of the focal individual's fertility (F_i) on its own phenotype (p_i) and on its neighbors' phenotypes (p_j), respectively.

In many structured populations, fitness is a function of both fertility and demographic effects such as local competition for limited resources with related individuals (Taylor 1992; Queller 1994). The scaled-relatedness coefficient r is defined to account for these demographic effects (Frank 1998) and can be calculated under a variety of life-history and demographic scenarios (Lehmann and Rousset 2010). As result of this accounting, $\beta_{F_i p_i}$ and $\beta_{F_i p_j}$ measure the effects of the phenotypes on fertility only. For example, in appendix A2, available online, we provide a derivation of the scaled-relatedness coefficient r for a population with overlapping generations and an island model of migration between D different demes, each of which has n individuals (Wright 1943). In that case, we find that

$$r = \frac{R - R^*}{1 - R^*},$$

where R is the unscaled-relatedness coefficient defined in terms of probabilities of identity by descent and

$$R^* = \left[(1-m)^2 + \frac{m^2}{D-1}\right]\left(\frac{1}{n} + \frac{n-1}{n}R\right);$$

R in this example is Wright's F_{ST} coefficient (Wright 1949; Rousset 2004). For an island model with overlapping generations, $r \approx s/n$, where s is the probability that adults survive from one generation to the next and D and n are large and m is small (see app. A2). Finally, we make the

additional assumption that r is determined only by the demographic parameters of the population and not by individual phenotypes, which implies that r is a constant. This assumption is justified by our assumption of weak selection below (Rousset 2004).

Both $\beta_{\mathrm{F}_i\mathrm{P}_i}$ and $\beta_{\mathrm{F}_i\mathrm{P}_j}$ depend on how individuals respond to each others' actions. To model these behavioral responses, we generalize the modeling approach of a recent series of models (McNamara et al. 1999; Taylor and Day 2004; André and Day 2007; Akçay et al. 2009; Akçay and Roughgarden 2011) that consider the evolution of behaviors in a two-tiered dynamic. Consider a social interaction with N individuals, where each individual i carries out an action a_i, which is a real, positive number. For example, a_i can denote how much individual i contributes to offspring provisioning in a cooperatively breeding group. Instead of treating the actions a_i as being directly determined by the genetic makeup of the individuals, we assume that the actions that each individual carries out are determined through a behavioral dynamic that operates at the timescale of the social interaction. We assume that this behavioral dynamic is fast and quickly settles on some equilibrium action a_i^* for each individual i. This "behavioral equilibrium" (denoted by $\mathbf{a}^* = (a_1^*, a_2^*, \ldots a_N^*)$) is a function of the decision-making mechanism of the individuals, that is, the proximate mechanism of behavior. For example, McNamara et al. (1999) model the proximate mechanism as a linear-response rule, where the action of an individual is a linear function of its opponent's action, while Akçay et al. (2009) model the proximate mechanism as a motivation to maximize a behavioral objective or social preference function.

We assume that the proximate mechanism is modulated by a genetically encoded trait that we term the "motivational trait," since it affects the decision making of the individuals. The motivational trait is the phenotype p, whose evolution we track in equation (1). In our model, the motivational trait p is a proxy for neurophysiological or endocrinological traits, such as the expression pattern of neuropeptide receptors (e.g., oxytocin or vasopressin in mammals) in key brain regions or the functional responses of specific neural circuits to external stimuli. Recent research has shown that such physiological traits in humans and other animals affect many types of social behaviors, such as pair bonding (Young et al. 2011), parental care (Donaldson and Young 2008), trust in economic games (Baumgartner et al. 2008), and the tendency to aggregate in groups (Goodson and Wang 2006).

By tracking the evolution of the motivational trait p instead of the equilibrium action a^*, we can study the evolution of the proximate mechanism that generates behaviors instead of the behaviors alone. In principle, different proximate mechanisms (e.g., linear-response rules or behavioral objectives) might produce the same outcome in a specific behavioral context (e.g., parental care). However, in different behavioral contexts (e.g., food sharing), the outcomes produced by these proximate mechanisms can be divergent. Studying the proximate mechanisms underlying behaviors would allow us to understand the evolution of behavioral correlations across different social contexts.

Figure 1 illustrates how, by determining properties of neurophysiology, the motivational trait p modulates individual decision-making processes and thus affects the actions that individuals choose at the behavioral equilibrium. The behavioral equilibrium is a function of the combination of the social partners' motivational traits, or mathematically, $\mathbf{a}^*(p_1, \ldots, p_N)$. The action of each individual at the behavioral equilibrium (e.g., the level of helping), in turn, determines the payoff that each individual gets (e.g., the amount of resources an individual has as a result of each partner's level of helping). We denote as

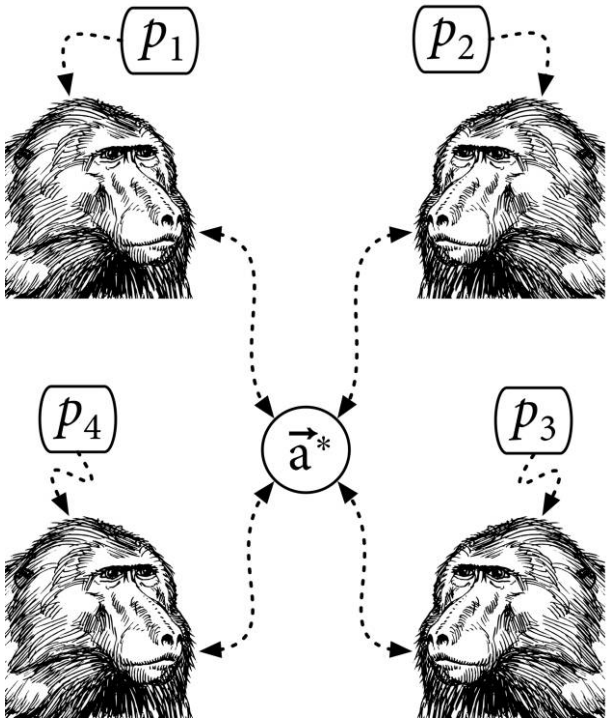


Figure 1: Relationship between motivational traits p_i and the actions individuals choose at the behavioral equilibrium. For each individual i in the social interaction, the motivational trait p_i determines some aspects of that individual's neurophysiology, which in turn affect the decision-making process of that individual. The behavioral equilibrium actions, $\mathbf{a}^* = (a_1^*, \ldots, a_N^*)$, are the outcome of the behavioral dynamic defined by the decision-making mechanisms of all the individuals involved in the interaction. Illustration from Pearson Scott Foresman, released into public domain at Wikimedia Commons.

$u_i(a_1, a_2, \ldots, a_N)$ individual i's payoff from the social interaction. In general, a focal individual's fertility F_i will be some function of the payoff u_i, possibly integrated over some time period and possibly including stochastic effects. To keep the analysis tractable, we assume that the fertility of individual i is proportional to the payoff u_i evaluated at the behavioral equilibrium $\mathbf{a}^*$,

$$F_i(p_1, \ldots, p_N) \equiv u_i(a_1^*(p_1, \ldots, p_N), \ldots, a_N^*(p_1, \ldots, p_N)).$$

This expression signifies that ultimately, the fertility of a focal individual is a function of both its own and its partners' motivational traits.

Assuming that the effect of the motivational trait on fertility is weak (i.e., that selection is weak; Taylor and Frank 1996), the partial-regression coefficient of the focal individual i's fertility on its own motivational trait is the direct effect of the change in the focal individual's motivational trait plus the indirect effect arising from other individuals in the group responding to the focal individual,

$$\beta_{F_i p_i} = \frac{\partial F_i}{\partial p_i} = \frac{\partial a_i^*}{\partial p_i}\left(\frac{\partial u_i}{\partial a_i} + \sum_{j \neq i} \rho_{ij} \frac{\partial u_i}{\partial a_j}\right)_{\mathbf{a}=\mathbf{a}^*}, \tag{2}$$

where the term $\rho_{ij} = (\partial a_j^*/\partial p_i)/(\partial a_i^*/\partial p_i) = \partial a_j^*/\partial a_i^*$ quantifies how individual j's equilibrium action changes in response to individual i's equilibrium action. Thus, ρ_{ij} describes the relative behavioral response of j to i; we term it the "response coefficient" of j to i (Akçay et al. 2009). The mechanistic details of a specific behavioral model, such as coordinated punishment (Boyd et al. 2010) or prosocial preferences (Levine 1998; Falk and Fischbacher 2006; Akçay et al. 2009), determine the value of the response coefficient. In general, ρ_{ij} may capture the equilibrium effect of many different behavioral models, including social norms and systems of rewards and punishments (Levin 2009). We emphasize that even though we treat the response coefficient ρ_{ij} as an index here, it is actually determined by how individuals settle on their equilibrium actions, and therefore it will evolve as the population distribution of the motivational trait evolves. We will turn to the evolution of ρ itself shortly.

For the effect on fertility of a change in a social partner's motivational trait p_j, we have

$$\beta_{F_i p_j} = \frac{\partial F_i}{\partial p_j} = \frac{\partial a_j^*}{\partial p_j}\left(\frac{\partial u_i}{\partial a_j} + \sum_{k \neq j} \rho_{jk} \frac{\partial u_i}{\partial a_k}\right)_{\mathbf{a}=\mathbf{a}^*}. \tag{3}$$

Again, this partial-regression coefficient consists of the direct effect on i's payoff of changing social partner j's action plus the effects of responses such a change elicits from all other individuals (including focal individual i).

Assuming that we are interested only in small deviations in the motivational-trait distribution from a monomorphic population (a population composed of individuals with the same trait), we can set $\rho_{ij} = \rho$ for all $j \neq i$ and $\partial a_i^*/\partial p_i = \partial a_j^*/\partial p_j$. We further define a benefit $b \equiv \partial u_i/\partial a_j$ for $i \neq j$ ("other-only" benefit; Pepper 2000) and a cost $c \equiv -\partial u_i/\partial a_i$, where the derivatives are evaluated at $\mathbf{a}^* = (a^*, \ldots, a^*)$; b and c are generalizations of the benefits and costs, respectively, in a linear public-goods game. Note that b and c are defined locally at the current phenotypic value of the population and the behavioral equilibrium it produces. As the behavioral equilibrium changes because of shifts in the population motivational trait, the values of b and c change as well.

Using the definitions of ρ, b, and c, we can rewrite the change in the population-average breeding value in equation (1) as

$$\Delta\overline{G} \propto k\big(b\rho(N-1) - c + r(N-1) \times \{b[\rho(N-2)+1] - \rho c\}\big). \tag{4}$$

By setting $\Delta\overline{G} = 0$ in equation (1), we obtain the first-order condition for a given motivational trait to be evolutionarily stable (ES) as

$$\frac{b}{c} = \frac{1 + r\rho(N-1)}{(N-1)[r + \rho + (N-2)r\rho]}. \tag{5}$$

A motivational trait that satisfies this first-order condition is a candidate evolutionarily stable strategy (ESS; Maynard Smith and Price 1973). Equation (5) is a kind of Hamilton's rule (Hamilton 1964) incorporating behavioral responses among N interacting individuals. Expressions similar to equation (5) are derived by Lehmann and Keller (2006, eq. [4]) and by McGlothlin et al. (2010, eq. [18]) using IGEs, whereas Bijma and colleagues (Bijma et al. 2007; Bijma and Wade 2008) develop a related expression (see eq. [15] in Bijma and Wade 2008) based on a different approach to partitioning the Price equation.

Importantly, equation (5) is exactly symmetric in r and ρ, meaning that behavioral responses and relatedness play mathematically analogous roles in determining evolutionary stability. This does not mean, however, that one can collapse them into a single index (e.g., an index of assortment) without loss of generality, since the two appear separately in equation (5). Hence, both behavioral responses and relatedness must be considered when determining the total selection pressure on a given social behavior.

The first-order condition (eq. [5]) is necessary but not sufficient for the candidate ESS to be the stable outcome of an evolutionary dynamic; such stability requires that certain second-order conditions hold as well. In particular, the second-order ESS condition must be satisfied to ensure that the candidate ESS is in fact a fitness maximum, as opposed to a minimum. This condition can be difficult to calculate exactly in structured populations, since it re-

quires determining how selection changes relatedness (Ajar 2003; Rousset 2004), and we refrain from calculating it here. Another second-order condition, called convergence stability (CS; Eshel and Motro 1981; Christiansen 1991), is required to ensure that a population that is near a candidate ESS evolves toward the candidate ESS through successive invasion and fixation of mutations. In appendix A4, available online, we derive a general expression for the CS condition. The CS condition depends crucially on how the response coefficient ρ changes with the evolving motivational trait, which in turn is a function of the proximate mechanism that produces the behavioral responses. In "The Evolution of Behavioral Responses through Prosocial Preferences," we model a proximate mechanism based on goal-oriented motivations in public-goods games to study the coevolution of behavioral responses and investment in a public good.

Behavioral Responses and the Levels of Selection

In deriving equation (4), we used the kin-selection partition of the Price equation; an alternative partition is the "group-selection" (or "multilevel-selection") partition that decomposes the effect of selection into within-group and between-group components (Hamilton 1975; Queller 1992; see app. A1, available online). Upon rearranging the terms in equation (4) for $\Delta\overline{G}$, we find that

$$\Delta\overline{G} \propto k\{[b(N-1)-c][1+\rho(N-1)][1+r(N-1)] - (N-1)(b+c)(1-\rho)(1-r)\}, \quad (6)$$

where the within-group component of selection is

$$-(N-1)(b+c)(1-\rho)(1-r) \quad (7)$$

and the between-group component is

$$[b(N-1)-c][1+\rho(N-1)][1+r(N-1)]. \quad (8)$$

The within-group component is always negative for $\rho < 1$ and $r < 1$ and vanishes when either is equal to 1. Hence, selection at the within-group level opposes an increase in the amount of cooperation, but its force gets weaker as ρ or r increases and disappears when either ρ or r equals 1. Within-group selection vanishes with $\rho = 1$ because this condition eliminates variation in the actions a^* and hence within-group variation in fitness. Thus, even when individuals are highly responsive to one another's actions, within-group selection opposes an increase in the amount of cooperation. The between-group selection component, on the other hand, is positive when $b/c > 1/(N-1)$ (assuming that $r > -1/(N-1)$ and $\rho > -1/(N-1)$; see appendixes A1 and A3, available online, for limits on r and ρ), which is true so long as the total benefit of the social trait outweighs the cost. This means that if cooperation is potentially beneficial, then between-group selection always favors an increase in the amount of cooperation, regardless of levels of responsiveness and relatedness. In other words, setting $\Delta\overline{G} > 0$ and using equation (6) together imply that the amount of cooperation increases if and only if between-group selection is stronger than within-group selection. A special case of this result was discovered by Wade (1980), who considered fixed altruistic behaviors ($\rho = 0$).

In contrast to the group-selection partition, the direct and indirect fertility effects ($\beta_{F_{p_i}p_i}$ and $\beta_{F_{p_i}p_j}$, respectively) in the kin-selection partition, equation (1), can each be positive or negative, depending on ρ (eq. [4]; app. A1). In the kin-selection terminology due to Hamilton (1964), which defines "altruism" as a behavior with a negative direct fitness effect and positive indirect fitness effects (Rousset 2004; Lehmann and Keller 2006), this means that whether cooperation is "altruistic" depends on both behavioral responsiveness and demographic factors that affect the scaled-relatedness coefficient r. Thus, cooperation in a public-goods game may be "mutually beneficial" (West et al. 2007) and not "altruistic" when behavioral responsiveness is high, since the direct effect on fertility (and consequently fitness) may be positive. The distinction between mutually beneficial and altruistic cooperation is conceptually important in social evolution, and a number of recent reviews attempt to clarify this issue in detail (e.g., Lehmann and Keller 2006; West et al. 2007), because it has often generated confusion (Kerr et al. 2004). In this regard, the group-selection partition may be a more natural framework for public-goods scenarios than the kin-selection partition, since the direction of between- and within-group selection does not depend on ρ or r. In fact, a public-goods scenario could be defined by payoffs that yield positive between-group selection and negative within-group selection.

Evolutionary Stability of Group-Optimal Behaviors

Exactly Group-Optimal Behaviors

During an ETI, selection within groups is eliminated and between-group selection determines whether social behaviors increase or decrease in frequency. Continued between-group selection will eventually lead to a convergence-stable outcome with social behaviors that maximize the aggregate payoff to the group (see eq. [6]). With a monomorphic population, maximizing the aggregate group payoff implies that

$$\frac{b}{c} = \frac{1}{N-1} \quad (9)$$

(see app. A6, available online). The b/c ratio that satisfies the group-optimality condition (9) will not, in general,

maximize a single individual's payoff, if one kept all other individuals' actions constant. For example, a focal individual might selfishly withhold its contribution to a public good and increase its own payoff while hurting the aggregate payoff. However, when the social behaviors are flexible, we also have to take into account how other individuals will respond to any changes in the focal individual's behavior, which is done in equation (5).

Combining conditions (5) and (9), we can see that a group-optimal outcome can be ES when

$$r + \rho - r\rho = 1. \tag{10}$$

Equation (10) is satisfied only when r or ρ or both are equal to 1. In other words, evolutionary stability of a group-optimal behavior requires either perfect relatedness, corresponding to clonal groups, or perfect responsiveness, corresponding to perfect behavioral matching of actions. Either of these conditions is sufficient; that is, group-optimal outcomes can be ES when $\rho = 1$ regardless of r, and vice versa. The first condition ($r = 1$) is well known, while a special case of the second condition has recently been discovered in the two-person continuous prisoner's dilemma game without population structure (André and Day 2007).

Almost-Group-Optimal Behaviors

The requirement that either $r = 1$ or $\rho = 1$ might still seem restrictive, but it is essentially an artifact of requiring exact group optimality. In reality, empirically distinguishing between exactly group-optimal behaviors and those that are almost, but not precisely, group optimal is likely to be difficult. This implies that the scope for what "looks" group optimal might be even broader than previously expected. To see this, we can consider an outcome $a^*(\varepsilon)$ that is approximately group optimal and induces a ratio $b/c = (1 + \varepsilon)/(N - 1)$, where $\varepsilon > 0$. As $\varepsilon \to 0$, the outcome $a^*(\varepsilon)$ approaches the exact group-optimal outcome. A b/c ratio that is lower than $(1 + \varepsilon)/(N - 1)$ implies an outcome closer to group optimality than is $a^*(\varepsilon)$. We can show that a b/c ratio smaller than $(1 + \varepsilon)/(N - 1)$ will be ES whenever

$$\rho > \frac{1 - r(1 + \varepsilon)}{1 - r + \varepsilon[1 + r(N - 2)]}. \tag{11}$$

Since the right-hand side is always less than 1 for all $N \geq 2$ and $1 > r > -1/(N - 1)$, we can find a $\rho < 1$ that will make it possible for an outcome arbitrarily close to the group-optimal outcome to be ES. Furthermore, we can see in figure 2 that the threshold ρ required to make an outcome ES is strictly decreasing with increased relatedness in the population. Figure 2 also shows that a combination of moderate r and moderate ρ is sufficient to bring the ES outcome within 10% of group optimality, suggesting that the scope for approximately group-optimal behaviors can be much wider than previously recognized. This also suggests that partial ETIs with meager amounts of within-group selection require only moderate levels of relatedness and behavioral responsiveness.

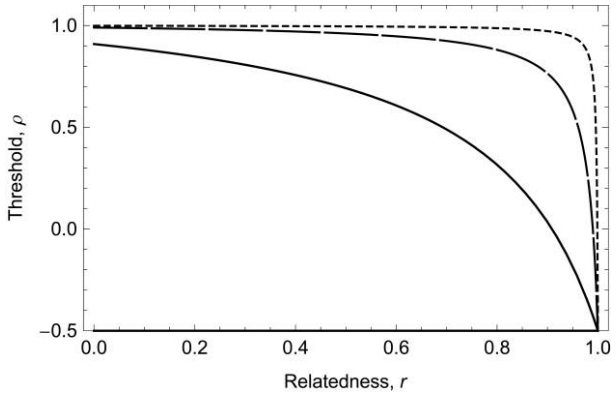


Figure 2: Threshold ρ from equation (11) required to make an outcome evolutionarily stable with a given level of divergence from the group-optimal outcome. Here, we assume that $N = 10$. The solid curve is for a b/c ratio 10% higher than the group-optimal ratio, the dashed curve is for a b/c ratio 1% higher, and the dotted curve is for a b/c ratio 0.1% higher.

The Evolution of Behavioral Responses through Prosocial Preferences

In the previous section, we discussed the effect of the response coefficient ρ on selection acting on social behaviors, but we did not ask how ρ itself evolves. Previous work shows that direct selection on ρ as an independent trait occurs only to second order in the strength of selection (André and Day 2007; Akçay et al. 2009). However, the behavioral response ρ is not independent of the social action a^*: both are produced by the proximate behavioral mechanism. Hence, as the motivational traits underlying the proximate mechanisms evolve, both a^* and ρ will coevolve with the motivational traits. Therefore, to study how the response coefficient ρ evolves, we have to specify how both the equilibrium actions and the behavioral responses are produced by the proximate mechanism of behavior. In this section, we do this by using a model for the evolution of prosocial preferences as the proximate cause of social behavior.

In particular, we concentrate on the evolution of the response coefficient $\rho = 1$, which ensures the stability of group-optimal behaviors, from lower values of ρ. Our approach in this section is somewhat inverted, relative to

conventional ESS analysis. Instead of searching for a general ES and CS outcome and determining when such an outcome is group optimal, we find the conditions under which the group-optimal outcome is ES and CS.

We analyze public-goods games where each individual has the option to invest in an action that benefits everyone in the group (including the focal individual) but incurs a personal cost. Thus, the payoff to an individual i is given by: $u_i = \mathcal{B}(a_1, \ldots, a_N) - \mathcal{C}(a_i)$, where $\mathcal{B}$ is the public benefit and $\mathcal{C}$ denotes the private cost. (Note that these functions are different from b and c defined above.) This model is a stylized description of many important social interactions, including cooperative hunting and defense or provisioning of a common brood of offspring.

Extending a recent model by Akçay et al. (2009), we model individuals' motivations for investment in the public good as deriving from innate goals. We represent these innate goals mathematically by an objective function, $x_i(a_1, a_2, \ldots, a_N)$ for individual i. The objective function fulfills a role similar to that of a utility or preference function in economics (e.g., Heifetz et al. 2007*b*) and is the proximate cause of behavior: each individual "wants" to achieve the maximum value of its objective function, given what others do. The shape of the objective function of individual x_i is, in turn, determined by the motivational trait p_i that is the target of natural selection. The behavioral equilibrium $\mathbf{a}^*$ for this maximization process satisfies

$$\left.\frac{\partial x_i}{\partial a_i}\right|_{\mathbf{a}=\mathbf{a}^*} = 0 \tag{12}$$

for all $i \in \{1, \ldots, N\}$ (see app. A3 for stability conditions). To obtain the response coefficient, we choose some $j \neq i$, differentiate $\partial x_j/\partial a_j = 0$ with respect to the motivational trait of the focal individual p_i, and solve for $\rho = (\partial a_j/\partial p_i)/(\partial a_i/\partial p_i)$ evaluated at $\mathbf{a}^*$ in a monomorphic population, to obtain

$$\rho = -\frac{\partial^2 x_j/\partial a_i \partial a_j}{(N-2)(\partial^2 x_j/\partial a_i \partial a_j) + (\partial^2 x_j/\partial a_j^2)}. \tag{13}$$

Expression (13) reduces to equation (9) of Akçay et al. (2009) for $N = 2$.

The objective function x allows a natural characterization of the variation in prosocial preferences. First, individuals might vary in how they value the public benefit relative to their private cost. Second, individuals might be more or less cost averse, depending on how much public benefit they receive, and thus vary in the direction and strength of this aversion. To account for these two effects, we assume that the objective function is determined by $\mathcal{B}$ and $\mathcal{C}$ and is parametrized by two motivational traits θ and ϕ, such that

$$x_i(a_1, \ldots, a_N) = \theta\mathcal{B} - (1 - \phi\mathcal{B})\mathcal{C}. \tag{14}$$

Individuals with higher θ place greater value on the public good and are more unconditionally prosocial. Individuals with positive ϕ discount their costs as the provision of the public good increases, similar to the behavior found in some public-goods experiments (Fischbacher et al. 2001). A pattern of conditional discounting of private costs, which results in a conditional willingness to contribute, might be termed "community reciprocity," reflecting the fact that individuals' motivations are reciprocal not to any particular individual but to the overall provision of public good. The response coefficient ρ, given by equation (13), evolves as both θ and ϕ evolve.

We first determine, for given a payoff function, the values of the phenotypic traits θ^* and ϕ^* that generate an objective function that leads to an ES group-optimal outcome. These traits must satisfy two conditions: first, the group-optimal outcome must be a proper behavioral equilibrium and satisfy equation (12) evaluated at θ^* and ϕ^*; and second, the objective functions must yield a response coefficient $\rho = 1$ at the group-optimal behavioral equilibrium. These conditions allow us to find a unique (θ^*, ϕ^*) pair that satisfies the first-order condition for evolutionary stability at the group-optimal behavioral equilibrium.

To illustrate, we set $N = 3$ and use the following public-benefit and private-cost functions:

$$\mathcal{B}(a_1, a_2, a_3) = (1 - \nu)\sum_{i=1}^{3}\sqrt{a_i} + \nu\sqrt{a_1 a_2 a_3},$$
$$\mathcal{C}(a) = a^2. \tag{15}$$

The parameter ν $(-1 < \nu < 1)$ measures how different individuals' contributions to the public benefit interact with each other. When $\nu > 0$, investments by different individuals interact synergistically with each other: the more one individual invests, the more valuable another individual's contribution becomes. The opposite holds when $\nu < 0$ and individuals' investments become substitutes for each other. The parameter ν therefore captures an important aspect of the ecology of the interaction. For example, cooperative hunting might require simultaneous, synergistic efforts from multiple individuals because of the requirement to close all escape routes for the prey. This would make for a positive ν. Conversely, when provisioning offspring, the increased effort by one individual would increase the food coming into the nest and would decrease value of additional food from another individual. This would create a negative ν. In economics, these two possibilities are termed complementary and substitutable inputs, respectively; hence, we call ν the complementarity parameter. Note that

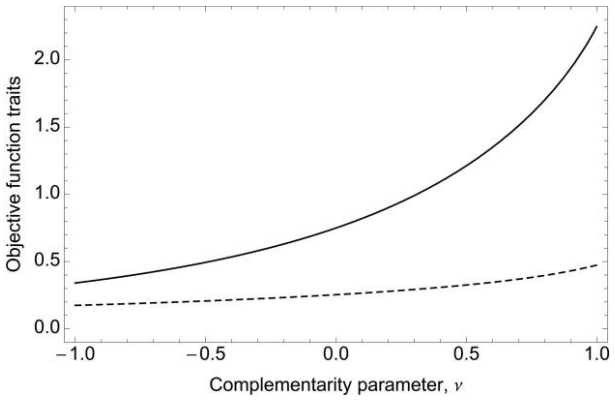


Figure 3: Values of θ^* (solid curve) and ϕ^* (dashed curve) that result in an evolutionarily stable group-optimal outcome, plotted as functions of the parameter ν that modulates whether different individuals' contributions to the public benefit are substitutes ($\nu < 0$) or complements ($\nu > 0$).

$\mathcal{B}$ is always positive (for positive a), regardless of the sign of ν.

Figure 3 shows θ^* and ϕ^* as functions of the parameter ν; the more complementary the inputs from different individuals (higher ν), the higher the valuation of the public benefit relative to the private cost (θ^*). Furthermore, over the entire range of ν, the trait for the interaction between the public benefit and private cost, ϕ^*, is positive and increases with ν. This means that an objective function satisfying $\rho = 1$ at the group-optimal behavioral equilibrium exhibits the community-reciprocity property. As the public benefit becomes more complementary, the level of community reciprocity at the group-optimal outcome also increases.

How do prosocial preferences with community reciprocity maintain a response coefficient of $\rho = 1$ proximately? Suppose a focal individual i has a positive ϕ trait. When its partner j increases investment a_j in the public good, this raises the value of $\mathcal{B}$. From equation (14), one can see that the focal individual i will then discount its own cost more. This, in turn, would make the focal individual more willing to invest and would give rise to a positive response coefficient (the reverse argument applies when the partner invests less). Perfectly matching responses, $\rho = 1$, simply constitute an extreme case of such positive responses.

However, as mentioned above, the first-order condition for evolutionary stability does not guarantee that a population will in fact evolve to the candidate ESS when it is initially away from it; this requires the convergence-stability (CS) condition to be satisfied as well (Eshel and Motro 1981; Christiansen 1991). Evaluating the CS condition for the group-optimal θ^* and ϕ^* traits found above, we find that whether θ^* and ϕ^* are CS depends on the relatedness coefficient r, which must be above a threshold value to satisfy the CS condition (fig. 4). One interpretation of this result is that even though $\rho = 1$ guarantees that the first-order ES condition is satisfied for a group-optimal outcome, the evolution of ρ to this value crucially depends on the relatedness being high enough. Thus, relatedness still plays an important role in determining whether natural selection can drive a population to reach these trait values. However, one can also observe from figure 4 that the value of r required to render the group-optimal traits CS is much less than 1 and that it decreases as the complementarity parameter ν increases. Hence, the evolution of behaviors that maximize group benefit is possible for a range of organisms much wider than only those that form clonal groups. Furthermore, ecological scenarios with synergistic effects (such as cooperative hunting) permit the evolution of group-optimal outcomes under a wider range of demographic conditions than do those with substitute effects (such as offspring provisioning).

Discussion

In this article, we provide a general analysis, based on the Price equation, of the evolution of flexible social behaviors in structured populations. We show that behavioral responses and genetic relatedness (or similarity) have symmetric but independent effects on evolutionary stability; thus, they cannot be collapsed into a single index (e.g., an index of assortment; Fletcher and Doebeli 2009). Incorporating behavioral responses into a group-selection perspective elegantly emphasizes that social behaviors, such

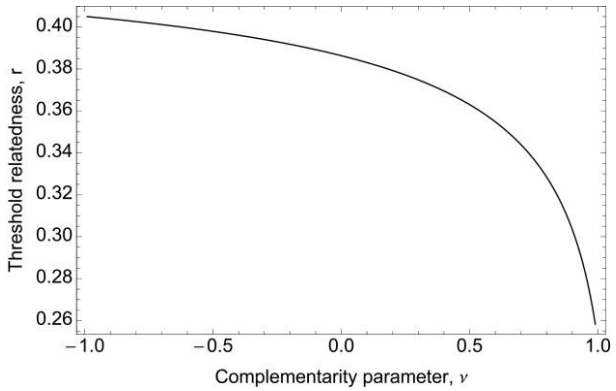


Figure 4: Threshold value of relatedness required to make the group-optimal trait values convergence stable, as a function of the complementarity parameter ν. For this plot, we fixed the interaction term ϕ at its group-optimal value and looked at the convergent stability of the valuation of the public good θ when approaching its group-optimal value.

as cooperation, that benefit others at a cost to self are always counterselected at the within-group level and positively selected at the between-group level, regardless of how tightly coordinated individuals' actions are. We also show that the classic result that cooperation increases in populations if and only if between-group selection is stronger than within-group selection (Wade 1980) generalizes to the case with behavioral responses. Furthermore, we show that behavioral responses and relatedness can interact synergistically in promoting the evolution of social behavior and specifically the evolution of group-optimal traits that might lead to ETIs. This synergistic relationship becomes particularly apparent in considerations of whether or not a group-optimal outcome is convergence stable, since relatedness must exceed a threshold value in order for convergence stability to be achieved. By applying our general model to a proximate mechanism based on behavioral objectives or social preferences, we also show how behavioral responses can evolve through the evolution of the valuation of public goods versus private costs and how this evolution depends on the level of synergism in contributions to the ecological benefit.

The Scope for Group Optimality

Our model provides a new analysis of the conditions that can lead to group optimality. In this respect, it is useful to compare our model to recent work by Gardner and Grafen (2009), who analyze the premise that natural selection is always expected to lead to group-optimal outcomes. Linking optimization of group fitness to a Price equation formalism, they find a formal justification for this premise only with clonal groups when $r = 1$. In the general case when $r \neq 1$, their analysis finds that group adaptations are not expected but are also not ruled out. Since Gardner and Grafen focus exclusively on genetically fixed behaviors, their analysis does not include the possible effects of evolving behavioral responses. In contrast, our model explicitly incorporates the effects of behavioral responses generated by proximate mechanisms, which allows us to consider the evolution of proximate mechanisms that eliminate within-group conflict (by yielding $\rho = 1$) even without complete clonality. This allows us to derive more specific conditions for when group-optimal behaviors are evolutionarily and convergent stable; such behaviors satisfy Gardner and Grafen's definition of a group adaptation (for an alternative definition, see Sober and Wilson 2011). We find that group adaptations sensu Gardner and Grafen are possible without clonal groups so long as relatedness meets a threshold that depends on the ecology of the interaction. Thus, we argue that their conclusion that "between-group selection can lead to group adaptation, but only in rather special circumstances" (Gardner and Grafen 2009, p. 668) underestimates the scope for group optimality. It is true that there is no unqualified justification for group fitness maximization, but as we show, there exists ample scope for the evolution of traits that result in either exactly or approximately group-optimal outcomes without requiring clonal groups.

One of the reasons that group optimality is important is its connection to ETIs. In order for an ETI to occur, either a demographic mechanism that ensures high relatedness (such as group formation by single propagules; Rainey and Kerr 2010) or a behavioral mechanism that ensures high behavioral responsiveness (such as strong prosocial preferences or policing) is needed to eliminate within-group selection. Once within-group selection is eliminated, extended between-group selection can drive a population toward a convergence-stable outcome with social behaviors that maximize group fitness. By illuminating the synergistic relationship between behavioral responsiveness and relatedness in determining when group optimality is convergence stable, our work suggests conditions under which ETIs themselves might be stable. Of course, a complete model of an ETI must detail how individuals specialize on different tasks (Gavrilets 2010) or how some types reproduce and others do not, such as in the germ-soma distinction (Michod et al. 2006); explicitly incorporating behavioral responses into models of task specialization is thus an important next step in future work.

Proximate Mechanisms and Objective Functions

The distinction between proximate and ultimate causes of behavior has been much emphasized in evolutionary biology with the recognition that proximate mechanisms can constrain adaptation of social behaviors (West et al. 2007). However, there has been little explicit modeling of the role of proximate mechanisms in social evolution. In this regard, our model provides a connection between social-evolution theory and the conceptual work in evolutionary genetics that focuses on dissecting the genotype-phenotype map and determining how this map evolves (i.e., the evolution of epistasis and pleiotropy; Hansen 2006; Wagner and Zhang 2011). In our model, the genotype-phenotype map is produced by the proximate behavioral mechanism, which translates heritable motivational traits (e.g., prosocial preferences) into actions. Given that the phenotype of interest is often fitness, the map also includes how actions translate into fitness through ecological payoffs. Just as the structure of gene regulatory networks determines how mutations affect gene expression, the cognitive and ecological constraints produced by a particular behavioral mechanism determine how responsive individuals can be to one another, given a particular demographic context

and relatedness. At the same time, our results also highlight the fact that behavioral responses through proximate mechanisms not only can constrain but also can facilitate the evolution of certain behaviors, such as those that maximize group benefit.

On a more practical level, our approach provides a toolkit to answer empirically salient questions about how specific behavioral mechanisms evolve. Here, we illustrate this by considering a proximate mechanism based on social preferences and asking how individuals evolve to value public goods and private costs in a public-goods game. We find that the objective functions that result in group-optimal outcomes involve discounting of the private cost of each individual as the provision of the public good increases (i.e., a positive ϕ trait). We also show that relatedness and complementarity of benefits enhances this discounting. Although direct measurements of motivations are scarce to nonexistent, some indirect evidence for such a pattern exists. For example, Fischbacher et al. (2001) performed an experiment that asked for a contribution schedule to a public good as a function of the average contributions of others and found that the most common pattern among Swiss undergraduates was monotonically increasing contributions; this is consistent with the pattern that would be produced by a positive ϕ trait. We believe that experiments in humans and animals directly aimed at capturing motivational states will be helpful in elucidating the underlying mechanism of decision making in social dilemmas.

Indirect Genetic Effects and Behavioral Responses in Structured Populations

There are important connections between our model and indirect-genetic-effects (IGE) models of quantitative genetics (e.g., Moore et al. 1997; Wolf et al. 1999; Bijma and Wade 2008; McGlothlin et al. 2010). The IGE approach measures the strength of selection on social behaviors by modeling the trait value of a focal individual as the sum of a direct genetic effect due to the focal individual and an IGE due to social partners. Using this framework, McGlothlin et al. (2010) derive an expression for the change in the mean of phenotype due to both individual and social effects on fitness analogous to our equation (4). In appendix A9, available online, we show that their coefficient of IGE, ψ, fulfills a role similar to ρ and that the change in mean phenotype from our equation (4) and that from equation (18) of McGlothlin et al. (2010) become equivalent when $\psi = \rho/[1 + \rho(N - 2)]$, their nonsocial-selection gradient β_N is the cost $-c$, and their social-selection gradient β_S is the benefit b. Thus, it is possible to map between our model and the IGE framework without changing the first-order evolutionary-stability predictions. However, our formulation highlights the exact symmetry between the scaled relatedness r and ρ. In contrast, r and ψ are not symmetric for $N > 2$ in phenotypic IGE models (McGlothlin et al. 2010), and other IGE models wrap behavioral responses into effects on variance components (see app. A9; Griffing 1981*a*; Bijma et al. 2007; Bijma and Wade 2008). The exact symmetry between r and ρ is an important feature because it reflects the fact that correlations between individuals' behavior have the same structural effect on selection regardless of whether those correlations are due to behavioral responses on the timescale of an interaction or to demographic factors that operate on a longer timescale. Moreover, we explicitly model the dependence of ρ on the proximate mechanisms that generate behavioral responses (such as the mechanism modeled in eq. [12]). This approach allows us to directly address the evolution of the phenotypic correlations given by ρ rather than treating it as an exogenously determined index, as the IGE literature has done so far.

In simple panmictic populations, there has been much work in both biology and economics on the evolution of flexible, responsive, or contingent behaviors. The most closely related models in biology are models of the continuous iterated prisoner's dilemma (CIPD; McNamara et al. 1999; Wahl and Nowak 1999; André and Day 2007) that look at the evolution of linear-response rules; given a behavioral mechanism of linear responses, the response slope is analogous to ρ. In economics, the most related field is that of "indirect evolution" (Güth 1995; Dekel et al. 2007; Heifetz et al. 2007*a*; Alger and Weibull 2010), which allows individuals to choose actions based on individual preferences and studies the evolution of those preferences. This work has focused mostly on the informational constraints required for players to act in a way that does not maximize their immediate self-interest. Unlike either the CIPD approach or most of the indirect-evolution literature, we embed the behavioral model in a structured population that allows for multiple levels of selection and selection among kin. Importantly, our general framework can be used with any behavioral model that allows one to calculate expected equilibrium behavioral outcomes and behavioral responses at that equilibrium.

In addition, our unified framework can also be used to study how group size and demographic parameters separately affect behavioral responsiveness and relatedness. For example, although larger groups are often less cooperative than smaller ones because of the decrease of relatedness with increasing group size, we can show that the opposite may be true under special conditions (app. A8, available online). Future work should therefore focus on how group-beneficial outcomes might evolve in response to environmental variability and a number of demographic

variables, including population size, dispersal rate, extinction risk, and carrying capacity (Lehmann and Rousset 2010).

Conclusion

We show that behavioral responses can significantly expand the scope for group-optimal behaviors that are associated with ETIs. Our results emphasize the important interplay between behavioral responses and relatedness and show that the two can reinforce each other to sustain higher levels of group benefit than would result from each in isolation. Our model also highlights the crucial role of proximate mechanisms of behavior in determining the magnitude and evolution of behavioral responses. A promising direction for future work is to combine models such as this one and experiments in humans and animals aimed at capturing motivational states: experiments can elucidate the underlying neurological mechanisms of decision making in social dilemmas, whereas models can illuminate relevant selection pressures that shape the evolution of these mechanisms.

Acknowledgments

E.A. was supported as a Postdoctoral Fellow at the National Institute for Mathematical and Biological Synthesis (NIMBioS). J.V. is supported as an Omidyar Fellow at the Santa Fe Institute. This work was also supported by a NIMBioS short-term visitor grant. NIMBioS is sponsored by the National Science Foundation, the U.S. Department of Homeland Security, and the U.S. Department of Agriculture through National Science Foundation award EF-0832858, with additional support from the University of Tennessee, Knoxville.

Literature Cited

Ajar, E. 2003. Analysis of disruptive selection in subdivided populations. BMC Evolutionary Biology 3:22. doi:10.1186/1471-2148-3-22.

Akçay, E., and J. Roughgarden. 2011. The evolution of payoff matrices: providing incentives to cooperate. Proceedings of the Royal Society B: Biological Sciences 278:2198–2206.

Akçay, E., J. Van Cleve, M. W. Feldman, and J. E. Roughgarden. 2009. A theory for the evolution of other-regard integrating proximate and ultimate perspectives. Proceedings of the National Academy of Sciences of the USA 106:19061–19066. doi:10.1073/pnas.0904357106.

Alexander, R. D., and G. Borgia. 1978. Group selection, altruism, and the levels of organization of life. Annual Review of Ecology and Systematics 9:449–474.

Alger, I., and J. W. Weibull. 2010. Evolutionary stability, co-operation and Hamilton's rule. Carleton Economic Paper 10-11. Department of Economics, Carleton University, Ottawa, ON.

André, J.-B., and T. Day. 2007. Perfect reciprocity is the only evolutionarily stable strategy in the continuous iterated prisoner's dilemma. Journal of Theoretical Biology 247:11–22. doi:10.1016/j.jtbi.2007.02.007.

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

Baumgartner, T., M. Heinrichs, A. Vonlanthen, U. Fischbacher, and E. Fehr. 2008. Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. Neuron 58:639–650.

Bijma, P., and M. J. Wade. 2008. The joint effects of kin, multilevel selection and indirect genetic effects on response to genetic selection. Journal of Evolutionary Biology 21:1175–1188. doi:10.1111/j.1420-9101.2008.01550.x.

Bijma, P., W. M. Muir, and J. A. M. Van Arendonk. 2007. Multilevel selection 1: quantitative genetics of inheritance and response to selection. Genetics 175:277–288. doi:10.1534/genetics.106.062711.

Boyd, R., H. Gintis, and S. Bowles. 2010. Coordinated punishment of defectors sustains cooperation and can proliferate when rare. Science 328:617–620. doi:10.1126/science.1183665.

Brosnan, S., and F. B. M. de Waal. 2002. A proximate perspective on reciprocal altruism. Human Nature 13:129–152.

Christiansen, F. B. 1991. On conditions for evolutionary stability for a continuously varying character. American Naturalist 138:37–50.

Dekel, E., J. C. Ely, and O. Yilankaya. 2007. Evolution of preferences. Review of Economic Studies 74:685–704.

Donaldson, Z. R., and L. J. Young. 2008. Oxytocin, vasopressin, and the oxytocin, vasopressin, and the neurogenetics of sociality. Science 322:900–904.

Eshel, I., and U. Motro. 1981. Kin selection and strong evolutionary stability of mutual help. Theoretical Population Biology 19:420–433.

Falk, A., and U. Fischbacher. 2006. A theory of reciprocity. Games and Economic Behavior 54:293–315. doi:10.1016/j.geb.2005.03.001.

Fischbacher, U., S. Gächter, and E. Fehr. 2001. Are people conditionally cooperative? evidence from a public goods experiment. Economics Letters 71:397–404. doi:10.1016/S0165-1765(01)00394-9.

Fletcher, J., and M. Doebeli. 2009. A simple and general explanation for the evolution of altruism. Proceedings of the Royal Society B: Biological Sciences 276:13–19.

Frank, S. A. 1998. Foundations of social evolution. Princeton University Press, Princeton, NJ.

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

Gardner, A., and A. Grafen. 2009. Capturing the superorganism: a formal theory of group adaptation. Journal of Evolutionary Biology 22:659–671.

Gavrilets, S. 2010. Rapid transition towards the division of labor via evolution of developmental plasticity. PLoS Computational Biology 6:e1000805. doi:10.1371/journal.pcbi.1000805.

Goodson, J. L., and Y. Wang. 2006. Valence-sensitive neurons exhibit divergent functional profiles in gregarious and asocial species. Proceedings of the National Academy of Sciences of the USA 45:17013–17017.

Griffing, B. 1967. Selection in reference to biological groups. I. Individual and group selection applied to populations of unordered groups. Australian Journal of Biological Sciences 20:127–139.

———. 1981*a*. A theory of natural selection incorporating interaction among individuals. I. The modeling process. Journal of Theoretical Biology 89:635–658. doi:10.1016/0022-5193(81)90033-3.

———. 1981*b*. A theory of natural selection incorporating interaction among individuals. II. Use of related groups. Journal of Theoretical Biology 89:659–677. doi:10.1016/0022-5193(81)90034-5.

Güth, W. 1995. An evolutionary approach to explaining cooperative behavior by reciprocal incentives. International Journal of Game Theory 24:323–344. doi:10.1007/BF01243036.

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

———. 1975. Innate social aptitudes of man: an approach from evolutionary genetics. Pages 133–153 *in* R. Fox, ed. Biosocial anthropology. Malaby, London.

Hansen, T. F. 2006. The evolution of genetic architecture. Annual Review of Ecology, Evolution, and Systematics 37:123–157. doi:10.1146/annurev.ecolsys.37.091305.110224.

Heifetz, A., C. Shannon, and Y. Spiegel. 2007*a*. The dynamic evolution of preferences. Economic Theory 32:251–286. doi:10.1007/s00199-006-0121-7.

———. 2007*b*. What to maximize if you must. Journal of Economic Theory 133:31–57. doi:10.1016/j.jet.2005.05.013.

Kerr, B., P. Godfrey-Smith, and M. W. Feldman. 2004. What is altruism? Trends in Ecology & Evolution 19:135–140. doi:10.1016/j.tree.2003.10.004.

Lehmann, L., and L. Keller. 2006. The evolution of cooperation and altruism: a general framework and a classification of models. Journal of Evolutionary Biology 19:1365–1376. doi:10.1111/j.1420-9101.2006.01119.x.

Lehmann, L., and F. Rousset. 2010. How life history and demography promote or inhibit the evolution of helping behaviours. Philosophical Transactions of the Royal Society B: Biological Sciences 365:2599–2617. doi:10.1098/rstb.2010.0138.

Leigh, E. G., Jr. 1977. How does selection reconcile individual advantage with the good of the group? Proceedings of the National Academy of Sciences of the USA 74:4542–4546.

———. 2010. The group selection controversy. Journal of Evolutionary Biology 23:6–19.

Levin, B. R., M. Lipsitch, and S. Bonhoeffer. 1999. Population biology, evolution, and infectious disease: convergence and synthesis. Science 283:806–809.

Levin, S. A. 2009. Games, groups, norms, and societies. Pages 143–153 *in* S. A. Levin, ed. Games, groups, and the global good. Springer Series in Game Theory. Springer, Berlin.

Levine, D. K. 1998. Modeling altruism and spitefulness in experiments. Review of Economic Dynamics 1:593–622. doi:10.1006/redy.1998.0023.

Maynard Smith, J., and G. R. Price. 1973. The logic of animal conflict. Nature 246:15–18.

Maynard Smith, J., and E. Szathmáry. 1995. The major transitions in evolution. Oxford University Press, Oxford.

McGlothlin, J. W., A. J. Moore, J. B. Wolf, and E. D. Brodie III. 2010. Interacting phenotypes and the evolutionary process. III. Social evolution. Evolution 64:2558–2574. doi:10.1111/j.1558-5646.2010.01012.x.

McNamara, J. M., C. E. Gasson, and A. I. Houston. 1999. Incorporating rules for responding into evolutionary games. Nature 401:368–371.

Michod, R. E. 2005. On the transfer of fitness from the cell to the multicellular organism. Biology and Philosophy 20:967–987. doi:10.1007/s10539-005-9018-2.

———. 2006. The group covariance effect and fitness trade-offs during evolutionary transitions in individuality. Proceedings of the National Academy of Sciences of the USA 103:9113–9117. doi:10.1073/pnas.0601080103.

Michod, R. E., Y. Viossat, C. A. Solari, M. Hurand, and A. M. Nedelcu. 2006. Life-history evolution and the origin of multicellularity. Journal of Theoretical Biology 239:257–272. doi:10.1016/j.jtbi.2005.08.043.

Miller, M. B., and B. L. Bassler. 2001. Quorum sensing in bacteria. Annual Review of Microbiology 55:165–199.

Moore, A. J., E. D. Brodie III, and J. B. Wolf. 1997. Interacting phenotypes and the evolutionary process. I. Direct and indirect genetic effects of social interactions. Evolution 51:1352–1362.

Pepper, J. W. 2000. Relatedness in trait group models of social evolution. Journal of Theoretical Biology 206:355–368. doi:10.1006/jtbi.2000.2132.

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. 1992. Quantitative genetics, inclusive fitness, and group selection. American Naturalist 139:540–558.

———. 1994. Genetic relatedness in viscous populations. Evolutionary Ecology 8:70–73. doi:10.1007/BF01237667.

Rainey, P. B., and B. Kerr. 2010. Cheats as first propagules: a new hypothesis for the evolution of individuality during the transition from single cells to multicellularity. BioEssays 32:872–880. doi:10.1002/bies.201000039.

Rousset, F. 2004. Genetic structure and selection in subdivided populations. Princeton University Press, Princeton, NJ.

Sober, E., and D. S. Wilson. 2011. Adaptation and natural selection revisited. Journal of Evolutionary Biology 24:462–468.

Taylor, P. D. 1992. Altruism in viscous populations: an inclusive fitness model. Evolutionary Ecology 6:352–356.

Taylor, P. D., and T. Day. 2004. Stability in negotiation games and the emergence of cooperation. Proceedings of the Royal Society B: Biological Sciences 271:669–674. doi:10.1098/rspb.2003.2636.

Taylor, P. D., and S. A. Frank. 1996. How to make a kin selection model. Journal of Theoretical Biology 180:27–37. doi:10.1006/jtbi.1996.0075.

Wade, M. J. 1980. Kin selection: its components. Science 210:665–667.

Wagner, G. P., and J. Zhang. 2011. The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. Nature Reviews Genetics 12:204–213. doi:10.1038/nrg2949.

Wahl, L. M., and M. A. Nowak. 1999. The continuous Prisoner's Dilemma: I. Linear reactive strategies. Journal of Theoretical Biology 200:307–321. doi:10.1006/jtbi.1999.0996.

West, S. A., A. S. Griffin, and A. Gardner. 2007. Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. Journal of Evolutionary Biology 20:415–432.

West-Eberhard, M. J. 1987. Flexible strategy and social evolution. Pages 35–51 *in* Y. Ito, J. L. Brown, and J. Kikkawa, eds. Animal societies: theories and facts. Japan Scientific Societies, Tokyo.

Williams, G. C. 1966. Adaptation and natural selection. Princeton University Press, Princeton NJ.

Wilson, D. S., and E. Wilson. 2007. Rethinking the theoretical foundation of sociobiology. Quarterly Review of Biology 82:327–348.

Wolf, J. B., E. D. Brodie III, and A. J. Moore. 1999. Interacting phenotypes and the evolutionary process. II. Selection resulting from social interactions. American Naturalist 153:254–266. doi: 10.1086/303168.

Wright, J., and I. Cuthill. 1989. Manipulation of sex differences in parental care. Behavioral Ecology and Sociobiology 25:171–181.

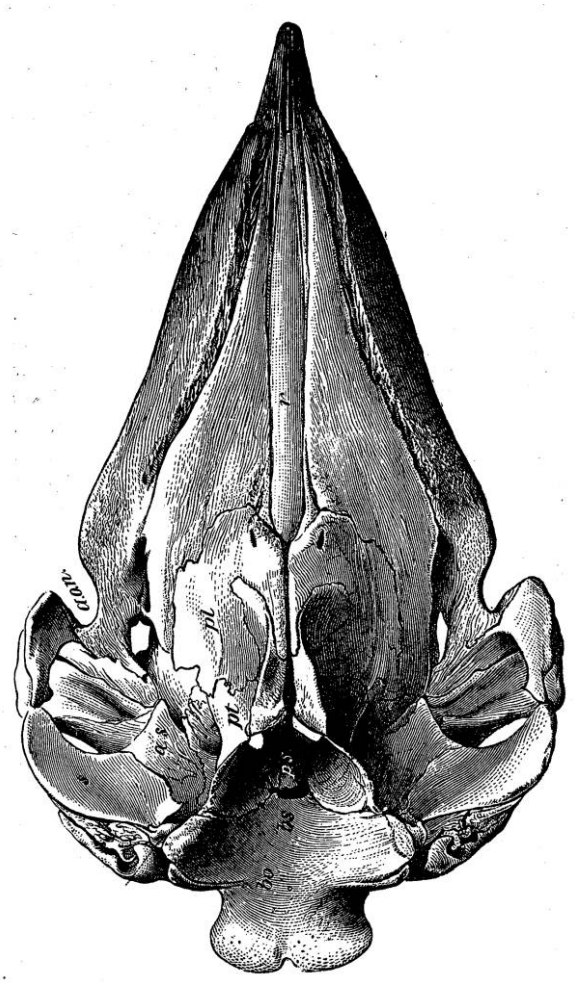
Wright, S. 1943. Isolation by distance. Genetics 28:114–138.

———. 1949. The genetical structure of populations. Annals of Eugenics 15:323–354. doi:10.1111/j.1469-1809.1949.tb02451.x.

Wynne-Edwards, V. C. 1962. Animal dispersion in relation to social behaviour. Oliver & Boyd, Edinburgh.

Young, K. A., K. L. Gobrogge, Y. Liu, and Z. Wang. 2011. The neurobiology of pair bonding: insights from a socially monogamous rodent. Frontiers in Neuroendocrinology 32:53–69.

Associate Editor: Peter D. Taylor
Editor: Ruth G. Shaw



Skull of adult *Physeter macrocephalus*, seen from above. From "The Sperm Whales, Giant and Pygmy" by Theodore Gill (*American Naturalist*, 1871, 4:725-743).

 DOI: 10.1086/663691

Appendix A from E. Akçay and J. Van Cleve, "Behavioral Responses in Structured Populations Pave the Way to Group Optimality"

(Am. Nat., vol. 179, no. 2, p. 257)

Methods and Additional Analysis

A1. Genetic Response to Selection with Behavioral Responses

We start with the "kin-selection" partition of fitness, using the Price equation (Price 1970; Queller 1992; Bijma and Wade 2008), which gives the genetic response to selection as

$$\Delta\overline{G} = \mathrm{Cov}(G_i, w) = \beta_{w,p_i}\mathrm{Cov}(G_i, p_i) + \sum_{j\neq i}\beta_{w,p_j}\mathrm{Cov}(G_i, p_j).$$

Here, G_i is the additive genetic (breeding) value of the focal individual and β_{w,p_i} and β_{w,p_j} are partial-regression coefficients of a focal individual i's fitness on its own motivational trait (p_i) and its neighbors' motivational trait (p_j), respectively. Following common assumptions in quantitative-genetic approaches (Moore et al. 1997; Wolf et al. 1999; Bijma and Wade 2008), we assume additive effects of genotype and environment on the trait. We further use the relation $\mathrm{Cov}(G_i, G_j) = R\,\mathrm{Var}(G_i)$, where R is the coefficient of relatedness between group members and $\mathrm{Var}(G_i)$ is the additive genetic variance (Queller 1992; Frank 1998), to obtain

$$\Delta\overline{G} = \mathrm{Var}(G_i)[\beta_{w,p_i} + (N-1)R\beta_{w,p_j}]. \tag{A1}$$

Equation (A1) holds for populations with arbitrarily complex demographic structures, but care must be taken in evaluating the fitness effects (the partial-regression coefficients) and relatedness. In many structured populations, local competition for limited resources with related individuals can counteract the effect of a motivational trait on a focal individual's fitness (Taylor 1992; Queller 1994; West et al. 2002; Rousset 2004; Lehmann and Rousset 2010). This effect of local competition can be found in the partial-regression coefficients β_{w,p_i} and β_{w,p_j}, which give the effect of the motivational trait (p_i or p_j) on the fertility of the focal individual plus the effect of local competition, which is a function of the total change in fertility in the population. The local-competition term is a function of demographic parameters such as population size and dispersal rate. Generally, one can rearrange equation (A1) so that a scaled-relatedness coefficient r includes the effect of local competition and the partial-regression coefficients β_{F_i,p_i} and β_{F_i,p_j} measure only the effects of the motivational trait on fertility F. We present an example of how this is done in an island model in appendix A2. In this way, we can write equation (1):

$$\Delta\overline{G} \propto \mathrm{Var}(G_i)[\beta_{F_i,p_i} + (N-1)r\beta_{F_i,p_j}].$$

(Queller 1994; Lehmann and Rousset 2010). Although strong local competition may result in a negative value of this scaled-relatedness coefficient, r is generally greater than $-1/(n-1) \geq -1/(N-1)$ for many well-studied demographic scenarios; n is the local population or deme size (Lehmann and Rousset 2010), and $n \geq N$, since social groups are embedded in demes.

Substituting equations (2) and (3) into equation (1), we can write, for the genetic change in the population after one round of selection,

$$\Delta\overline{G} \propto \mathrm{Var}(G_i)\left[\frac{\partial a_i}{\partial p_i}\left(\frac{\partial u_i}{\partial a_i} + \sum_{j\neq i}\rho_{ij}\frac{\partial u_i}{\partial a_j}\right) + (N-1)r\frac{\partial a_j}{\partial p_j}\left(\frac{\partial u_i}{\partial a_j} + \sum_{k\neq j}\rho_{jk}\frac{\partial u_i}{\partial a_k}\right)\right]. \tag{A2}$$

When derivatives are evaluated in a monomorphic population, $\partial a_i/\partial p_i = \partial a_j/\partial p_j$, and we can simplify this equation by setting $k = \mathrm{Var}(G_i)(\partial a_i/\partial p_i)$ and writing

$$\Delta\overline{G} \propto k\left[\left(\frac{\partial u_i}{\partial a_i} + \sum_{j\neq i}\rho_{ij}\frac{\partial u_i}{\partial a_j}\right) + (N-1)r\left(\frac{\partial u_i}{\partial a_j} + \sum_{k\neq j}\rho_{jk}\frac{\partial u_i}{\partial a_k}\right)\right]. \tag{A3}$$

If we accept by convention that $\partial a_i/\partial p_i > 0$, that is, that an increase in the phenotypic value leads to an increase in the behavioral action, then $k > 0$, and hence the sign of $\Delta\overline{G}$ is determined by the sign of the term in the brackets. In particular, the value of p that sets the brackets in equation (A3) equal to 0 and $\Delta\overline{G} = 0$ is a candidate evolutionarily stable strategy. Using the brackets in equation (A3), we are ready to ask when group-beneficial behaviors can be evolutionarily stable (ES).

Using the definitions of b and c given in the main text and noting that in a monomorphic population, $\rho_{ij} = \rho$ for all $j \neq i$ ($\rho_{ii} = 1$ by definition), we can write

$$\frac{\partial u_i}{\partial a_i} + \sum_{j\neq i}\rho_{ij}\frac{\partial u_i}{\partial a_j} = \frac{\partial u_i}{\partial a_i} + (N-1)\rho\frac{\partial u_i}{\partial a_j} = b\rho(N-1) - c \tag{A4}$$

and

$$\frac{\partial u_i}{\partial a_j} + \sum_{k\neq j}\rho_{jk}\frac{\partial u_i}{\partial a_k} = \frac{\partial u_i}{\partial a_j} + \rho\left[\frac{\partial u_i}{\partial a_i} + (N-2)\frac{\partial u_i}{\partial a_k}\right] = b[\rho(N-2)+1] - \rho c. \tag{A5}$$

Equation (A4) represents the direct fertility effect of a small change in the motivational trait of the focal individual on itself, while equation (A5) represents the indirect fertility effect of a small change in the motivational trait of a neighbor related to the focal individual. The magnitude and sign of these direct and indirect effects depend on the level of responsiveness measured by ρ. When there is no responsiveness, $\rho = 0$, the direct and indirect fertility effects are $-c$ and b, respectively, as found in classic models of kin selection (Hamilton 1964). Substituting equations (A4) and (A5) into equation (A3) and rearranging gives us

$$\Delta\overline{G} \propto k\big(b\rho(N-1) - c + r(N-1)\{b[\rho(N-2)+1] - \rho c\}\big), \tag{A6}$$

which is equation (4). Setting equation (4) equal to 0 and rearranging yields equation (5). In order to ensure that the ratio b/c has a unique solution in equation (5), b/c must be monotonically decreasing; conditions for such a decrease are given in appendix A5.

The Group-Selection Partition

The group-selection partition of fitness is given by

$$\Delta\overline{G} = \mathrm{Cov}(G_i, w) = \beta_{w,p}\mathrm{Cov}(G_i, p) + \beta_{w,\Delta p}\mathrm{Cov}(G_i, \Delta p_i), \tag{A7}$$

where p is the mean motivational trait of the group and Δp is the deviation of the focal individual's motivational trait from the group mean. The first term then corresponds to selection among groups, and the second term gives selection within groups. The partial-regression coefficients $\beta_{w,p}$ and $\beta_{w,\Delta p_i}$ can be written as

$$\beta_{w,p} = \beta_{w,p_i} + (N-1)\beta_{w,p_j}, \tag{A8}$$

$$\beta_{w,\Delta p_i} = \beta_{w,p_i} - \beta_{w,p_j} \tag{A9}$$

(Bijma and Wade 2008), where $\beta_{w,p}$ is the partial regression of the focal individual's fitness on the mean group phenotype and $\beta_{w,\Delta p_i}$ is the partial regression of the focal individual's fitness on the difference between the focal individual's phenotype and the mean group phenotype. After substituting equations (A8) and (A9) into equation (A7), using the same rearrangement of fertility and demographic effects as in equation (1), and using the definitions of b, c, and ρ, we can write the genetic response to selection as

$$\Delta\overline{G} \propto k\{[b(N-1) - c][1 + \rho(N-1)][1 + r(N-1)] - (N-1)(b+c)(1-\rho)(1-r)\},$$

which is the same as equation (6).

A2. Scaled-Relatedness Coefficient in an Island Model with Overlapping Generations

In equation (1), the relatedness coefficient r is assumed to be appropriately defined so as to account for demographic effects such as local competition. In this section, we work out a simple example explicitly to illustrate how this can be done. We assume that our population obeys a finite-island model of migration (Wright 1943) with overlapping generations (Taylor and Irwin 2000), with the following life cycle: (1) n haploid adult individuals randomly mate and produce a large number of offspring; (2) offspring migrate or disperse independently to each of the $D - 1$ other demes in the population with probability $m/(D - 1)$ or remain in their natal deme with probability $1 - m$; (3) each adult survives to the next generation with probability s, and adults who die are replaced by offspring who mature into new adults; (4) all adults express a social trait that affects their fertility and the fertility of the neighbors in their deme.

Under weak selection, the expected change in the average frequency of a mutant allele, $\mathrm{E}(\Delta q)$, is proportional to the selection gradient S (Rousset 2004); in an island model of population structure, S is given by

$$S = \frac{\partial w_i}{\partial p_i} + \frac{\partial w_i}{\partial p_0} Q_0 + \frac{\partial w_i}{\partial p_1} Q_1, \tag{A10}$$

where w_i is the expected number of offspring of a focal individual i (fitness), p_i is the focal individual's phenotype, p_0 is the average phenotype of the focal individual's deme (excluding the focal individual), and p_1 is the average phenotype of all demes, excluding the focal deme. Further, Q_0 is the probability of identity by state for two alleles drawn randomly without replacement from the same deme, and Q_1 is the probability of identity by state for two alleles drawn from different demes. If we assume that total population size is fixed (i.e., inelastic demography), then the sum of all partial derivatives of w_i is 0 (Rousset and Billiard 2000). Thus, we can rewrite S as

$$S = (1 - Q_1)\left(\frac{\partial w_i}{\partial p_i} + \frac{\partial w_i}{\partial p_0} R\right), \tag{A11}$$

where $R = (Q_0 - Q_1)/(1 - Q_1)$ is Wright's F_{ST} (Wright 1943, 1949) and is a measure of relatedness. The fitness of a focal individual is given by (to first order in the strength of selection)

$$\begin{aligned} w_i = s &+ \frac{(1 - s)(1 - m)F(p_i, p_0)}{(1 - m)F(p_{0,R}, p_{0,R}) + mF(p_1, p_1)} \\ &+ \frac{(1 - s)mF(p_i, p_0)}{(1 - m)F(p_1, p_1) + [m/(D - 1)][(D - 2)F(p_1, p_1) + F(p_{0,R}, p_{0,R})]}. \end{aligned} \tag{A12}$$

A social interaction between individual i and $N - 1$ other individuals j_1 through j_{N-1} in its deme with phenotypes p_i and p_{j_1} through $p_{j_{N-1}}$, respectively, changes the fertility of individual i, which is measured by $F(p_i, p_{j_1}, \ldots, p_{j_{N-1}})$. Since we assume weak selection and continuous phenotypes, the fitness of the focal individual depends on its own phenotype, p_i, the average phenotype in its deme (excluding itself), p_0, and the average phenotype across all demes except the focal deme, p_1. Note that $p_{0,R}$ is the average phenotype in the focal deme including the focal individual and is equal to $(1/n)p_i + [(n - 1)/n]p_0$.

To calculate S, we must calculate the derivatives in equation (A11). The derivative of the focal individual's fitness with respect to its own phenotype is given by

$$\frac{\partial w_i}{\partial p_i} = \frac{1 - s}{F_{\mathrm{r}}}\left\{\frac{\partial F_i}{\partial p_i} - \frac{1}{n}\left[(1 - m)^2 + \frac{m^2}{D - 1}\right]\left[\frac{\partial F_i}{\partial p_i} + (N - 1)\frac{\partial F_i}{\partial p_j}\right]\right\}, \tag{A13}$$

where the derivatives $\partial F_i/\partial p_i$ and $\partial F_i/\partial p_j$ are evaluated at $p_i = p_j = p_{\mathrm{r}}$ and $F_{\mathrm{r}} = F(p_{\mathrm{r}}, p_{\mathrm{r}}, \ldots, p_{\mathrm{r}})$, the fertility of the resident phenotype. The derivatives of the fertility functions, $\partial F_i/\partial p_i$ and $\partial F_i/\partial p_j$, correspond to the partial-regression coefficients β_{F_i, p_i} and β_{F_i, p_j}, respectively, in equation (1). From equations (A4) and (A5), we have

$$\frac{\partial F}{\partial p_i} = \frac{\partial a_i}{\partial p_i}[b\rho(N - 1) - c] \tag{A14}$$

and

$$\frac{\partial F}{\partial p_j} = \frac{\partial a_i}{\partial p_i}\{b[\rho(N-2)+1]-\rho c\}, \tag{A15}$$

respectively. Combining equations (A14) and (A15) with equation (A13) yields

$$\frac{\partial w_i}{\partial p_i} = \frac{(1-s)(\partial a_i/\partial p_i)}{F_r}\left\{b\rho(N-1)-c-\frac{1+(N-1)\rho}{n}\left[(1-m)^2+\frac{m^2}{D-1}\right][b(N-1)-c]\right\}, \tag{A16}$$

$$\frac{\partial w_i}{\partial p_0} = \frac{(1-s)(\partial a_i/\partial p_i)}{F_r} \times \left((N-1)\{b[\rho(N-2)+1]-\rho c\}-\frac{(n-1)[1+(N-1)\rho]}{n}\left\{(1-m)^2+\frac{m^2}{D-1}[b(N-1)-c]\right\}\right). \tag{A17}$$

Using equations (A16) and (A17) in equation (A11), we obtain

$$S = \frac{(1-s)(\partial a_i/\partial p_i)(1-Q_1)}{F_r}\Bigg(b\rho(N-1)-c+(N-1)\{b[\rho(N-2)+1]-\rho c\}R - [1+(N-1)\rho][b(N-1)-c]\left[(1-m)^2+\frac{m^2}{D-1}\right]\left(\frac{1}{n}+\frac{n-1}{n}R\right)\Bigg). \tag{A18}$$

Using equation (A18), we can rewrite the condition $S = 0$ as

$$\frac{b}{c} = \frac{1+r\rho(N-1)}{(N-1)[r+\rho+r\rho(N-2)]},$$

which is the Hamilton-type rule given in equation (5); here,

$$r = \frac{R-R^*}{1-R^*} \tag{A19}$$

is the scaled relatedness corrected for the effect of local competition and

$$R^* = \left[(1-m)^2+\frac{m^2}{D-1}\right]\left(\frac{1}{n}+\frac{n-1}{n}R\right). \tag{A20}$$

Solving the appropriate equilibrium equation for R when $D \to \infty$,

$$R = s^2R + 2s(1-s)(1-m)\left(\frac{1}{n}+\frac{n-1}{n}R\right) + (1-s)^2(1-m)^2\left(\frac{1}{n}+\frac{n-1}{n}R\right)$$

(Taylor and Irwin 2000), yields

$$R = \frac{(1-m)[1-m(1-s)+s]}{1+2m(n-1)-m^2(n-1)(1-s)+s} \tag{A21}$$

for the relatedness coefficient. Plugging this value of R into equations (A19) and (A20) yields a scaled-relatedness coefficient of

$$r = \frac{2(1-m)s}{n[2-m(1-s)]+2(1-m)s},$$

which simplifies to $r = s/n$ to first order in both $1/n$ and m.

A3. Stability Condition for the Behavioral Equilibrium

In this section, we derive necessary and sufficient conditions for the behavioral equilibrium $\mathbf{a}^*$ to be stable under the dynamical system given by

$$\frac{da_i}{dt} = \frac{\partial x_i}{\partial a_i}. \tag{A22}$$

The behavioral equilibrium $\mathbf{a}^*$ is defined as the N-tuple $(a_1^*, \ldots, a_N^*)$ that sets the right-hand side of equation (A22) equal to 0. In order for $\mathbf{a}^*$ to be a stable rest point of equation (A22), the eigenvalues of the Jacobian matrix,

$$\mathbf{J} = \begin{pmatrix} \frac{\partial^2 x_1}{\partial a_1^2} & \frac{\partial^2 x_1}{\partial a_1 \partial a_2} & \cdots & \frac{\partial^2 x_1}{\partial a_1 \partial a_N} \\ \frac{\partial^2 x_2}{\partial a_1 \partial a_2} & \frac{\partial^2 x_2}{\partial a_2^2} & \cdots & \frac{\partial^2 x_2}{\partial a_2 \partial a_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 x_N}{\partial a_1 \partial a_N} & \frac{\partial^2 x_N}{\partial a_2 \partial a_N} & \cdots & \frac{\partial^2 x_N}{\partial a_N^2} \end{pmatrix}, \tag{A23}$$

must have negative real part. Since we are interested in evolutionary stability, we can evaluate equation (A23) at a monomorphic equilibrium where all individuals are genetically identical. This means that $\partial^2 x_i/\partial a_i^2 = \delta_1$ and $\partial^2 x_i/\partial a_i \partial a_j = \delta_2$ for some δ_1 and δ_2 for all i and j in $(1, \ldots, N)$. Thus, we can rewrite $\mathbf{J}$ as

$$\begin{pmatrix} \delta_1 & \delta_2 & \cdots & \delta_2 \\ \delta_2 & \delta_1 & \cdots & \delta_2 \\ \vdots & \vdots & \ddots & \vdots \\ \delta_2 & \delta_2 & \cdots & \delta_1 \end{pmatrix}. \tag{A24}$$

We next calculate the eigenvalues of $\mathbf{J}$ by solving $\det(\lambda\mathbf{I} - \mathbf{J}) = 0$ for λ, where $\mathbf{I}$ is the $N \times N$ identity matrix. Using elementary row and column operations and expanding the determinant via minors, we can show that

$$\begin{aligned} \det(\lambda\mathbf{I} - \mathbf{J}) &= \det \begin{pmatrix} \lambda - \delta_1 & -\delta_2 & \cdots & -\delta_2 \\ \delta_1 - \delta_2 - \lambda & \lambda + \delta_2 - \delta_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \delta_1 - \delta_2 - \lambda & 0 & \cdots & \lambda + \delta_2 - \delta_1 \end{pmatrix} \\ &= (\lambda - \delta_1) \det \begin{pmatrix} \lambda + \delta_2 - \delta_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda + \delta_2 - \delta_1 \end{pmatrix} \\ &\quad + (N-1)\delta_2 \det \begin{pmatrix} \delta_1 - \delta_2 - \lambda & 0 & \cdots & 0 \\ \delta_1 - \delta_2 - \lambda & \lambda + \delta_2 - \delta_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \delta_1 - \delta_2 - \lambda & 0 & \cdots & \lambda + \delta_2 - \delta_1 \end{pmatrix} \\ &= (\lambda - \delta_1)(\lambda + \delta_2 - \delta_1)^{N-1} - (N-1)\delta_2(\lambda + \delta_2 - \delta_1)^{N-1} \\ &= (\lambda + \delta_2 - \delta_1)^{N-1}[\lambda - \delta_1 - (N-1)\delta_2]. \end{aligned} \tag{A25}$$

Setting equation (A25) equal to 0 shows that the two eigenvalues are $\lambda_1 = \delta_1 - \delta_2$ and $\lambda_2 = \delta_1 + (N-1)\delta_2$. We assume that $\delta_1 < 0$, since this means that the behavioral equilibrium is a local peak of each individual i's objective function x_i with respect to its own action a_i for all i. Thus, in order for λ_1 and λ_2 to be negative, the following inequality must be met:

$$\delta_1 < \delta_2 < \frac{-\delta_1}{N-1}. \tag{A26}$$

If we define the response coefficient ρ as in equation (13), then the condition in equation (A26) means that ρ must satisfy the inequality

$$\frac{-1}{N-1} < \rho < 1. \tag{A27}$$

A4. Deriving the Second-Order Convergence-Stability (CS) Condition

In this section, we derive the second-order CS condition. In order to calculate the CS condition, we set $p_i = p_j = p_R$ for all i, j, reflecting the requirement that we look at how the selection gradient changes with different motivational traits p_R fixed in the population. Assuming that the genetic variance in the trait does not change with the trait value, the CS condition is that the derivative of the terms in the brackets in equation (A2) with respect to the resident motivational trait p_R be less than 0; that is,

$$\frac{d}{dp_R}\left[\frac{\partial a_i}{\partial p_i}\sum_j \rho_{ij}\frac{\partial u_i}{\partial a_j} + (N-1)r\frac{\partial a_j}{\partial p_j}\sum_k \rho_{jk}\frac{\partial u_i}{\partial a_k}\right]_{p_R} < 0, \tag{A28}$$

where the subscript on the bracket indicates that all derivatives, as well as ρ_{ij} and ρ_{jk}, are to be evaluated at $p_1 = p_2 = \ldots = p_N = p_R$ and the corresponding behavioral equilibrium actions a^*. Expanding this derivative, we get

$$\begin{aligned}&\frac{d}{dp_R}\frac{\partial a_i}{\partial p_i}\left(\sum_j \rho_{ij}\frac{\partial u_j}{\partial a_j}\right) + (N-1)r\frac{d}{dp_R}\frac{\partial a_j}{\partial p_i}\left(\sum_k \rho_{jk}\frac{\partial u_i}{\partial a_k}\right)\\ &+ \frac{\partial a_i}{\partial p_i}\sum_j\left(\frac{d\rho_{ij}}{dp_R}\frac{\partial u_i}{\partial a_j} + \rho_{ij}\frac{d}{dp_R}\frac{\partial u_i}{\partial a_j}\right) + \frac{\partial a_j}{\partial p_j}(N-1)r\sum_k\left(\frac{d\rho_{jk}}{dp_R}\frac{\partial u_i}{\partial a_k} + \rho_{jk}\frac{d}{dp_R}\frac{\partial u_i}{\partial a_k}\right).\end{aligned} \tag{A29}$$

Note that because we are changing all motivational traits at the same time, the total derivative of, say, ρ_{ij} with respect to p_R can be written as

$$\frac{d\rho_{ij}}{dp_R} = \frac{d\rho_{ij}}{dp_i} + \frac{d\rho_{ij}}{dp_j} + \sum_{k\neq i,j}\frac{d\rho_{ij}}{dp_k}.$$

The total derivatives with respect to p_i, p_j, and p_k are evaluated as illustrated in the following example:

$$\frac{d\rho_{ij}}{dp_i} = \frac{\partial\rho_{ij}}{\partial p_i} + \sum_k \frac{\partial a_i}{\partial p_i}\rho_{ik}\frac{\partial\rho_{ij}}{\partial a_k},$$

where we have used the definition $\rho_{ik} \equiv (\partial a_k/\partial p_i)/(\partial a_i/\partial p_i)$. To evaluate the sum, we note that there are three unique cases for $\partial\rho_{ij}/\partial a_k$: $k = i$, $k = j$, and $k \neq i, j$. Hence, we can write

$$\frac{d\rho_{ij}}{dp_i} = \frac{\partial\rho_{ij}}{\partial p_i} + \frac{\partial a_i}{\partial p_i}\left(\rho_{ii}\frac{\partial\rho_{ij}}{\partial a_i} + \rho_{ij}\frac{\partial\rho_{ij}}{\partial a_j} + \sum_{k\neq i,j}\rho_{ik}\frac{\partial\rho_{ij}}{\partial a_k}\right).$$

Noting that $\rho_{ii} = 1$ by definition and using symmetry, we can simplify $d\rho_{ij}/dp_i$ as

$$\frac{d\rho_{ij}}{dp_i} = \rho'_{p_1} + \frac{\partial a_i}{\partial p_i}[\rho'_1 + \rho\rho'_2 + (N-2)\rho\rho'_3],$$

where $\partial\rho_{ij}/\partial p_i \equiv \rho'_{p_1}$, $\rho'_1 \equiv \partial\rho_{ij}/\partial a_i$, $\rho'_2 \equiv \partial\rho_{ij}/\partial a_j$, and $\rho'_3 \equiv \partial\rho_{ij}/\partial a_k$ for $k \neq i,j$. Using this logic, we can rewrite the following term in expression (A29) as

$$\begin{aligned}
\sum_j \frac{d\rho_{ij}}{dp_R}\frac{\partial u_i}{\partial a_j} &= \sum_j \left(\frac{d\rho_{ij}}{dp_i} + \frac{d\rho_{ij}}{dp_j} + \sum_{k\neq i,j}\frac{d\rho_{ij}}{dp_k}\right)\frac{\partial u_i}{\partial a_j} \\
&= \sum_{j\neq i} \left(\frac{d\rho_{ij}}{dp_i} + \frac{d\rho_{ij}}{dp_j} + \sum_{k\neq i,j}\frac{d\rho_{ij}}{dp_k}\right)\frac{\partial u_i}{\partial a_j} \\
&= \sum_{j\neq i} b\Bigg(\rho'_{p_1} + \frac{\partial a_i}{\partial p_i}[\rho'_1 + \rho\rho'_2 + (N-2)\rho\rho'_3] + \rho'_{p_2} + \frac{\partial a_j}{\partial p_j}[\rho\rho'_1 + \rho'_2 + (N-2)\rho\rho'_3] \\
&\qquad + \sum_{k\neq i,j}\rho'_{p_3} + \frac{\partial a_k}{\partial p_k}\{\rho\rho'_1 + \rho\rho'_2 + [1 + (N-3)\rho]\rho'_3\}\Bigg) \\
&= b(N-1)\left\{\rho'_{p_1} + \rho'_{p_2} + (N-2)\rho'_{p_3} + \frac{\partial a_i}{\partial p_i}[1 + (N-1)\rho][\rho'_1 + \rho'_2 + (N-2)\rho'_3]\right\},
\end{aligned}$$

where derivatives of $\rho_{ii} = 1$ vanish in the second line, $\partial\rho_{ij}/\partial p_j \equiv \rho'_{p_2}$, and $\partial\rho_{ij}/\partial p_k \equiv \rho'_{p_3}$ for $k \neq i,j$. Completing the other terms in expression (A29) in the same manner and performing some algebra, one can write the CS condition for a monomorphic population as

$$\begin{aligned}
&\left[\frac{\partial^2 a_i}{\partial p_i^2} + (N-1)\frac{\partial^2 a_i}{\partial p_i \partial p_j}\right]\big(b\rho(N-1) - c + r(N-1)\{b[\rho(N-2)+1] - \rho c\}\big) \\
&+ \frac{\partial a_i}{\partial p_i}\Bigg[\left\{\frac{\partial a_i}{\partial p_i}[(N-1)\rho + 1][(N-2)\rho'_3 + \rho'_2 + \rho'_1] + (N-2)\rho'_{p_3} + \rho'_{p_2} + \rho'_{p_1}\right\} \\
&\qquad \times (N-1)\{b + r[(N-2)b - c]\} \\
&\qquad + \frac{\partial a_i}{\partial p_i}[(N-1)\rho + 1] \\
&\qquad \times\big((N-1)[(N-2)b'_3 + b'_2 + b'_1]\{\rho + r[(N-2)\rho + 1]\} \\
&\qquad\quad + [(N-1)b'_1 - c'][1 + (N-1)r\rho]\big)\Bigg] < 0.
\end{aligned} \tag{A30}$$

The following definitions, with $j \neq i$ and $k \neq i,j$, were used in expression (A30):

$$b'_1 \equiv \frac{\partial^2 u_i}{\partial a_i \partial a_j},$$

$$b'_2 \equiv \frac{\partial^2 u_i}{\partial a_j^2},$$

$$b'_3 \equiv \frac{\partial^2 u_i}{\partial a_k \partial a_j},$$

$$c' \equiv -\frac{\partial^2 u_i}{\partial a_i^2}.$$

A5. Condition for Decreasing *b*/*c*

When the benefits-to-costs ratio b/c decreases monotonically as a function of action at a monomorphic behavioral equilibrium a^*, there is a unique a^* that satisfies the first-order ES condition in equation (5). This also means that lower levels of investment in cooperation given by smaller values of a^* correspond uniformly to higher values of b/c, compared to higher levels of cooperation. Recall that $b = \partial u_i/\partial a_j$ and $c = -\partial u_i/\partial a_i$, where the partial derivatives are evaluated at a monomorphic behavioral equilibrium $a_1^* = \ldots = a_N^* = a^*$. The derivative of b/c evaluated at a^* is

$$\frac{\partial}{\partial a^*}\left(\frac{b}{c}\right) = \left(\frac{\partial u_i}{\partial a_i}\sum_{k=1}^{N}\frac{\partial^2 u_i}{\partial a_j \partial a_k} - \frac{\partial u_i}{\partial a_j}\sum_{k=1}^{N}\frac{\partial^2 u_i}{\partial a_i \partial a_k}\right)\left(\frac{\partial u_i}{a_i}\right)^{-2}. \tag{A31}$$

Since we evaluate all the partial derivatives assuming that the behavioral equilibrium is monomorphic, we can use symmetry, the notation of section A4 for the partial derivatives of u_i, and equation (A31) to write $\partial/\partial a^*(b/c) < 0$ as

$$b[(N-1)b_1' - c'] + c[b_1' + b_2' + (N-2)b_3'] < 0. \tag{A32}$$

It is common to assume that marginal benefits decrease with individual actions $b_2' < 0$ and $b_3' < 0$ and that marginal costs increase with individual actions, $c' > 0$ (Akçay et al. 2009). Given these assumptions, a sufficient condition for $\partial/\partial a^*(b/c) < 0$ is that

$$b_1' < \min\left(\frac{c'}{N-1}, -b_2' - (N-2)b_3'\right). \tag{A33}$$

The actions of the focal individual and its partners produce payoffs in a complementary or positively synergistic way when $b_1' > 0$, and the actions produce payoffs in a substitutable or negatively synergistic way when $b_1' < 0$. Thus, the sufficient condition for $\partial/\partial a^*(b/c) < 0$ requires that complementarity not be too high when compared to the rate of decrease in marginal benefits and the rate of increase in marginal costs.

A6. Characterizing the Group-Optimal Outcome

In a monomorphic population, the behavioral-equilibrium outcome will be symmetric, that is, $a_i^* = a^*$ for all i. Hence, the group-optimal outcome is the symmetric outcome that maximizes the payoff u_i. Formally, the following has to hold at the group-optimal outcome:

$$\sum_j \frac{\partial u_i}{\partial a_j}\bigg|_{a_1 = a_2 = \ldots = a_N = a^*} = 0. \tag{A34}$$

In terms of b and c, condition (A34) implies

$$\frac{b}{c} = \frac{1}{N-1}, \tag{A35}$$

which is equation (9).

A7. Group-Optimal Objective Function Traits

Suppose that the objective function is given by a general functional form f of the public good $\mathcal{B}(a_1, \ldots, a_N)$ and private cost $\mathcal{C}(a_i)$, that is,

$$x_i = (a_1, \ldots, a_N) = f(\mathcal{B}, \mathcal{C}). \tag{A36}$$

We know that at the group-optimal outcome $\rho = 1$, so we substitute the objective function in equation (A36) into equation (13) for ρ and set equation (13) equal to 1 to obtain

$$\frac{\partial f}{\partial \mathcal{C}}\frac{d\mathcal{C}}{da_i^2} + \left[(N-1)\frac{\partial^2 \mathcal{B}}{\partial a_j \partial a_i} + \frac{\partial^2 \mathcal{B}}{\partial a_i^2}\right]\frac{\partial f}{\partial \mathcal{B}} + (N+1)\frac{\partial \mathcal{B}}{\partial a_i}\frac{d\mathcal{C}}{da_i}\frac{\partial^2 f}{\partial \mathcal{B}\partial \mathcal{C}} + N\left(\frac{\partial \mathcal{B}}{\partial a_i}\right)^2\frac{\partial^2 f}{\partial \mathcal{B}^2} = 0. \tag{A37}$$

Here, we have again used the symmetry assumption, which implies $\partial^2\mathcal{B}/\partial a_j \partial a_i = \partial^2\mathcal{B}/\partial a_k \partial a_i$ for all $k, j \neq i$. Furthermore, we also know that this behavioral outcome satisfies equation (12), which yields

$$\frac{\partial f}{\partial \mathcal{B}}\frac{\partial \mathcal{B}}{\partial a_i} + \frac{\partial f}{\partial \mathcal{C}}\frac{d\mathcal{C}}{da_i} = 0. \tag{A38}$$

Substituting for f the expression on the right-hand side of equation (14) into equations (A37) and (A38) and solving for θ and ϕ, we obtain

$$\theta = \frac{(d\mathcal{C}/da_i)\Psi - \mathcal{C}\Omega}{(\partial\mathcal{B}/\partial a_i)\Psi + \mathcal{B}\Omega},$$

$$\phi = \frac{\Omega}{(\partial\mathcal{B}/\partial a_i)\Psi + \mathcal{B}\Omega}, \tag{A39}$$

where

$$\Psi = (N+1)\frac{d\mathcal{C}}{da_i}\frac{\partial \mathcal{B}}{\partial a_i},$$

$$\Omega = \frac{d^2\mathcal{C}}{da_i^2}\frac{\partial \mathcal{B}}{\partial a_i} - \frac{d\mathcal{C}}{da_i}\left[(N-1)\frac{\partial^2\mathcal{B}}{\partial a_i \partial a_j} + \frac{\partial^2\mathcal{B}}{\partial a_i^2}\right].$$

When we evaluate the right-hand sides of equations (A39) at the group-optimal outcome defined by $b/c = 1/(N-1)$, we find the θ and ϕ traits that make the group-optimal outcome a behavioral equilibrium and also satisfy the first-order ES condition ($\rho = 1$ at the group-optimal outcome).

A8. Effects of Group Size on the Evolution of Cooperation in Public-Goods Dilemmas

Group size N is important in determining the evolutionarily stable benefits-to-costs ratio. When the ratio of benefits to costs increases with increasing group size N, it will be harder to evolve increased levels of cooperation in larger groups than in smaller groups, and vice versa when the benefits-to-costs ratio is decreasing with N. In order to correct for the effect of group size on the total benefit individuals receive from the social interaction, we define $\hat{b} = (N-1)b$, which is the total benefit accruing to a single individual from all other individuals in the group. In terms of $\hat{b}$, the evolutionary-stability condition in equation (5) can be rewritten as

$$\frac{\hat{b}}{c} = \frac{1 + r\rho(N-1)}{r + \rho + (N-2)r\rho}. \tag{A40}$$

Condition (A40) simplifies to $\hat{b}/c = 1$ when $r = 1$ or $\rho = 1$ and the outcome is group optimal. If there is either no behavioral responsiveness, $\rho = 0$, or no effect of relatedness, $r = 0$, then condition (40) becomes analogous to the result from Hamilton's rule: $\hat{b}/c = 1/r$ for $\rho = 0$ and $\hat{b}/c = 1/\rho$ for $r = 0$. As expected in all of these limits, the evolutionarily stable $\hat{b}/c$ fraction is independent of group size N. Assuming that r and ρ can be held constant as N changes, the derivative of condition (A40) with respect to N is proportional to $-r(1-r)\rho(1-\rho)$, which is always negative. This suggests that larger groups tend to be more cooperative than smaller groups when r and ρ are independent of N. From condition (A40), the $\hat{b}/c$ ratio becomes, for large N,

$$1 + \frac{(1-r)(1-\rho)}{r\rho N},$$

which indicates that for large group sizes $\hat{b}/c$ is close to the group-optimal outcome and is insensitive to both r and ρ (so long as $r \gg 0$ and $\rho \gg 0$).

However, both the response coefficient ρ and the relatedness r are likely to be decreasing functions of N. Responsiveness likely decreases with N as coordination becomes more difficult in larger groups. Relatedness decreases in general with increases in population size n and will decrease with group size N so long as N is correlated with population size. This suggests that the evolutionarily stable value of $\hat{b}/c$ will usually increase with

N and that larger groups will tend to be less cooperative than smaller ones. Nonetheless, the evolutionarily stable $\hat{b}/c$ can decrease with N when

$$-\frac{1+(N-1)\rho}{r\rho(1-r)}r'n'-\frac{1+(N-1)r}{r\rho(1-\rho)}\rho'<1, \tag{A41}$$

where n' and ρ' are derivatives with respect to N and r' is a derivative with respect to n. Both fractions in inequality (A41) are positive, while r' and ρ' are negative; hence, the left-hand side is a positive number. Inequality (A41) can be satisfied if r' and ρ' have sufficiently small absolute values and the denominators in both fractions are not too small. In other words, the evolutionarily stable ratio of benefits to costs can decrease with group size N when both r and ρ are only weakly decreasing in N and neither r nor ρ is close to 0 or 1. Thus, if either relatedness or behavioral responsiveness is too low or too high, smaller groups will tend to be more cooperative than larger ones, while moderately related populations with moderate levels of behavioral responsiveness may result in larger groups being more cooperative than smaller ones.

A9. Mapping between Indirect Genetics Effects and Behavioral Responses

The current framework for modeling behavioral responses in a structured population uses a response coefficient ρ that is defined as the effect of a change in the action of a focal individual on the action of one of its social partners (Akçay et al. 2009). In this framework, actions are quantities defined on a behavioral timescale and are what individuals actually do in the interaction. Which actions an individual chooses to perform are affected by its motivational trait and the motivational traits of its social partners. In models that use indirect genetics effects (IGEs) to study social interactions, a regression coefficient ψ is used to measure directly the effect of the phenotype (such as the behavior) of a social partner on the phenotype of a focal individual (Moore et al. 1997; Wolf et al. 1999; McGlothlin et al. 2010). Since both the current framework and the IGE framework as implemented by McGlothlin et al. (2010) use the Price equation to measure the evolutionary change in behavioral phenotype, we can create a mapping between responsiveness (ρ) and the strength of IGEs (ψ) by comparing the conditions required for an increase in the mean phenotype.

Since our framework considers selection on a single trait, we present single-trait versions of the equations from McGlothlin et al. (2010). We begin with the phenotype of a focal individual, equation (B5) from McGlothlin et al. (2010),

$$z=\hat{\psi}\{(a+e)[1-(N-2)\psi]+(N-1)\psi(a'+e')\}, \tag{A42}$$

where a is the additive genetic value, e is an environmental effect, N is group size, and $\hat{\psi}=[1-(N-2)\psi-(N-1)\psi^2]^{-1}$. A prime in equation (A42) denotes values for social partners, and the overbar represents an average over the social partners of the focal individual. The average phenotype z in the group of the focal individual (excluding the focal individual) is given by

$$z=\hat{\psi}[(a'+e')+\psi(a+e)]. \tag{A43}$$

The evolutionary change in mean phenotype is given by equation (17) of McGlothlin et al. (2010), which in the current notation is

$$\Delta z=\mathrm{Cov}(A,z)\beta_{\mathrm{N}}+\mathrm{Cov}(A,z)\beta_{\mathrm{S}}, \tag{A44}$$

where $A=\tilde{\psi}a$ is the total breeding value from equation (B7) of McGlothlin et al. (2010) and $\tilde{\psi}=[1-(N-$

$1)\psi]^{-1}$. In equation (A44), β_N is the nonsocial-selection gradient and β_S is the social-selection gradient. Substituting equations (A42) and (A43) into equation (A44) yields

$$\begin{aligned}\Delta z &= \hat{\psi}\tilde{\psi}\big[\mathrm{Cov}\big(a,(a+e)[1-(N-2)\psi]+(N-1)\psi(a'+e')\big)\beta_N \\ &\quad + (N-1)\,\mathrm{Cov}\big(a,(a'+e')+\psi(a+e)\big)\beta_N\big)\big] \\ &= \hat{\psi}\tilde{\psi}\big[\mathrm{Cov}\big(a,(a+e)\{[1-(N-2)\psi]\beta_N+(N-1)\psi\beta_S\}\big) \\ &\quad + (N-1)\,\mathrm{Cov}\big(a,(a'+e')(\psi\beta_N+\beta_S)\big)\big] \\ &= \hat{\psi}\tilde{\psi}\,\mathrm{Var}\,(G)\{[1-(N-2)\psi]\beta_N+(N-1)\psi\beta_S+r(N-1)(\psi\beta_N+\beta_S)\},\end{aligned} \tag{A45}$$

which is equation (18) in McGlothlin et al. (2010). If we use the mapping of ψ to ρ given by

$$\rho = \frac{\psi}{1-(N-2)\psi}, \tag{A46}$$

then we can factor out $1-(N-2)\psi$ from equation (A45) to obtain

$$\Delta z = \frac{\hat{\psi}\tilde{\psi}\,\mathrm{Var}\,(G)}{1-(N-2)\psi}\left[\beta_N+(N-1)\rho\beta_S+r(N-1)\left(\rho\beta_N+\frac{\beta_S}{1-(N-2)\psi}\right)\right]. \tag{A47}$$

But from our mapping in equation (A46),

$$\psi = \frac{\rho}{1+\rho(N-2)}, \tag{A48}$$

so

$$\frac{1}{1-(N-2)\psi} = \frac{\rho}{\psi} = 1+\rho(N-2), \tag{A49}$$

and equation (A47) becomes

$$\Delta z = \frac{\hat{\psi}\tilde{\psi}\,\mathrm{Var}\,(G)}{1-(N-2)\psi}\big(\beta_S\rho(N-1)+\beta_N+r(N-1)\{\beta_S[1+\rho(N-2)]+\rho\beta_N\}\big). \tag{A50}$$

Given the mapping in equation (A46) and that $-1/(N-1)<\rho<1$, from appendix A3, ψ must obey the inequality

$$-1<\psi<\frac{1}{N-1}, \tag{A51}$$

which was suggested by McGlothlin et al. (2010) on the basis of either of $\hat{\psi}$ or $\tilde{\psi}$ having a denominator of 0. We can show that $\hat{\psi}\tilde{\psi}/[1-(N-2)\psi]>0$, using the inequality in equation (A51), which means that we can compare the terms inside the outermost parentheses of equation (A50) to the terms in the outermost parentheses in the equation for $\Delta\overline{G}$, equation (4). If we map the nonsocial-selection gradient of McGlothlin et al. (2010) to $-c$, $\beta_N=-c$, and the social-selection gradient to b, $\beta_S=b$, then equation (A50) is proportional to equation (4); this proves that the mapping of ψ to ρ in equation (A46) is the right mapping, since it yields identical conditions from the two frameworks for when the mean phenotype increases as a result of selection.

In fact, the reason that ψ in the IGE model of McGlothlin et al. (2010) does not map exactly onto ρ is that the McGlothlin et al. (2010) model (and other models of IGEs, e.g., Moore et al. 1997; Wolf et al. 1999; Wolf and Moore 2010) assumes that the phenotype of the focal individual feeds back to affect the phenotypes of the social partners. Without this feedback or reciprocal effect, ψ maps directly onto ρ. We can show this simply by writing expressions for the focal phenotype z and average phenotype z and then calculating Δz again using equation (A44). The phenotypes without feedback are

$$z = a+e+(N-1)\psi(a'+e') \tag{A52}$$

and

$$z = (a' + e')[1 + (N - 2)\psi] + \psi(a + e) \tag{A53}$$

(Moore et al. 1997; McGlothlin and Brodie 2009), and the total breeding value $A = a[1 + (N - 1)\psi]$. Plugging equations (A52) and (A53) and A into equation (A44) and rearranging yields

$$\Delta z = \mathrm{Var}(G)[1 + (N - 1)\psi]\big(\beta_S\psi(N - 1) + \beta_N + r(N - 1)\{\beta_S[1 + (N - 2)\psi] + \psi\beta_N\}\big),$$

which leads to equation (5) when $\beta_N = -c$ and $\beta_S = b$. Thus, ψ in this IGE specification yields the same conditions for an increase in mean phenotype as ρ.

Finally, we can also relate ρ to the variances and covariance of direct and social breeding values used in the variance-component IGE models of Griffing (1967, 1981*a*, 1981*b*) and Bijma and colleagues (Bijma et al. 2007; Bijma and Wade 2008). For example, the equation for the response to selection in the $r \neq 0$ and $g \neq 0$ elements of table 4 in Bijma and Wade (2008) is equivalent to equation (4) when the direct-selection coefficient $\beta_{W_{D},P}$ of Bijma and Wade (2008) is $-c$, the social-selection coefficient $\beta_{W_{S},P}$ is b, and the response coefficient is

$$\rho = \frac{\sigma_{A_{DS}} + (N - 1)\sigma^2_{A_S}}{\sigma^2_{A_D} + (N - 1)\sigma_{A_{DS}}}, \tag{A54}$$

where $\sigma^2_{A_D}$ and $\sigma^2_{A_S}$ are the variances of the direct and social breeding values, respectively, and $\sigma_{A_{DS}}$ is the covariance of direct and social breeding values.

Literature Cited Only in the Appendix

McGlothlin, J. W., and E. D. Brodie III. 2009. How to measure indirect genetic effects: the congruence of trait-based and variance-partitioning approaches. Evolution 63:1785–1795.

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 13:814–825.

Taylor, P. D., and A. J. Irwin. 2000. Overlapping generations can promote altruistic behavior. Evolution 54: 1135–1141.

West, S. A., I. Pen, and A. S. Griffin. 2002. Cooperation and competition between relatives. Science 296:72–75.

Wolf, J. B., and A. J. Moore. 2010. Interacting phenotypes and indirect genetic effects. Pages 225–245 *in* D. F. Westneat and C. W. Fox, eds. Evolutionary behavioral ecology. Oxford University Press, New York.