A Synthesis of Game Theory and Quantitative Genetic Models of Social Evolution

Joel W. McGlothlin[1,*], Erol Akçay[2], Edmund D. Brodie III[3], Allen J. Moore[4], and Jeremy Van Cleve[5]

[1]Department of Biological Sciences, Virginia Tech
[*]corresponding author, email: joelmcg@vt.edu
[2]Department of Biology, University of Pennsylvania
[3]Department of Biology, University of Virginia
[4]Department of Entomology, University of Georgia
[5]Department of Biology, University of Kentucky

Abstract

Two popular approaches for modeling social evolution, evolutionary game theory and quantitative genetics, ask complementary questions but are rarely integrated. Game theory focuses on evolutionary outcomes, with models solving for evolutionarily stable equilibria, whereas quantitative genetics provides insight into evolutionary processes, with models predicting short-term responses to selection. Here we draw parallels between evolutionary game theory and interacting phenotypes theory, which is a quantitative genetic framework for understanding social evolution. First, we show how any evolutionary game may be translated into two quantitative genetic selection gradients, nonsocial and social selection, which may be used to predict evolutionary change from a single round of the game. We show that synergistic fitness effects may alter predicted selection gradients, causing changes in magnitude and sign as the population mean evolves. Second, we show how evolutionary games involving plastic behavioral responses to partners can be modeled using indirect genetic effects, which describe how trait expression changes in response to genes in the social environment. We demonstrate that repeated social interactions in models of reciprocity generate indirect effects and conversely, that estimates of parameters from indirect genetic effect models may be used to predict the evolution of reciprocity. We argue that a pluralistic view incorporating both theoretical approaches will benefit empiricists and theorists studying social evolution. We advocate the measurement of social selection and indirect genetic effects in natural populations to test the predictions from game theory, and in turn, the use of game theory models to aid in the interpretation of quantitative genetic estimates.

Key words: evolutionary game theory, hawk-dove game, indirect genetic effects, interacting phenotypes, reciprocal altruism, social selection

Introduction

Social interactions are ubiquitous in nature (Frank 2007). In nearly every species, conspecifics interact in some way during their life cycle in contexts such as territoriality, mating, and parental care, and these social interactions can have profound implications for evolution (West-Eberhard 1979, 1983; Moore et al. 1997; Frank 1998). Darwin (1859) realized the importance of social context for evolution early on, and proposed the mechanisms of sexual selection and family-level selection to account for elaborate courtship displays and non-reproductive castes in social insects, respectively. However, when theoretical population genetics began to develop in the twentieth century to provide a mathematical basis for Darwin's theory, it was slow to incorporate social evolution. Stalwarts of the Modern Synthesis like Fisher (1930) and Haldane (1932) described some scenarios involving social evolution qualitatively but did not offer a mathematical analysis (Karlin and Matessi 1983).

In the 1960's, behavior began to be viewed separately given its unique property as a both a target and agent of selection, attracting the attention of formal mathematical theory. Hamilton's (1963; 1964a; 1964b) seminal work, which showed how genetic relatedness and social fitness effects combine to predict the evolution of social traits, catalyzed the development of the mathematical theory of social evolution and the fields of behavioral ecology and sociobiology in the decades to follow. Hamilton's early models of the evolution of altruistic behavior were explicitly genetic, and his concept of inclusive fitness involved viewing selection through what was later termed the "gene's eye view" (Dawkins 1976). However, later developments in social evolution theory often ignored genetics and relied on the "phenotypic gambit" (Grafen 1984), the assumption that the approach to phenotypic evolutionary optima is usually unconstrained by genetics (Eshel 1996). The most influential theoretical approach to phenotypic social evolution was evolutionary game theory and its central concept, the evolutionarily stable strategy (ESS) (Maynard Smith and Price 1973; Maynard Smith 1982; Dugatkin and Reeve 1998; Hofbauer and Sigmund 1998; McElreath and Boyd 2007; McNamara and Leimar 2020). Evolutionary game theory offered a powerful approach to predict evolutionary outcomes when individuals interact, particularly in the presence of frequency-dependent selection, was widely adopted by behavioral ecologists (Krebs and Davies 1978) and continues to be influential in behavioral research (McNamara and Leimar 2020).

In the 1980s and 1990s, a body of theory incorporating explicit genetic considerations into social evolution models began to be developed. Lande (1980, 1981) and Kirkpatrick (1982) both formalized Fisher's model of sexual selection and showed that the genetic correlations that arise between sexually selected traits and mating preferences can sometimes drive runaway evolution. Cheverud (1984), Lynch (1987), and Kirkpatrick and Lande (1989) incorporated maternal effects, which had long been studied by quantitative geneticists (Dickerson 1947; Willham 1963, 1972; Falconer 1965), into general evolutionary models. Echoing earlier work by Griffing (1967, 1977, 1981), Moore et al. (1997) extended maternal effects models to accommodate generalized social interactions between individuals, including unrelated individuals. Such "interacting phenotype" models depend on "indirect genetic effects" (IGEs), the modification of an individual's phenotypic expression in response to genetically variable traits in a social interactant. IGEs, which are

analogous to genetic maternal effects but which arise between any interacting conspecifics, may drastically alter responses to selection (Moore et al. 1997; Wolf et al. 1998; McGlothlin et al. 2010).

Another key aspect of quantitative genetic models of social evolution is social selection, which is caused when one individual's traits affect the fitness of another and is thus also amenable to quantitative genetic modeling (Wolf et al. 1999). The term "social selection" has been used in different ways by many authors (West-Eberhard 1979, 1983, 2014; Roughgarden et al. 2006; Roughgarden 2012; Lyon and Montgomerie 2012), focusing on different behavioral mechanisms and contexts (e.g., competition for social partners versus negotiation of social dynamics). Common to all conceptions of social selection is that the fitness of a focal individual is a function of its social environment, which includes the traits of its partners. In quantitative genetics, social selection can be modeled as a selection gradient measuring the fitness effect of social traits on the fitness of a focal individual (Wolf et al. 1999), which is equivalent to neighbor-modulated or group-level selection (Queller 1992; Bijma et al. 2007). Regardless of how it is represented, social selection may lead to evolutionary change when related individuals interact or in the presence of IGEs (Bijma and Wade 2008; McGlothlin et al. 2010; Akçay and Van Cleve 2012).

Despite the long tradition of studying social evolution via evolutionary game theory and quantitative genetics, the two frameworks have remained mostly separate, at least in practice. This divide has persisted in part because the goals of the two approaches differ slightly. While game theory models usually focus on solving for an equilibrium ESS (Maynard Smith 1982; McNamara and Leimar 2020), the goal of a quantitative genetic model of social evolution is usually to develop predictions for short-term evolutionary response to selection (Lande 1980; Moore et al. 1997). Yet this apparent difference masks a deep complementarity. Solving for an ESS requires writing equations for short-term evolutionary change and calculating what phenotypes can resist it. Quantitative genetics is typically concerned with changes in trait values without regard to whether such change is near or far from an evolutionary equilibrium, but can also be used to predict which traits will be stable in the face of such change. Conversely, evolutionary game theory places emphasis on modeling the behavioral processes that result in differential fitness outcomes, while quantitative genetics remains largely agnostic about such processes. Thus, although evolutionary game theory and quantitative genetics usually differ in emphasis, they fundamentally describe the same processes. A synthesis of the two frameworks can provide general insight into social evolution by combining models that focus on predicting equilibria with those that describe the process required to reach such equilibria (Moore and Boake 1994).

Here we demonstrate parallels between evolutionary game theory and quantitative genetics that are not widely appreciated in social evolution. Our treatment complements earlier syntheses (Aoki 1983, 1984; Gomulkiewicz 1998; Lion 2018; Lehtonen 2018; Queller 1992, 2011; Abrams et al. 1993; Taylor 1996), but differs in that we aim to explicitly incorporate the parameters of ESS models into interacting phenotype theory and vice versa. First, we express fitness effects in a general evolutionary game in terms of quantitative genetic selection gradients, showing how payoff matrices relate to parameters that may be estimated in natural populations. Next, we consider how both short-term response to selection and long-term evolutionary outcomes may be affected by the presence of indirect genetic effects. We discuss two specific games, the prisoner's dilemma and the hawk-dove game, as examples. Finally, we discuss the relevance of the IGE

parameter ψ, which measures the strength and direction of phenotypic social interactions, for evolutionary games that include reciprocity (Trivers 1971; Axelrod and Hamilton 1981).

Game Theory and Social Selection

In this section, we connect fitness effects in evolutionary game theory to the two selection gradients in interacting phenotype theory, the nonsocial (or direct) selection gradient β_N, which measures the effects of an individual's traits on its own fitness, and the social (or indirect) selection gradient β_S, which measures the effects of the traits of others on an individual's fitness (Wolf et al. 1999). Throughout, we assume that individuals interact in dyads, but our results can easily be extended to larger groups. Our treatment builds upon previous work by Taylor (1996), Taylor and Frank (1996), Queller (2011), Van Cleve (2017), Araya-Ajoy et al. (2020), and others, but we provide additional insights into the estimation of parameters in wild populations.

In game theory, the fitness costs and benefits of an evolutionary game are described by the payoff matrix, which defines the fitness consequences of an individual employing (or "playing") a particular strategy in a given social context. The values in the payoff matrix can be written in the form $V(z|z')$, which represents the expected payoff of an individual playing strategy z against an individual (or group) playing strategy z'. In a simple two-player game where an individual may take one of two strategies, we can define z as a binary trait where one strategy is $z = 1$, which corresponds to a player using strategy "1" with probability one, and the second strategy is $z = 0$, which corresponds to a player using strategy "0" with probability one and strategy "1" with probability zero. A general payoff matrix that can accommodate a wide variety of two-player games can be written as

$$\begin{aligned} V(1|1) &= \pi_1 + \pi_2 + \pi_3 & V(1|0) &= \pi_1 \\ V(0|1) &= \pi_2 & V(0|0) &= 0. \end{aligned} \tag{1}$$

Here, the parameter π_1 represents the additive payoff to the focal individual of playing strategy $z = 1$ and π_2 represents the additive payoff of facing strategy $z' = 1$. Parameter π_3 represents nonlinearity or synergy, i.e., the additional payoff that accrues only when both players play strategy "1." The payoffs in Equation 1 are assumed to be normalized with respect to the payoff $V(0|0)$, which is set to zero. This decision is arbitrary; fitness payoffs may be written by normalizing with respect to any of the other cells of the matrix. In this way, this notation can encompass any two-player, two-strategy game (Queller 2011; Van Cleve 2017). The payoffs in Equation 1 can be used to write a general equation for the absolute fitness of an individual:

$$W(z, z') = W_0 + (\pi_1 + \pi_2 + \pi_3)zz' + \pi_1 z(1 - z') + \pi_2(1 - z)z' = W_0 + \pi_1 z + \pi_2 z' + \pi_3 zz', \tag{2}$$

where W_0 represents baseline fitness outside the context of the game. Using this fitness function, we can interpret z as before, a binary trait with values zero and one, or allow z to be a continuous trait with values within the interval $[0, 1]$ where z is the probability of using strategy "1" and $1 - z$ is the probability of using

strategy "0". Further, we can use the fitness function $W(z, z')$ to calculate all the possible ESSs of the game using standard methods (Maynard Smith 1982; Hofbauer and Sigmund 1998).

Like evolutionary game theory, interacting phenotype theory also calculates fitness in a social context but using the notation and tools of quantitative genetics. Specifically, the interacting phenotype theory fitness function allows one to partition total selection into individual and socially mediated components (Wolf et al. 1999). Quantitative genetic models often start with relative fitness, which is defined as absolute fitness divided by mean fitness, or $w = \frac{W}{\overline{W}}$. For a one-trait model, relative fitness is modeled as

$$w(z, z') = w_0 + \beta_N z + \beta_S z' + \varepsilon, \tag{3}$$

where w_0 is an intercept, z represents the phenotype of the focal individual, z' represents the phenotype of its social partner, and ε is a normally-distributed error term with an expectation of zero. The effect of the focal phenotype on fitness is by the nonsocial selection gradient (β_N), and the effect of the social phenotype is captured by the social selection (β_S). These selection gradients are partial regression coefficients and represent partial fitness effects, i.e., the effect of a given phenotype on fitness while holding the other constant. When group means are used in place of z', this partitioning is identical to the approach known as contextual analysis (Goodnight et al. 1992; Heisler and Damuth 1987), and when that group mean excludes the focal individual, it is identical to the neighbor model of group selection (Lehtonen 2020; Okasha 2004, 2006). Nonsocial and social selection gradients can be estimated using a simple modification of the multiple regression method of Lande and Arnold (1983) and thus provide a useful tool for studying social evolution in natural populations (Wolf et al. 1999; Formica et al. 2011).

Fitness is the common currency of the two bodies of theory and so provides a translation between them. We first assume that the payoff matrix (Equation 1), which we have defined in terms of discrete binary strategies, can be used to define a continuous surface of fitness effects defined by continuous traits z and z'. These could represent the "mixed strategies" of game theory, where z represents the probability of performing strategy 1. In this case, the payoff matrix (Equation 1) would represent the four corners of a bounded fitness surface. However, z and z' could also represent other continuous traits that take values beyond the interval $[0, 1]$. In the latter case, the four points in the payoff matrix now define a fitness function that may be used to calculate the fitness of individuals with any trait value.

The selection gradients in Equation 3 can now be calculated by differentiating Equation 2 with respect to focal and social phenotypes. Following Taylor and Frank (1996, see also Lande and Arnold 1983; Charlesworth 1990; Iwasa et al. 1991; Taper and Case 1992; Abrams et al. 1993; Taylor 1996), the regression coefficients can be expressed as

$$\beta_N = \frac{1}{\overline{W}} \mathrm{E}\left(\frac{\partial W}{\partial z}\right) \tag{4a}$$

and

$$\beta_S = \frac{1}{\overline{W}} \mathrm{E}\left(\frac{\partial W}{\partial z'}\right) \tag{4b}$$

when we assume that z and z' are jointly normally distributed. Equations 4 represent expectations of partial derivatives of the absolute fitness function divided by population mean fitness. These expectations may be approximated by evaluating derivatives at the population mean, or

$$\beta_N \approx \frac{1}{\overline{W}} \frac{\partial W}{\partial z}\bigg|_{z=z'=\bar{z}} \tag{5a}$$

and

$$\beta_S \approx \frac{1}{\overline{W}} \frac{\partial W}{\partial z'}\bigg|_{z=z'=\bar{z}}. \tag{5b}$$

These expressions are exact when $W(z, z')$ is at most quadratic in either z or z', as is the case in Equation 2, and they hold more generally whenever the phenotypic variance is small (Abrams et al. 1993) or when selection is weak (Taylor 1996).

Using Equations 2 and 5, we can calculate the nonsocial and social selection gradients in the two-player game defined by Equation 1 as

$$\beta_N = \frac{1}{\overline{W}}(\pi_1 + \pi_3 \bar{z}) \tag{6a}$$

and

$$\beta_S = \frac{1}{\overline{W}}(\pi_2 + \pi_3 \bar{z}). \tag{6b}$$

where the average fitness is

$$\overline{W} = W_0 + (\pi_1 + \pi_2)\bar{z} + \pi_3\big(\bar{z}^2 + \mathrm{Cov}(z, z')\big). \tag{7}$$

The covariance term in Equation 7, which will be nonzero when individuals are related, when individuals assort on the basis of phenotype, or when there are indirect genetic effects (i.e., when one individual's phenotype is adjusted in response to its social partner), is equal to the interactant covariance ($C^{ij'}$) of Wolf et al. (1999).

Equations (6-7) show that the nonsocial and social selection gradients from interacting phenotype theory are not necessarily constant across generations but rather should change as the population mean changes. In the absence of nonlinear payoffs ($\pi_3 = 0$), these changes will derive solely from changes in mean fitness (equation 7), which will not influence the signs of the selection gradients and will be small when selection is weak. However, in any evolutionary game with nonlinear payoffs ($\pi_3 \neq 0$), both nonsocial and social gradients may change in sign as well as magnitude (Figure 1).

When nonlinear fitness effects are absent or weak (i.e., when $|\pi_3| < \pi_1$ and $|\pi_3| < \pi_2$), the interpretation of nonsocial and social selection gradients does not differ from the interpretation of the additive terms in the payoff matrix. For example, a trait that represents a cost to self ($\pi_1 < 0$ in the payoff matrix will always result in a negative nonsocial selection gradient (Figure 1). However, when fitness effects are strongly

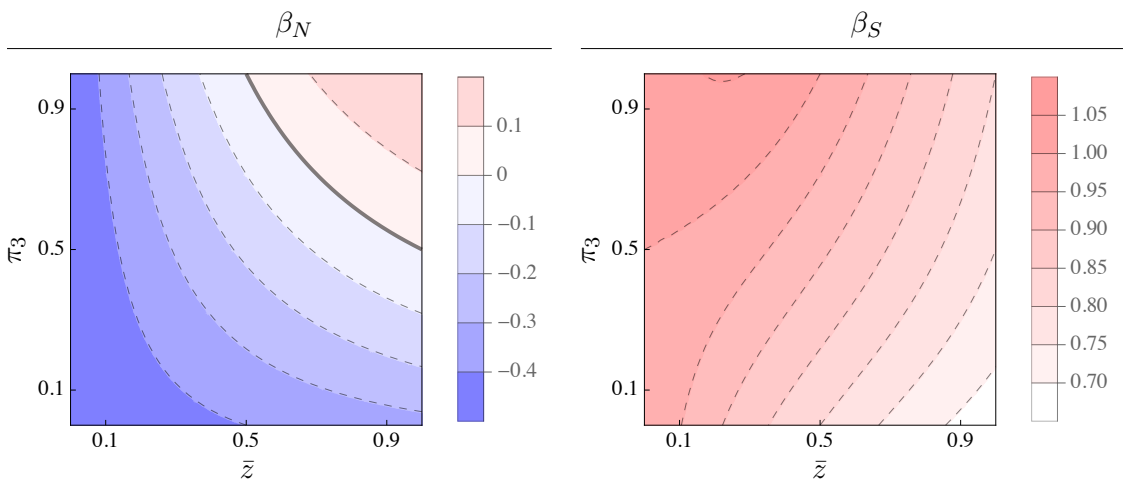

Figure 1: Dependence of nonsocial (β_N) and social (β_S) selection gradients on the trait mean ($\bar{z}$) and the nonlinear payoff π_3. Trait z is costly to self ($\pi_1 = -0.5$) but beneficial to others ($\pi_2 = 1.0$). Average fitness is calculated using Equation 7 where $W_0 = 1$ and the covariance term is set to zero. When synergistic effects are small ($\pi_3 < \pi_1$), the selection gradients change in magnitude with $\bar{z}$ but do not change in sign. When fitness effects are synergistic enough ($\pi_3 > \pi_1$), however, the nonsocial selection changes in sign once $\bar{z}$ is large enough. This effect results in the trait changing from representing a net cost to self to a net benefit to self.

synergistic (i.e., when $|\pi_3| > \pi_1$ or $|\pi_3| > \pi_2$), our interpretations of the two types of parameters may diverge. A strongly positive π_3 coupled with a negative π_1, for example, may result in a nonsocial selection gradient that is negative when the trait mean is low and positive when the trait mean is high (Figure 1). The interpretation here depends on perspective. A game theorist might view this trait as lacking "strategic dominance" (Binmore 2007) since the cost of switching strategies, $V(1|1) - V(0|1) = \pi_1 + \pi_3$ and $V(1|0) - V(0|0) = \pi_1$, have different signs, whereas a quantitative geneticist would note that the trait changes from representing a net cost to a net benefit depending on the trait mean.

Our results show that interacting phenotypes models potentially include two types of frequency dependence. The first occurs whenever there is social selection, i.e., whenever π_2 and β_S are nonzero. In this case, the fitness of a given phenotype depends upon the individual (or individuals in a group model) with which it interacts (Wolf et al. 1999). This contribution can be called "local frequency dependence" because it creates variance in fitness across pairs (or groups) of interactants existing in the same population at the same time. Local frequency dependence aligns with the concept of frequency dependence in game theory and population genetics, which is usually defined to include any type of socially dependent fitness effect (McNamara and Leimar 2020). Local frequency dependence should be present in any evolutionary game or social selection model. The second type occurs when the sign of selection gradients changes across generations with the trait mean as in Figure 1. This contribution can be called "global frequency dependence" because the total effect of selection on the trait, $\Delta\bar{z}$, depends upon the makeup of the whole population. Global frequency dependence aligns with the definition of frequency-dependent selection in quantitative genetics, i.e., a change in the direction of selection with a change in the phenotypic mean (Lande 1976). Global frequency dependence will arise only when nonlinear payoffs (π_3) exist.

Previous interacting phenotype models have obscured the contribution of nonlinear fitness effects, which has been cited as one of their primary limitations (Westneat 2012; Araya-Ajoy et al. 2020). As noted by Araya-Ajoy et al. (2020), one of the most important consequences of nonlinear fitness effects is that they can lead to evolutionary feedback. In other words, because the strength of selection may change with the trait mean in the presence of synergy, the rate of evolution of a social phenotype may potentially increase each generation. Although the inclusion of nonlinear effects does not change the predictions of short-term response to selection (Moore et al. 1997; Bijma and Wade 2008; McGlothlin et al. 2010), it is important to consider how nonlinear effects might change selection gradients when attempting to draw conclusions about social evolution over many generations.

An alternative approach (Queller 1985, 2011; Westneat 2012; Araya-Ajoy et al. 2020) partitions selection to include a separate gradient (β_I) that estimates the interactive (i.e., multiplicative) effect of focal and social traits (zz'). In this partitioning, the three selection gradients take on expectations that are independent of the mean, with $\beta_N = \frac{1}{W}\pi_1$, $\beta_S = \frac{1}{W}\pi_2$, and $\beta_I = \frac{1}{W}\pi_3$. However, we prefer the standard definitions of β_N and β_S in Equation 6. As we will show in the next section, using standard definitions allows the incorporation of fitness effects from any evolutionary game directly into equations for evolutionary response (Bijma and Wade 2008; McGlothlin et al. 2010). In addition, these definitions directly correspond to the selection gradients that would be measured in a social selection analysis of a natural population (Wolf et al. 1999).

With a sufficiently large dataset one could estimate nonlinear fitness effects in a natural population by expanding the Wolf et al. (1999) model to include quadratic terms (Lande and Arnold 1983; Phillips and Arnold 1989). Defining relative fitness as

$$w(z, z') = w_0 + \beta_N z + \beta_S z' + \frac{1}{2}\gamma_{NN}(z - \bar{z})^2 + \frac{1}{2}\gamma_{SS}(z' - \bar{z})^2 + \gamma_{NS}(z - \bar{z})(z' - \bar{z}) + \varepsilon \qquad (8)$$

has the advantage of estimating evolutionarily relevant selection gradients while estimating the total curvature of the fitness surface (γ_{ij} terms). Here, the nonsocial-social correlational selection gradient (γ_{NS}) captures the same fitness interaction as β_I (Westneat 2012; Araya-Ajoy et al. 2020), while the directional gradients include all fitness effects relevant to evolutionary response.

Response to Selection in Evolutionary Games

One advantage of defining selection gradients in terms of the payoff matrix is that it enables calculation of the predicted response to selection that arises from an evolutionary game. In general, the predicted response to selection for a socially influenced trait can be calculated from the Price (1970, 1972) equation using Equation 3 as

$$\Delta\bar{z} = \mathrm{Cov}(A, w) = \mathrm{Cov}(A, z)\beta_N + \mathrm{Cov}(A, z')\beta_S \qquad (9)$$

(Frank 1997; McGlothlin et al. 2010), where A is the total breeding value of trait z in an individual. This equation is useful because it partitions the relationship between breeding values and phenotypes (the two covariances) from the relationship between phenotypes and fitness (the two selection gradients). The two

covariance terms on the right-hand side of Equation 9 represent two types of genotype-phenotype relationships. The first describes the relationship between an individual's own genotype and phenotype, and the second describes whether an individual's genotype is associated with phenotypes found in its social environment.

To calculated these covariances, we must specify a genotype-phenotype model. One common model for the genetics of interacting phenotypes in quantitative genetics partitions a single phenotype z into the influence deriving from a focal individual's own additive genetic (a) and environmental (e) contributions and the effect of a social partner's phenotype (z'):

$$z = a + e + \psi z', \tag{10}$$

where ψ, which is a constant representing the importance of the social interaction for the expressed trait, theoretically ranges from -1 to 1 for a single-trait model (Moore et al. 1997). The coefficient ψ is a way to incorporate phenotypic responsiveness to the social environment into quantitative genetic models. Specifically, when ψ is nonzero, the expression of phenotype z depends in part on a response to the trait of its social partner (Moore et al. 1997). This can arise, for example, if the phenotype is behavioral and individuals observe and respond to each other's behavior in real time; ψ then represents this behavioral responsiveness (Akçay et al. 2009; Akçay and Van Cleve 2012). The model in Equation 10 may be used to develop expressions for both an individual's breeding value ($A = \frac{a}{1-\psi}$) and the covariances in Equation 9 (Moore et al. 1997; McGlothlin et al. 2010). Whenever z is heritable and $\psi \neq 0$, the total covariance between breeding values and phenotype will depend upon two sources: direct genetic effects caused by an individual's own genes and IGEs caused by the social partner's genes and mediated by ψ.

Because the social partner's trait is also influenced by genetic, environmental, and social components, Equation 10 can be expanded as

$$z = \frac{a + e + \psi(a' + e')}{1 - \psi^2}, \tag{11}$$

where the denominator now represents the degree of feedback between the two interactants (Moore et al. 1997). Equation 11 shows that phenotypic expression involves two genetic effects: a direct genetic effect proportional to a and an indirect genetic effect (IGE) proportional to a'. If relatives interact or pairs form nonrandomly, these effects may be correlated. We model the correlation between the breeding values of different individuals using a genetic relatedness coefficient r. The predicted response to selection in the presence of both relatedness and IGEs can be shown to be

$$\Delta\bar{z} = G\,\frac{(1 + r\psi)\beta_N + (r + \psi)\beta_S}{(1 - \psi)(1 - \psi^2)} \tag{12}$$

(McGlothlin et al. 2010; cf. Akçay and Van Cleve 2012), where G represents additive genetic variance.

An equally valid alternative to this trait-based model partitions the total breeding value (A) of an individual trait into direct and indirect components without reference to a specific phenotype of the social partner and uses variance components to predict the response to selection (Griffing 1967; Bijma et al. 2007; Bijma and Wade 2008; McGlothlin and Brodie 2009; Bijma 2014). While trait-based IGE models, such as

the one we develop here, decompose IGEs into effects of specific traits, variance-component models include terms that encompass all IGEs on a particular trait. We use a univariate trait-based model here to parallel game theory approaches, but this is easily generalizable to a multivariate scenario. Including effects of other traits on evolutionary response requires either a variance-component model (Bijma and Wade 2008) or a multivariate trait-based model (Moore et al. 1997).

Equation 12, when combined with our previous definitions of nonsocial and social selection gradients (Equation 6), allows the prediction of short-term evolutionary change from a wide variety of evolutionary games. In addition, the incorporation of both relatedness and phenotypic responsiveness into Equation 12 allows us to determine when these phenomena may affect both short-term evolutionary response and predicted evolutionary outcomes. In particular, ψ may be useful for modeling any game where strategies are allowed to change in response to the strategy of the opponent. Such phenotypic responsiveness is known variously as "(direct) reciprocity" (Trivers 1971; Axelrod and Hamilton 1981; Alexander 1985) or "negotiation" (McNamara et al. 1999). An alternative approach to the one we take here is to explicitly model how responsiveness affects the trait value over repeated rounds of interaction and incorporate those effects into the payoff matrix (Van Cleve and Akçay 2014; Van Cleve 2017). The variable used by Van Cleve and Akçay (2014) and Van Cleve (2017) to quantify responsiveness (ρ) is conceptually similar but not identical to ψ (Akçay and Van Cleve 2012). An advantage of our current approach is that it partitions the effect of selection on trait values from genetics, which is consistent with standard quantitative genetic theory and thus allows for the estimation of relevant parameters using quantitative genetic methods.

Equations (9-12) can also be used to predict the direction of evolution and evolutionary equilibria for a given model by setting $\Delta\bar{z}$ equal to zero and rearranging. For example, in the presence of both relatedness and IGEs, $\bar{z}$ should evolve in a positive direction when

$$\beta_N + \frac{r+\psi}{1+r\psi}\beta_S > 0 \tag{13a}$$

and should be at equilibrium when

$$\beta_N + \frac{r+\psi}{1+r\psi}\beta_S = 0 \tag{13b}$$

(McGlothlin et al. 2010, 2014; Akçay and Van Cleve 2012). In most cases, this approach should predict the ESS for a given evolutionary game (Taylor and Frank 1996). To explicitly illustrate how game theory and interacting phenotypes models can apply to the same behavioral scenarios, we analyze two simple evolutionary games using this approach below.

Example 1: Prisoner's Dilemma

One classic evolutionary game is the prisoner's dilemma (Rapoport and Chammah 1965), which is defined by the matrix

$$\begin{aligned} V(C|C) &= b - c + d & V(C|D) &= -c \\ V(D|C) &= b & V(D|D) &= 0. \end{aligned} \tag{14}$$

In this game, each player may either cooperate (C) with the other player, in which case it pays a fitness cost c, or defect (D), acting in its own interest and paying no fitness cost. Interacting with a cooperator leads to a fitness benefit b. The synergy term d is greater than zero when simultaneous cooperation yields additional fitness benefits. Most evolutionary games can be recovered from the prisoner's dilemma by varying the signs and relative magnitudes of b, c, and d (Van Cleve and Akçay 2014; Van Cleve 2017). Indeed, the payoff matrix represented in Equation 14 is identical to that in Equation 1 except for signs.

Using Equations 5, the nonsocial and social selection gradients in this game are

$$\beta_N = \frac{1}{\overline{W}}(-c + d\bar{z}) \tag{15a}$$

and

$$\beta_S = \frac{1}{\overline{W}}(b + d\bar{z}). \tag{15b}$$

As expected, nonsocial and social selection are proportional to the costs and benefits of cooperation in the absence of synergy ($d = 0$) and depend on the population mean in the presence of synergy ($d \neq 0$). In the absence of IGEs and relatedness, evolution depends only upon nonsocial selection, and cooperation will always be disfavored in the absence of positive synergy. Even with positive synergy, however, cooperation is unable to increase when $\bar{z} = 0$.

Adding IGEs and relatedness allows us to derive a version of Hamilton's rule that includes synergy. Cooperation will increase (i.e., $\Delta\bar{z} > 0$) when

$$-c + \frac{r + \psi}{1 + r\psi}b + \frac{(1 + r)(1 + \psi)}{1 + r\psi}d\bar{z} > 0. \tag{16}$$

As in previous versions of Hamilton's rule that did not explicitly incorporate synergy (McGlothlin et al. 2010, 2014), Equation 16 shows that relatedness and IGEs have completely symmetrical effects on the evolution of cooperation (Akçay and Van Cleve 2012; Van Cleve and Akçay 2014; Van Cleve 2017). The influence of fitness synergy on the evolution of altruism has been discussed at length elsewhere (e.g. Queller 1985). Two additional insights arise from including IGEs into this model, however. First, when $\psi > 0$, IGEs may facilitate the invasion of cooperation when $\bar{z}$ is close to zero by creating a phenotypic covariance between interactants and increasing the effect of the b term (i.e., the linear component of social selection).

Second, IGEs may act in concert with relatedness to enhance the effects of positive synergy as cooperation and $\bar{z}$ increase.

Example 2: The Hawk-Dove Game

We next consider the classic hawk-dove game, which involves two conspecifics competing over a resource (Maynard Smith and Price 1973; Maynard Smith 1982). When two doves meet, they divide the resource or decide the contest without aggression. When two hawks meet, they fight to determine the contest, with one hawk winning and the other paying a cost. When a hawk meets a dove, the dove flees and the hawk takes the resource. In this game, the payoffs can be expressed as

$$\begin{aligned} V(H|H) &= \frac{v-c}{2} & V(H|D) &= v \\ V(D|H) &= 0 & V(D|D) &= \frac{v}{2} \end{aligned} \tag{17}$$

where H represents the hawk strategy, D represents the dove strategy, v represents the value of the resource, and c represents the cost of fighting over that resource. Fitness can be expressed as a function by equating $z = 1$ with a pure hawk strategy and $z = 0$ with pure dove:

$$W = W_0 + \frac{v-c}{2}zz' + vz(1-z') + (1-z)(1-z')\frac{v}{2} = W_0 + \frac{v}{2}(1+z-z') - \frac{c}{2}zz' + \varepsilon \tag{18}$$

As before, we can translate this fitness function into selection gradients by taking partial derivatives of Equation 18, yielding

$$\beta_N = \frac{v - c\bar{z}}{2\overline{W}} \tag{19a}$$

and

$$\beta_S = \frac{-v - c\bar{z}}{2\overline{W}}. \tag{19b}$$

To predict evolutionary rates and solve for equilibria, we substitute Equations 19 into Equation 12. In the absence of IGEs and relatedness, the result reduces to

$$\Delta\bar{z} = G\beta_N = G\,\frac{v - c\bar{z}}{2\overline{W}}. \tag{20}$$

Solving for equilibrium,

$$\hat{z} = \frac{v}{c}, \tag{21}$$

which corresponds to the classic ESS for the continuous hawk-dove game (Maynard Smith 1982). Naturally this mixed-strategy equilibrium is stable only when $v < c$ (so that $\hat{z} < 1$). When $v \geq c$, $\hat{z}$ from Equation 21 is unstable, and $\bar{z}$ will evolve to its boundary condition at $\bar{z} = 1$, which is the ESS in the hawk-dove game when hawk is the dominant strategy.

Now imagine that the two individuals are allowed to adjust their behavior in response each other. Such negotiation can be modeled as an IGE. By substituting Equations 19 into Equation 12 and setting only relatedness to zero, we find a predicted response to selection of

$$\Delta\bar{z} = G\,\frac{v - c\bar{z} - \psi(v + c\bar{z})}{2\overline{W}(1-\psi)(1-\psi^2)}. \tag{22}$$

We can solve for the equilibrium phenotype by setting the left-hand side equal to zero and solving for $\bar{z}$:

$$\hat{z} = \frac{v(1-\psi)}{c(1+\psi)}. \tag{23}$$

Equation 23 shows that phenotypic adjustment changes the predicted equilibrium of a hawk-dove game. If $\psi > 0$, there is positive feedback, meaning that interactants will adjust their behavior based on what the other does, which results in lower equilibrium hawkishness than predicted in the phenotypic model. This result is equivalent to Grafen's (1979) result for relatedness, with ψ in place of r. In contrast, if $\psi < 0$, individuals become more hawkish when their opponent is more dovish, and vice versa. This results in a higher equilibrium hawkishness.

Indirect Genetic Effects and Reciprocity

Above, we have argued that IGEs may be used as a general way to model phenotypic adjustment in evolutionary games. Perhaps the most well-known type of phenotypic adjustment is reciprocity or reciprocal altruism, first discussed by Trivers (1971). A typical model of reciprocity involves sequential interactions between individuals where memory of one or more previous interactions determines the action taken in the current interaction. It has been argued that IGEs allow quantitative genetic models to incorporate reciprocity (Bleakley and Brodie 2009; McGlothlin et al. 2010), but this relationship has not been made mathematically.

In evolutionary game theory, reciprocity is often modeled using a repeated or iterated version of the prisoner's dilemma game (Axelrod and Hamilton 1981; McElreath and Boyd 2007). After meeting once and playing a prisoner's dilemma, two individuals play again with some probability. By allowing various repeated strategies to compete in a computer simulation, Axelrod and Hamilton (1981) showed that a strong strategy in this game is tit-for-tat, where each player mimics its partner's behavior from the previous round. A population that employs a tit-for-tat strategy can prevent the invasion of cheaters, because an individual that cheats on the first meeting is punished in later iterations.

This game appears to match a typical IGE model with one exception. The evolutionary game theory model assumes that interactions are repeated and sequential, while IGE models involve a single interaction with phenotypic feedback. To show the relationship between the two models, we can express the phenotypes for a pair of individuals that interact sequentially using a quantitative genetic model in which subsequent interactions are included as different traits. Suppose that in the first interaction, two individuals express a

phenotype that is unaffected by the other individual:

$$z_0 = a + e \qquad\qquad z_0' = a' + e'. \tag{24}$$

In Equations 24 and those that follow, the subscript refers to the number of times two individuals have previously met (i.e., z_0 and z_0' are phenotypes for two individuals that have not previously interacted). When two individuals interact again, each individual adjusts its phenotype based on what occurred in their last interaction. We use y to represent the fraction of the phenotype that is attributable to this adjustment, while $1-y$ is the fraction of the phenotype based on the individual's own genes and environment. A value of $y = 1$ would be equivalent to a perfect tit-for-tat response. For simplicity, we assume that y is a population parameter. However, the model could easily be expanded to allow each individual to have its own value of y. Thus, at the second interaction, the two phenotypes can be written as

$$z_1 = (1-y)(a+e) + yz_0' \qquad\qquad z_1' = (1-y)(a'+e') + yz_0. \tag{25}$$

Because all subsequent interactions follow the same form, this can be generalized as

$$z_t = (1-y)(a+e) + yz_{t-1}' \qquad\qquad z_t' = (1-y)(a'+e') + yz_{t-1}. \tag{26}$$

Now suppose that each additional interaction occurs with probability p. To make a direct comparison to the standard IGE model, we measure the phenotype as the average behavior across all repeated interactions between the two individuals. The average phenotype is the sum of values across all steps divided by the total number of interactions, $(1-p)^{-1}$, or

$$z_{\text{ave}} = (1-p)\sum_{t=0}^{\infty} p^t z_t. \tag{27}$$

In the Appendix, we show that Equation 27 can be expressed as

$$z_{\text{ave}} = \frac{a + e + py(a' + e')}{1 + py}. \tag{28}$$

The numerator in Equation 28 corresponds to the phenotypic definition from the standard IGE model (Equation 11), with the equivalency

$$\psi = py. \tag{29}$$

In other words, the the interaction effect ψ is equal to the probability of repeated interactions multiplied by the strength of reciprocal response. However, the sequential model differs slightly from the standard IGE model; namely, there is an additional factor of $(1-\psi)$ in the standard model. In Equation 28, as py approaches 1, the denominator approaches 2 and the numerator approaches the sum of the initial values of the two interactants. This shows that stronger, more likely interactions cause the phenotype to resemble the average of the initial values of the two interactants.

To derive an equation for the response to selection in this model, we take the population average of Equation 28,

$$\bar{z}_{\text{ave}} = \bar{a}, \tag{30}$$

which shows that unlike in the standard IGE model, the population mean equals the mean additive genetic value. This allows us to solve for a response to selection by using the additive genetic value as the breeding value in Equation 9, yielding

$$\Delta\bar{z}_{\text{ave}} = G\,\frac{(r+py)\beta_N + (1+rpy)\beta_S}{1+py}, \tag{31}$$

which, with $\psi = py$, differs from Equation 12 by a factor of $(1-py)^{-2}$.

Equation 31 shows that the predicted response to selection when repeated interactions occur will be proportional to the predicted response for a single interaction with simultaneous feedback; however, assuming $py > 0$, the rate of evolutionary change will be reduced. Thus, when we assume that interactions occur sequentially instead of simultaneously, likely a more realistic assumption, the runaway feedback effect that normally characterizes models with reciprocal IGEs (Moore et al. 1997; McGlothlin et al. 2010) disappears. The response to selection still depends on both direct genetic effects and IGEs in a similar way, but the rate of evolution no longer spirals to infinity as the strength of the IGE increases.

Despite the difference in predicted evolutionary rate between the two models, other conclusions depend only on the numerator and will not differ. This allows us to use estimates of ψ to make general predictions about the direction of evolution no matter what model the interactions follow. For example, the condition for the evolution of reciprocity with sequential interactions can be given as a form of Hamilton's rule:

$$\beta_N + \frac{r+py}{1+rpy}\beta_S > 0, \tag{32}$$

which with $\psi = py$ is identical to Equation 13a. When non-relatives interact, Equation 32 simplifies to

$$\beta_N + py\beta_S = 0, \tag{33}$$

which reproduces results from prior models of reciprocity and responsiveness (André and Day 2007; Akçay et al. 2009). Equations 32-33 indicate that, as expected, the evolution of reciprocity depends on py or approximately, ψ. The evolution of altruistic behavior is favored when repeated encounters with the same individual are more likely (higher p) and when the response more closely resembles tit-for-tat (higher y). This result corresponds to the standard result that the evolution of reciprocity via a tit-for-tat strategy is favored when subsequent encounters are more likely (McElreath and Boyd 2007). It is important to note, however, that our model assumes that the parameters p and y, and consequently, ψ, are fixed, and conclusions may differ when ψ is allowed to evolve (André and Day 2007; Akçay et al. 2009; Kazancioglu et al. 2012; André 2015). The assumption of fixed p and y, or ψ, allows prediction of short-term evolution.

Discussion

Our analysis highlights important connections between evolutionary game theory and quantitative genetic models of interacting phenotypes. Specifically, the fitness effects that arise from the payoff matrix in social games translate directly into the nonsocial and social selection gradients of interacting phenotypes theory from quantitative genetics. Various types of phenotypic modification can be modeled using the interaction coefficient ψ, which quantifies the strength of IGEs. These relationships make it straightforward to incorporate any type of interaction described by an evolutionary game into the predictive equations for evolutionary responses to selection. Thus, evolutionary models developed under the "phenotypic gambit" can be translated into explicitly genetic models simply by specifying the genotype-phenotype relationship typical of quantitative genetics, namely the partitioning of phenotype into direct and indirect genetic components and an environmental component.

The primary advantage of translating between game theory and quantitative genetics is that genetic models are typically formulated in terms of parameters that can be estimated in natural populations. For example, β_N and β_S can be estimated in natural populations using a modification of the standard regression method of Lande and Arnold (1983) that incorporates social phenotypes or group means as predictors (Heisler and Damuth 1987; Goodnight et al. 1992; Wolf et al. 1999). Studies that measure social selection have begun to accumulate (Formica et al. 2011; Farine and Sheldon 2015; Fisher and Pruitt 2019; Santostefano et al. 2020), and we hope this work encourages empiricists studying social phenotypes in natural populations, particularly those whose questions are motivated by evolutionary game theory, to estimate β_S for their traits of interest. Such estimates may be used to parameterize payoff matrices, providing strong tests of predictions from game theory. As we note above, correlational selection gradients may also be incorporated to test for the presence of synergistic interactions by using Equation 8.

We argue that the phenotypic modification captured in evolutionary games leads to IGEs, which may also be estimated in natural or laboratory populations. The parameters of variance-component models of IGEs, i.e., direct and indirect genetic (co)variance (Bijma et al. 2007; Bijma and Wade 2008), may be estimated using quantitative genetic animal models (Bijma 2010; Wilson et al. 2009, 2010, 2011). Methods also exist to estimate ψ, either by using inbred lines or test strains (Bleakley and Brodie 2009) or as functions of variance components (McGlothlin and Brodie 2009). Estimates of IGEs are increasingly common, particularly in populations of domestic animals (Wade et al. 2010), and a number of studies have estimated IGEs specifically for behavioral traits that have often been analyzed using game theory, such as aggression (Wilson et al. 2009, 2011; Saltz 2013; Alemu et al. 2014; Santostefano et al. 2017; Han et al. 2018; Lane et al. 2020). Despite this increased interest in IGEs, however, studies attempting to estimate ψ remain rare (but see Bleakley and Brodie 2009; Edenbrow et al. 2017). We hope that our results motivate empiricists to estimate this parameter more often, particularly when testing for the phenomenon of reciprocity.

Our results show that game theory holds lessons for quantitative geneticists as well. Nonsocial and social selection gradients represent net fitness effects of a phenotype. As such, they are often interpreted as quantitative genetic analogs of Hamilton's (1964a) costs and benefits. This equivalence may be accurate in some cases, but there are several scenarios in which the two diverge (McGlothlin et al. 2014; Hadfield

and Thomson 2017). As our results show, one such case arises when strong nonlinear fitness effects exist (see also Westneat 2012; Araya-Ajoy et al. 2020). We urge empiricists to exercise caution when mapping selection gradients to cost and benefit terms from evolutionary games, particularly when the correlational selection gradient γ_{NS} is nonzero. We note that selection gradients can always be correctly interpreted as the effective relative fitness cost or benefit of a measured phenotype in the current population, and they can be used to accurately predict short-term evolutionary response of socially influenced traits (Bijma and Wade 2008; McGlothlin et al. 2010) even when nonlinear effects are present.

The interaction coefficient ψ should also be interpreted with caution, particularly when predicting the rate of evolutionary change. Explicitly modeling phenotypic modification with sequential social interactions similar to prior models of direct reciprocity in the prisoner's dilemma shows that the runaway feedback effects predicted by previous IGE models (Moore et al. 1997; McGlothlin et al. 2010) often may be attenuated. However, our treatment shows that other conclusions of the general model hold true because the direction of evolutionary response does not change. In particular, the sequential interaction model leads to a version of Hamilton's rule identical to that of the standard model, indicating that ψ can be used to predict the evolution of reciprocity. In addition, methods to estimate ψ as a function of variance components are unaffected by the specific model of social interaction because feedback effects influence all variance components equally and thus cancel out (McGlothlin and Brodie 2009). Therefore, when it may be measured, ψ remains a useful estimate for the strength of reciprocity. Finally, ψ is likely to evolve (Chenoweth et al. 2010), which may also alter long-term evolutionary predictions (Akçay and Van Cleve 2012; Kazancioglu et al. 2012). A future contribution will address the conditions under which ψ will change in response to selection.

In conclusion, our results emphasize the benefits of unifying models of evolutionary outcomes and evolutionary process. Incorporating both types of theory leads to a holistic view of social evolution on microevolutionary and macroevolutionary scales. Further, both types of theory can contribute in different ways to the crucial feedback between theory and experiment. As we have argued, empirical quantitative genetics can estimate parameters to test and refine game theory models. At the same time, game theory can both provide conceptual context for the interpretation of theoretical and empirical quantitative genetic results and generate predictions about potential evolutionary outcomes that inform the design of microevolutionary studies of natural populations. We hope that our results will inspire further integration of game theory and quantitative genetics to lead to a richer understanding of the evolution of social phenotypes.

Acknowledgments

Many thanks to Kim Hughes and Anjanette Baker for their efforts organizing the 2020 AGA Presidential Symposium. We thank Yimen Araya-Ajoy and David Westneat for helpful conversations and two anonymous reviewers for comments on the manuscript. JVC acknowledges NSF CAREER Award DEB-1846260 for support.

Appendix

Here we will show how to obtain Equation 28 from Equation 27. First, we solve for the infinite sum, which we write as z_Σ. Substituting for the phenotype at $t = 0$, when only the individual's own genes and environment are important (Equation 24) yields

$$z_\Sigma = a + e + \sum_{t=1}^{\infty} p^t z_t. \tag{A.1}$$

Starting with $t = 1$ (Equation 26), the phenotype depends on a combination of an individual's own genes and environment (multiplied by the quantity $1 - y$) and a response to its social partner's previous phenotype (multiplied by y). Thus, the infinite series follows the form

$$z_\Sigma = a + e + p(1-y)(a+e) + pyz_0' + p^2(1-y)(a+e) + p^2yz_1' + \dots. \tag{A.2}$$

The infinite series in Equation A.2 is then

$$z_\Sigma = a + e + \frac{p(1-y)(a+e)}{1-p} + pyz_0' + p^2yz_1' + \dots. \tag{A.3}$$

Next, we expand the contribution of the social partner, closing the sum in a similar way:

$$z_\Sigma = a + e + \frac{p(1-y)(a+e)}{1-p} + py(a'+e') + \frac{p^2y(1-y)(a'+e')}{1-p} + p^2y^2z_0 + p^3y^2z_1 + \dots. \tag{A.4}$$

This leaves a portion of the phenotype due to feedback effects, which can be rewritten as

$$z_\Sigma = a + e + \frac{p(1-y)(a+e)}{1-p} + py(a'+e') + \frac{p^2y(1-y)(a'+e')}{1-p} + p^2y^2z_\Sigma, \tag{A.5}$$

allowing us to close the entire sum as

$$z_\Sigma = \frac{a+e+py(a'+e')}{1-p^2y^2} + \frac{p(1-y)(a+e) + p^2y(1-y)(a'+e')}{(1-p)(1-p^2y^2)}. \tag{A.6}$$

Simplifying,

$$z_\Sigma = \frac{1-py}{1-p}\left(\frac{a+e+py(a'+e')}{1-p^2y^2}\right). \tag{A.7}$$

Now, we take the average across all interactions by substituting Equation A.7 into Equation 27, yielding

$$z_{\text{ave}} = (1-py)\left(\frac{a+e+py(a'+e')}{1-p^2y^2}\right). \tag{A.8}$$

Finally, Equation A.8 easily simplifies to Equation 28.

References

Abrams, P. A., Y. Harada, and H. Matsuda, 1993. On the relationship between quantitative genetic and ESS models. Evolution 47:982–985.

Akçay, E. and J. Van Cleve, 2012. Behavioral responses in structured populations pave the way to group optimality. American Naturalist 179:257–269.

Akçay, E., J. Van Cleve, M. W. Feldman, and J. Roughgarden, 2009. A theory for the evolution of other-regard integrating proximate and ultimate perspectives. Proceedings of the National Academy of Sciences of the United States of America 106:19061–19066.

Alemu, S. W., P. Bijma, S. H. Møller, L. Janss, and P. Berg, 2014. Indirect genetic effects contribute substantially to heritable variation in aggression-related traits in group-housed mink (*Neovison vison*). Genetics Selection Evolution 46.

Alexander, R. D., 1985. A biological interpretation of moral systems. Zygon 20:3–20.

André, J.-B., 2015. Contingency in the evolutionary emergence of reciprocal cooperation. American Naturalist 185:303–316.

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.

Aoki, K., 1983. A quantitative genetic model of reciprocal altruism: A condition for kin or group selection to prevail. Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences 80:4065–4068. URL `<GotoISI>://WOS:A1983QW61000040`.

———, 1984. A quantitative genetic model of two-policy games between relatives. Journal of Theoretical Biology 109:111–126. URL `<GotoISI>://WOS:A1984TC21000008`.

Araya-Ajoy, Y. G., D. F. Westneat, and J. Wright, 2020. Pathways to social evolution and their evolutionary feedbacks. Evolution 74:1894–1907.

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

Bijma, P., 2010. Estimating indirect genetic effects: precision of estimates and optimum designs. Genetics 186:1013–1028.

———, 2014. The quantitative genetics of indirect genetic effects: a selective review of modelling issues. Heredity 112:61–9.

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.

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.

Binmore, K. G., 2007. Game Theory: A Very Short Introduction. Oxford University Press, New York.

Bleakley, B. H. and E. D. Brodie, III, 2009. Indirect genetic effects influence antipredator behavior in guppies: Estimates of the coefficient of interaction psi and the inheritance of reciprocity. Evolution 63:1796–806.

Charlesworth, B., 1990. Optimization models, quantitative genetics, and mutation. Evolution 44:520–538.

Chenoweth, S. F., H. D. Rundle, and M. W. Blows, 2010. Experimental evidence for the evolution of indirect genetic effects: changes in the interaction effect coefficient, psi (ψ), due to sexual selection. Evolution published online.

Cheverud, J. M., 1984. Evolution by kin selection: a quantitative genetic model illustrated by maternal performance in mice. Evolution 38:766–777.

Darwin, C., 1859. The Origin of Species. John Murray, London.

Dawkins, R., 1976. The Selfish Gene. Oxford Univ. Press, New York.

Dickerson, G. E., 1947. Composition of hog carcasses as influenced by heritable differences in rate and economy of gain. Iowa Agricultural Experiment Station Research Bulletin 354:489–524.

Dugatkin, L. A. and H. K. Reeve, 1998. Game Theory and Animal Behavior. Oxford Univ. Press, New York.

Edenbrow, M., B. H. Bleakley, S. K. Darden, C. R. Tyler, I. W. Ramnarine, and D. P. Croft, 2017. The evolution of cooperation: Interacting phenotypes among social partners. American Naturalist 189:630–643.

Eshel, I., 1996. On the changing concept of evolutionary population stability as a reflection of a changing point of view in the quantitative theory of evolution. Journal of Mathematical Biology 34:485–510.

Falconer, D. S., 1965. Maternal effects and selection response. Proceedings of the XIth International Congress of Genetics 3:763–774.

Farine, D. R. and B. C. Sheldon, 2015. Selection for territory acquisition is modulated by social network structure in a wild songbird. Journal of Evolutionary Biology 28:547–556.

Fisher, D. N. and J. N. Pruitt, 2019. Opposite responses to selection and where to find them. Journal of Evolutionary Biology 32:505–518.

Fisher, R. A., 1930. The Genetical Theory of Natural Selection. Clarendon Press, Oxford.

Formica, V. A., J. W. McGlothlin, C. W. Wood, M. E. Augat, R. E. Butterfield, M. E. Barnard, and E. D. Brodie III, 2011. Phenotypic assortment mediates the effect of social selection in a wild beetle population. Evolution 65:2771–2781.

Frank, S. A., 1997. The price equation, fisher's fundamental theorem, kin selection, and causal analysis. Evolution 51:1712–1729.

———, 1998. Foundations of Social Evolution. Princeton University Press, Princeton, NJ.

———, 2007. All of life is social. Current Biology 17:R648–R650.

Gomulkiewicz, R., 1998. Game theory, optimization, and quantitative genetics. Pp. 283–303, *in* L. A. Dugatkin and H. K. Reeve, eds. Game Theory and Animal Behavior. Oxford Univ. Press, Oxford.

Goodnight, C. J., J. M. Schwartz, and L. Stevens, 1992. Contextual analysis of models of group selection, soft selection, hard selection, and the evolution of altruism. American Naturalist 140:743–761.

Grafen, A., 1979. The hawk-dove game played between relatives. Animal Behaviour 27:905–907.

———, 1984. Natural selection, kin selection and group selection. Pp. 62–84, *in* J. R. Krebs and N. B. Davies, eds. Behavioural Ecology: An Evolutionary Approach. Blackwell Scientific Publications, Oxford.

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

———, 1977. Selection for populations of interacting genotypes. Pp. 413–434, *in* E. Pollak, O. Kempthorne, and T. B. Bailey, eds. Proceedings of the International Conference on Quantitative Genetics, August 16-21, 1976. Iowa State Univ. Press, Ames, IA.

———, 1981. A theory of natural selection incorporating interaction among individuals. I. The modeling process. Journal of Theoretical Biology 89:635–658.

Hadfield, J. D. and C. E. Thomson, 2017. Interpreting selection when individuals interact. Methods in Ecology and Evolution 8:688–699.

Haldane, J. B. S., 1932. The Causes of Evolution. Longmans, Green and Co., London.

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

———, 1964a. The genetical evolution of social behavior. I. Journal of Theoretical Biology 7:1–16.

———, 1964b. The genetical evolution of social behavior. II. Journal of Theoretical Biology 7:17–52.

Han, C. S., C. Tuni, J. Ulcik, and N. J. Dingemanse, 2018. Increased developmental density decreases the magnitude of indirect genetic effects expressed during agonistic interactions in an insect. Evolution 72:2435–2448.

Heisler, I. and J. Damuth, 1987. A method for analyzing selection in hierarchically structured populations. American Naturalist 130:582–602.

Hofbauer, J. and K. Sigmund, 1998. Evolutionary Games and Population Dynamics. Cambridge University Press, Cambridge, UK.

Iwasa, Y., A. Pomiankowski, and S. Nee, 1991. The evolution of costly mate preferences II. The handicap principle. Evolution 45:1431–1442.

Karlin, S. and C. Matessi, 1983. The eleventh R. A. Fisher Memorial Lecture: Kin selection and altruism. Proceedings of the Royal Society B 219:327–353.

Kazancioglu, E., H. Klug, and S. H. Alonzo, 2012. The evolution of social interactions changes predictions about interacting phenotypes. Evolution 66:2056–64.

Kirkpatrick, M., 1982. Sexual selection and the evolution of female mate choice. Evolution 36:1–12.

Kirkpatrick, M. and R. Lande, 1989. The evolution of maternal characters. Evolution 43:485–503.

Krebs, J. R. and N. B. Davies, 1978. Behavioural Ecology: An Evolutionary Approach. Blackwell Scientific, Oxford.

Lande, R., 1976. Natural selection and random genetic drift in phenotypic evolution. Evolution 30:314–334.

———, 1980. Sexual dimorphism, sexual selection, and adaptation in polygenic characters. Evolution 34:292–305.

———, 1981. Models of speciation by sexual selection on polygenic traits. Proceedings of the National Academy of Sciences of the USA 78:3721–3725.

Lande, R. and S. J. Arnold, 1983. The measurement of selection on correlated characters. Evolution 37:1210–1226.

Lane, S. M., A. J. Wilson, and M. Briffa, 2020. Analysis of direct and indirect genetic effects in fighting sea anemones. Behavioral Ecology 31:540–547.

Lehtonen, J., 2018. The Price equation, gradient dynamics, and continuous trait game theory. American Naturalist 191:146–153. URL <GotoISI>://WOS:000418197900013.

———, 2020. The Price equation and the unity of social evolution theory. Philosophical Transactions of the Royal Society B-Biological Sciences 375. URL <GotoISI>://WOS:000519217300006.

Lion, S., 2018. Theoretical approaches in evolutionary ecology: environmental feedback as a unifying perspective. American Naturalist 191:21–44. URL <GotoISI>://WOS:000418197900005.

Lynch, M., 1987. Evolution of intrafamilial interactions. Proceedings of the National Academy of Sciences of the USA 84:8507–8511.

Lyon, B. E. and R. Montgomerie, 2012. Sexual selection is a form of social selection. Philosophical Transactions of the Royal Society B: Biological Sciences 367:2266–2273.

Maynard Smith, J., 1982. Evolution and the Theory of Games. Cambridge Univ. Press, Cambridge.

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

McElreath, R. and R. Boyd, 2007. Mathematical Models of Social Evolution: A Guide for the Perplexed. Univ. of Chicago Press, Chicago.

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.

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.

McGlothlin, J. W., J. B. Wolf, E. D. Brodie III, and A. J. Moore, 2014. Quantitative genetic versions of Hamilton's rule with empirical applications. Philosophical Transactions of the Royal Society B 369:20130358.

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

McNamara, J. M. and O. Leimar, 2020. Game Theory in Biology. Oxford Univ. Press, Oxford.

Moore, A. J. and C. R. B. Boake, 1994. Optimality and evolutionary genetics: complementary procedures for evolutionary analysis in behavioral ecology. Trends in Ecology & Evolution 9:69–72.

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.

Okasha, S., 2004. Multilevel selection and the partitioning of covariance: A comparison of three approaches. Evolution 58:486–494. URL <GotoISI>://000220704800004.

———, 2006. Evolution and the Levels of Selection. Oxford Univ. Press, Oxford.

Phillips, P. C. and S. J. Arnold, 1989. Visualizing multivariate selection. Evolution 43:1209–1222.

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

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

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

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

———, 2011. Expanded social fitness and Hamilton's rule for kin, kith, and kind. Proceedings of the National Academy of Sciences of the USA 108 Suppl 2:10792–9.

Rapoport, A. and A. M. Chammah, 1965. Prisoner's Dilemma: A Study in Conflict and Cooperation. University of Michigan Press, Ann Arbor, MI.

Roughgarden, J., 2012. The social selection alternative to sexual selection. Philosophical Transactions of the Royal Society B: Biological Sciences 367:2294–2303.

Roughgarden, J., M. Oishi, and E. Akçay, 2006. Reproductive social behavior: cooperative games to replace sexual selection. Science 311:965–969.

Saltz, J. B., 2013. Genetic composition of social groups influences male aggressive behaviour and fitness in natural genotypes of *Drosophila melanogaster*. Proceedings of the Royal Society B Biological Sciences 280.

Santostefano, F., D. Garant, P. Bergeron, P.-O. Montiglio, and D. Réale, 2020. Social selection acts on behavior and body mass but does not contribute to the total selection differential in eastern chipmunks. Evolution 74:89–102.

Santostefano, F., A. J. Wilson, P. T. Niemela, and N. J. Dingemanse, 2017. Indirect genetic effects: a key component of the genetic architecture of behaviour. Scientific Reports 7.

Taper, M. L. and T. J. Case, 1992. Models of character displacement and the theoretical robustness of taxon cycles. Evolution 46:317–333.

Taylor, P. D., 1996. The selection differential in quantitative genetics and ESS models. Evolution 50:2106–2110.

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

Trivers, R. L., 1971. Evolution of reciprocal altruism. Quarterly Review of Biology 46:35–57.

Van Cleve, J., 2017. Stags, hawks, and doves: social evolution theory and individual variation in cooperation. Integrative and Comparative Biology 57:566–579.

Van Cleve, J. and E. Akçay, 2014. Pathways to social evolution: reciprocity, relatedness, and synergy. Evolution 68:2245–2258.

Wade, M. J., P. Bijma, E. D. Ellen, and W. Muir, 2010. Group selection and social evolution in domesticated animals. Evolutionary Applications 3:453–465.

West-Eberhard, M. J., 1979. Sexual selection, social competition, and evolution. Proceedings of the American Philosophical Society 123:222–234.

———, 1983. Sexual selection, social competition, and speciation. Quarterly Review of Biology 58:155–183.

———, 2014. Darwin's forgotten idea: The social essence of sexual selection. Neuroscience and Biobehavioral Reviews 46:501–508.

Westneat, D. F., 2012. Evolution in response to social selection: the importance of interactive effects of traits on fitness. Evolution 66:890–895.

Willham, R. L., 1963. The covariance between relatives for characters composed of components contributed by related individuals. Biometrics 19:18–27.

———, 1972. The role of maternal effects in animal breeding. III. Biometrical aspects of maternal effects in animals. Journal of Animal Science 35:1288–1293.

Wilson, A. J., U. Gelin, M.-C. Perron, and D. Réale, 2009. Indirect genetic effects and the evolution of aggression in a vertebrate system. Proceedings of the Royal Society of London B 276:533–541.

Wilson, A. J., M. B. Morrissey, M. J. Adams, C. A. Walling, F. E. Guinness, J. M. Pemberton, T. H. Clutton-Brock, and L. E. B. Kruuk, 2011. Indirect genetic effects and evolutionary constraint: an analysis of social dominance in red deer, *Cervus elaphus*. Journal of Evolutionary Biology 24:772–783.

Wilson, A. J., D. Reale, M. N. Clements, M. M. Morrissey, E. Postma, C. A. Walling, L. E. B. Kruuk, and D. H. Nussey, 2010. An ecologist's guide to the animal model. Journal of Animal Ecology 79:13–26.

Wolf, J. B., E. D. Brodie III, J. M. Cheverud, A. J. Moore, and M. J. Wade, 1998. Evolutionary consequences of indirect genetic effects. Trends in Ecology & Evolution 13:64–69.

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.