Relatedness and synergies of kind and scale in the evolution of helping

Jorge Peña[1,*] Georg Nöldeke[2] Laurent Lehmann[3]

December 4, 2014

[1] Department of Evolutionary Theory
Max Planck Institute for Evolutionary Biology
August-Thienemann-Str. 2, 24306 Plön, Germany
e-mail: pena@evolbio.mpg.de

[2] Faculty of Business and Economics
University of Basel
Peter Merian-Weg 6, CH-4002 Basel, Switzerland
e-mail: georg.noeldeke@unibas.ch

[3] Department of Ecology and Evolution
University of Lausanne
Le Biophore, CH-1015 Lausanne, Switzerland
e-mail: laurent.lehmann@unil.ch

* Corresponding author.

Abstract

Relatedness and synergy affect the selection pressure on cooperation and altruism. Although early work investigated the effect of these factors independently of each other, recent efforts have been aimed at exploring their interplay. Here, we contribute to this ongoing synthesis in two distinct but complementary ways. First, we integrate models of n-player matrix games into the direct fitness approach of inclusive fitness theory, hence providing a framework to consider synergistic social interactions between relatives in family and spatially structured populations. Second, we illustrate the usefulness of this framework by delineating three distinct types of helping traits ("whole-group", "nonexpresser-only" and "expresser-only"), which are characterized by different synergies of kind (arising from differential fitness effects on individuals expressing or not expressing helping) and can be subjected to different synergies of scale (arising from economies or diseconomies of scale). We find that relatedness and synergies of kind and scale can interact to generate nontrivial evolutionary dynamics, such as cases of bistable coexistence featuring both a stable equilibrium with a positive level of helping and an unstable helping threshold. This broadens the qualitative effects of relatedness (or spatial structure) on the evolution of helping.

Keywords. evolution of helping, relatedness, synergy, inclusive fitness, evolutionary games

1 Introduction

Explaining the evolution of helping (cooperation and altruism) has been a main focus of research in evolutionary biology over the last fifty years (e.g., Sachs et al., 2004; West et al., 2007). In this context, Hamilton's seminal papers established the importance of relatedness (genetic assortment between individuals) by showing that an allele for helping can be favored by natural selection as long as $-c+rb > 0$ is satisfied, where c is the fitness cost to an average carrier from expressing the allele, b is the fitness benefit to such a carrier stemming from a social partner expressing the allele, and r is the relatedness between social partners (Hamilton, 1964a,b, 1970). Additional factors, including different forms of reciprocity (i.e., conditional behaviors and responsiveness under multimove interactions, e.g., Trivers, 1971; Axelrod and Hamilton, 1981) and synergy (i.e., nonadditive effects of social behaviors on material payoffs, either positive or negative, e.g., Queller, 1985; Sumpter, 2010), modify the fitness costs and benefits in Hamilton's rule (Axelrod and Hamilton, 1981; Day and Taylor, 1997; Lehmann and Keller, 2006a; Gardner et al., 2011; Van Cleve and Akçay, 2014) and hence fundamentally influence the evolutionary dynamics of helping.

Because of their ubiquity, relatedness and synergy occupy a central role among the factors affecting the selection pressure on helping. Both are clearly present in the cooperative enterprises of most organisms. First, real populations are characterized by limited gene flow at least until the stage of offspring dispersal (Clobert et al., 2001), with the consequence that most social interactions necessarily occur between relatives of varying degree. Second, social exchanges often feature at least one of two different forms of synergy, which we call in this article "synergies of kind" and "synergies of scale".

Synergies of kind (implicit in what Queller, 2011 calls "kind selection") arise when the expression of a social trait benefits recipients in different ways, depending on whether or not (or more generally, to which extent) recipients express the social trait themselves. A classical example of a positive synergy of kind is collective hunting (Packer and Ruttan, 1988), where the benefits of a successful hunt go to cooperators (hunters) but not to defectors (solitary individuals). Examples of negative synergies of kind are eusociality in Hymenoptera, by which sterile workers help queens to reproduce (Bourke and Franks, 1995), and self-destructive cooperation in bacteria, where expressers lyse while releasing virulence factors that benefit nonexpressers (Fröhlich and Madeo, 2000; Ackermann et al., 2008).

Synergies of scale (Corning, 2002) result from economies or diseconomies of scale in the production of a social good, so that the net effect of several individuals behaving socially can be more or less than the sum of individual effects. For instance, enzyme production in microbial cooperation is likely to be nonlinear, as in the cases of invertase hydrolyzing disaccharides into glucose in the budding yeast *Saccharomyces cerevisiae* (Gore et al., 2009) or virulence factors triggering gut inflammation (and hence removal of competitors) in the pathogen *Salmonella typhimurium* (Ackermann et al., 2008). In the former case, the relationship between growth rate and glucose concentration in yeast has been reported to be sublinear, i.e., invertase production has diminishing returns or negative synergies of scale (Gore et al., 2009, fig. 3.*c*); in the latter case, the relationship between the level of expression of virulence factors and inflammation intensity appears to be superlinear, i.e., it exhibits increasing returns or positive synergies of scale (Ackermann et al., 2008, fig. 2.*d*).

Previous theoretical work has investigated the effects of relatedness and synergy on the evolution of helping either independently of each other or by means of simplified models that neglect crucial interactions between the two factors. For instance, the effects of demography on relatedness and the scale of competition in family and spatially structured populations have often been explored under the assumption of additive payoff effects (e.g., Taylor, 1992; Taylor and Irwin, 2000; Lehmann et al., 2006; Gardner and West, 2006), while synergistic interactions have usually been investigated under the assumption that individuals are unrelated (e.g., Motro, 1991; Leimar and Tuomi, 1998; Hauert et al.,

2006). In the cases where relatedness and synergy have been considered to operate in conjunction, it has been customary to model social interactions by means of a two-player Prisoner's Dilemma, modified by adding a synergy parameter D to the payoff of mutual cooperation (Grafen, 1979; Queller, 1984, 1985, 1992; Fletcher and Zwick, 2006; Lehmann and Keller, 2006a,b; Ohtsuki, 2010; Gardner et al., 2011; Ohtsuki, 2012; Taylor and Maciejewski, 2012; Van Cleve and Akçay, 2014). In this framework, $D > 0$ (positive synergy) implies positive frequency-dependent selection, while $D < 0$ (negative synergy) implies negative frequency-dependent selection. The value of relatedness only matters in determining whether or not positive synergy leads to bistability, resp., whether or not negative synergy leads to coexistence.

Although illuminating in some aspects, such two-player models cannot capture patterns of synergy and resulting frequency dependence where positive (resp. negative) synergies of kind and negative (resp. positive) synergies of scale do not combine additively. Such situations are however likely to be common in nature. For example, collective hunting often features both positive synergies of kind and negative synergies of scale (Packer and Ruttan, 1988), while the production of virulence factors in *S. typhimurium* features both negative synergies of kind and positive synergies of scale (Ackermann et al., 2008). Models of two-player matrix games between relatives miss these patterns of synergy (and possible interactions between relatedness and synergy) because such games are linear, and only nonlinear games (which necessarily involve at least three-party interactions) can accommodate both negative and positive synergies without conflating them into a single parameter. Although previous work has explored instances of n-player games between relatives (e.g., Boyd and Richerson, 1988; Eshel and Motro, 1988; Archetti, 2009; Van Cleve and Lehmann, 2013; Marshall, 2014) this has been done only for specific population or payoff structures, and hence not in a comprehensive manner.

In this article, we study the interplay between relatedness and synergies of kind and scale in models of n-player social interactions between relatives. In order to do so, we first present a general framework that integrates n-player matrix games (e.g., Kurokawa and Ihara, 2009; Gokhale and Traulsen, 2010) into the "direct fitness" approach (Taylor and Frank, 1996; Rousset, 2004) of social evolution theory. This framework allows us to deliver a tractable expression for the selection gradient (or gain function) determining the evolutionary dynamics, which differs from the corresponding expression for n-player games between unrelated individuals only in that "inclusive gains from switching" rather than solely "direct gains from switching" must be taken into account.

We then use the theoretical framework to investigate the interaction between relatedness, synergies of kind, and synergies of scale in the evolution of helping. We show the importance of distinguishing between three different kinds of helping traits (which we call "whole-group", "nonexpresser-only" and "expresser-only"), that are characterized by different types of synergies of kind (none for "whole-group", negative for "nonexpresser-only", positive for "expresser-only"), and can be subjected to different synergies of scale. Our analysis demonstrates that the interplay between relatedness and synergy can lead to patterns of frequency dependence, evolutionary dynamics, and bifurcations that cannot arise when considering synergistic interactions between unrelated individuals. Thereby, our approach illustrates how relatedness and synergy combine nontrivially to affect the evolution of social behaviors.

2 Modeling framework

2.1 Population structure (demography)

We consider a homogeneous haploid population subdivided into a finite and constant number of groups, each with a constant number $N \geq 1$ of adult individuals (see table 1 for a list of symbols). The following events occur cyclically and span a demographic time period. Each adult individual gives birth to offspring and then survives with a constant probability, so that individuals can be semelparous (die after

reproduction) or iteroparous (survive for a number of demographic time periods). After reproduction, offspring dispersal occurs. Then, offspring in each group compete for breeding spots vacated by the death of adults so that exactly N individuals reach adulthood in each group.

Dispersal between groups may follow a variety of schemes, including the island model of dispersal (Wright, 1931; Taylor, 1992), isolation by distance (Malécot, 1975; Rousset, 2004), hierarchical migration (Sawyer and Felsenstein, 1983; Lehmann and Rousset, 2012), a model where groups split into daughter groups and compete against each other (Gardner and West, 2006; Lehmann et al., 2006; Traulsen and Nowak, 2006), and several variants of the haystack model (e.g., Matessi and Jayakar, 1976; Godfrey-Smith and Kerr, 2009). We leave the exact details of the life history unspecified, but assume that they fall within the scope of models of spatially homogeneous populations with constant population size (see Rousset, 2004, ch. 6).

2.2 Social interactions (games and payoffs)

Each demographic time period, individuals interact socially by participating in a game between n players. Interactions can occur among all adults in a group ($n = N$), among a subset of such individuals ($n < N$) or among offspring before dispersal ($n > N$). Individuals may either express a social behavior (e.g., cooperate in a Prisoner's Dilemma) or not (e.g., defect in a Prisoner's Dilemma). We denote these two possible actions by A ("cooperation") and B ("defection") and also refer to A-players as "expressers" and to B-players as "nonexpressers". The game is symmetric so that, from the point of view of a focal individual, any two co-players playing the same action are exchangeable. We denote by a_k the material payoff to an A-player when $k = 0, 1, \ldots, n-1$ co-players choose A (and hence $n-1-k$ co-players choose B). Likewise, we denote by b_k the material payoff to a B-player when k co-players choose A.

We assume that individuals implement mixed strategies, i.e., they play A with probability z (and hence play B with probability $1-z$). The set of available strategies is then the interval $z \in [0, 1]$. At any given time only two strategies are present in the population: residents who play A with probability z and mutants who play A with probability $z + \delta$. Let us denote by $z_\bullet$ the strategy (either z or $z + \delta$) of a focal individual, and by $z_{\ell(\bullet)}$ the strategy of the ℓ-th co-player of such focal. The expected payoff π to the focal is then

$$\pi\left(z_\bullet, z_{1(\bullet)}, z_{2(\bullet)}, ..., z_{n-1(\bullet)}\right) = \sum_{k=0}^{n-1} \phi_k\big(z_{1(\bullet)}, z_{2(\bullet)}, \ldots, z_{n-1(\bullet)}\big)\left[z_\bullet a_k + (1 - z_\bullet) b_k\right], \tag{1}$$

where ϕ_k is the probability that exactly k co-players play action A. A first-order Taylor-series expansion about the average strategy $z_\circ = \sum_{\ell=1}^{n-1} z_{\ell(\bullet)}/(n-1)$ of co-players shows that, to first order in δ, the probability ϕ_k is given by a binomial distribution with parameters $n-1$ and $z_\circ$, i.e.,

$$\phi_k\big(z_{1(\bullet)}, z_{2(\bullet)}, \ldots, z_{n-1(\bullet)}\big) = \binom{n-1}{k} z_\circ^k (1 - z_\circ)^{n-1-k} + O(\delta^2). \tag{2}$$

Substituting (2) into (1) and discarding second and higher order terms, we obtain

$$\pi\left(z_\bullet, z_\circ\right) = \sum_{k=0}^{n-1} \binom{n-1}{k} z_\circ^k (1 - z_\circ)^{n-1-k}\left[z_\bullet a_k + (1 - z_\bullet) b_k\right] \tag{3}$$

for the payoff of a focal individual as a function of the focal's strategy $z_\bullet$ and the average strategy $z_\circ$ of the focal's co-players (see also Rousset, 2004, p. 95 and Van Cleve and Lehmann, 2013, p. 85).

2.3 Gain function and convergence stability

Consider a population of residents playing z in which a single mutant $z+\delta$ appears due to mutation, and denote by ρ the fixation probability of the mutant. We take the phenotypic selection gradient $\mathcal{S} = (\mathrm{d}\rho/\mathrm{d}\delta)_{\delta=0}$ as a measure of evolutionary success of the mutant (Rousset and Billiard, 2000, p. 819; Van Cleve, 2014, p. 17); indeed, $\mathcal{S} > 0$ entails that the mutant has a fixation probability greater than neutral under weak selection ($|\delta| \ll 1$). In order to evaluate the fixation probability, we assume that each demographic time period the material payoff to an individual determines its own fecundity (number of offspring produced before competition) or that of its parent (if interactions occur among offspring) by letting the average fecundity of an adult relative to a baseline be equal to the average payoff of a focal actor (i.e., payoffs from the game have "fecundity effects" as opposed to "survival effects", e.g., Taylor and Irwin, 2000). With this and our demographic assumptions, $\mathcal{S}$ is proportional to the "gain function" given by

$$\mathcal{G}(z) = \underbrace{\left.\frac{\partial \pi(z_\bullet, z_\circ)}{\partial z_\bullet}\right|_{z_\bullet = z_\circ = z}}_{\text{"direct" effect, } -\mathcal{C}(z)} + \kappa \underbrace{\left.\frac{\partial \pi(z_\bullet, z_\circ)}{\partial z_\circ}\right|_{z_\bullet = z_\circ = z}}_{\text{"indirect" effect, } \mathcal{B}(z)} = -\mathcal{C}(z) + \kappa \mathcal{B}(z) \tag{4}$$

(see, e.g., Van Cleve and Lehmann, 2013, eq. 7).

Equation (4) shows that the gain function $\mathcal{G}(z)$ is determined by three components. First, the "direct" effect $-\mathcal{C}(z)$, that describes the change in average payoff of the focal resulting from the focal infinitesimally changing its own strategy. Second, the "indirect" effect $\mathcal{B}(z)$, that describes the change in average payoff of the focal resulting from the focal's co-players changing their strategy infinitesimally. Third, the indirect effect is weighted by the "scaled relatedness coefficient" κ, which is a measure of relatedness between the focal individual and its neighbors, demographically scaled so as to capture the effects of local competition on selection (Queller, 1994; Lehmann and Rousset, 2010; Akçay and Van Cleve, 2012). We discuss these three components of the gain function in more detail in the following section.

Knowledge of equation (4) is sufficient to characterize convergent stable strategies (Eshel and Motro, 1981; Eshel, 1983; Taylor, 1989; Christiansen, 1991; Geritz et al., 1998; Rousset, 2004). In our context, candidate convergent stable strategies are either "singular points" (i.e., values $z^* \in (0,1)$ such that $\mathcal{G}(z^*) = 0$), or the two pure strategies $z = 1$ (always play A) and $z = 0$ (always play B). In particular, a singular point z^* is convergent stable if $\mathrm{d}\mathcal{G}(z)/\mathrm{d}z|_{z=z^*} < 0$. Regarding the endpoints, $z = 1$ (resp. $z = 0$) is convergent stable if $\mathcal{G}(1) > 0$ (resp. $\mathcal{G}(0) < 0$). In this article we focus on convergence stability, and thus do not consider the possibility of disruptive selection, which can be ruled out by assuming that the evolutionary dynamics proceeds strictly through a sequence of mutant invasions and fixation events (e.g., the "substitution process"; Gillespie, 1991; the "trait substitution process"; Metz et al., 1996; or the "trait substitution sequence"; Champagnat et al., 2006).

2.4 Inclusive gains from switching

From equation (4), the condition for a mutant to be favored by selection can be written as $-\mathcal{C}(z)+\kappa\mathcal{B}(z) > 0$. This can be understood as a scaled form of the marginal version of Hamilton's rule (Lehmann and Rousset, 2010) with $\mathcal{C}(z)$ corresponding to the marginal direct costs and $\mathcal{B}(z)$ to the marginal indirect benefits of expressing an increased probability of playing action A. These marginal costs and benefits are not measured in terms of actual fitness (number of adult offspring, which are the units of measurement of b and c in Hamilton's rule as given in the introduction, see e.g., Rousset, 2004, p. 113), but in terms of fecundity via payoffs in a game. The scaled relatedness coefficient κ is also not equal to the regression definition of relatedness present in the standard Hamilton's rule, except for special cases where

competition is completely global (Queller, 1994).

The coefficient κ is a function of demographic parameters such as migration rate, group size, and vital rates of individuals or groups, but is independent of the evolving trait z (Van Cleve and Lehmann, 2013). For instance, in the island model with overlapping generations, $\kappa = 2s(1-m)/(N[2-m(1-s)]+2(1-m)s)$, where m is the migration rate and s is the probability of surviving to the next generation (Taylor and Irwin, 2000, eq. A10; Akçay and Van Cleve, 2012, app. A2). In broad terms, we have (i) $\kappa > 0$ for population structures characterized by positive assortment and relatively global competition, (ii) $\kappa = 0$ for infinitely large panmictic populations or for viscous populations with local competition exactly compensating for increased assortment of strategies (Taylor, 1992), and (iii) $\kappa < 0$ for population structures characterized by negative assortment and/or very strong local competition (e.g., density-dependent competition occurs before dispersal). Scaled relatedness coefficients have been evaluated for many life cycle conditions (see table 1 of Lehmann and Rousset, 2010, table 1 of Van Cleve and Lehmann, 2013, and references therein; see also app. A for values of κ under different variants of the haystack model).

In contrast to κ, which depends only on population structure, the other two components of the gain function are solely determined by the payoff structure of the social interaction. In the following, we show how $-\mathcal{C}(z)$ and $\mathcal{B}(z)$ can be expressed in terms of the payoffs a_k and b_k of the game. Doing so delivers an expression for $\mathcal{G}(z)$ that can be analyzed with the same techniques applicable for games between unrelated individuals. This expression provides the foundation for our subsequent analysis.

Imagine a focal individual playing B in a group where k of its co-players play A. Suppose that this focal individual unilaterally switches its action to A while its co-players hold fixed their actions, thus changing its payoff from b_k to a_k. As a consequence, the focal experiences a "direct gain from switching" given by

$$d_k = a_k - b_k,\ k = 0, 1, \ldots, n-1. \tag{5}$$

At the same time, each of the focal's co-players playing A experiences a change in payoff given by $\Delta a_{k-1} = a_k - a_{k-1}$ and each of the focal's co-players playing B experiences a change in payoff given by $\Delta b_k = b_{k+1} - b_k$. Hence, taken as a block, the co-players of the focal experience a change in payoff given by

$$e_k = k\Delta a_{k-1} + (n-1-k)\Delta b_k,\ k = 0, 1, \ldots, n-1, \tag{6}$$

where we let $a_{-1} = b_{n+1} = 0$ for mathematical convenience. From the perspective of the focal, this change in payoffs represents an "indirect gain from switching" the focal obtains if co-players are related.

In appendix B, we show that the partial derivatives appearing in (4) can be expressed as expected values of the direct and indirect gains from switching, so that the direct and indirect effects are respectively given by

$$-\mathcal{C}(z) = \sum_{k=0}^{n-1} \binom{n-1}{k} z^k (1-z)^{n-1-k} d_k, \tag{7}$$

and

$$\mathcal{B}(z) = \sum_{k=0}^{n-1} \binom{n-1}{k} z^k (1-z)^{n-1-k} e_k. \tag{8}$$

Hence, defining the "inclusive gains from switching" as

$$f_k = d_k + \kappa e_k,\ k = 0, 1, \ldots, n-1, \tag{9}$$

the gain function can be written as the expected value of the inclusive gains from switching:

$$\mathcal{G}(z) = \sum_{k=0}^{n-1} \binom{n-1}{k} z^k (1-z)^{n-1-k} f_k. \tag{10}$$

An immediate consequence of equation (10) is that matrix games between relatives are mathematically equivalent to "transformed" games between unrelated individuals, where "inclusive payoffs" take the place of standard, or personal, payoffs. Indeed, consider a game in which a focal playing A (resp. B) obtains payoffs

$$a'_k = a_k + \kappa \left[k a_k + (n-1-k) b_{k+1} \right], \; k = 0, 1, \ldots, n-1 \tag{11}$$

$$b'_k = b_k + \kappa \left[k a_{k-1} + (n-1-k) b_k \right], \; k = 0, 1, \ldots, n-1 \tag{12}$$

when k of its co-players play A. Using equations (5)–(6) we can rewrite equation (9) as $f_k = a'_k - b'_k$, so that the inclusive gains from switching are identical to the direct gains from switching in a game with payoff structure given by equations (11)–(12). The payoffs a'_k (resp. b'_k) can be understood as inclusive payoffs consisting of the payoff obtained by a focal playing A (resp. B) plus κ times the sum of the payoffs obtained by its co-players.

This observation has two relevant consequences. First, the results developed in Peña et al. (2014) for nonlinear n-player matrix games between unrelated individuals, which are based on the observation that the right side of (10) is a polynomial in Bernstein form (Farouki, 2012), also apply here, provided that (i) the inclusive gains from switching f_k are used instead of the standard (direct) gains from switching d_k in the formula for the gain function, and (ii) the concept of evolutionary stability is read as meaning convergence stability. For a large class of games, these results allow to identify convergence stable points from a direct inspection of the sign pattern of the inclusive gains from switching f_k. Second, we may interpret the effect of relatedness on selection as inducing the payoff transformation $a_k \to a'_k$, $b_k \to b'_k$. For $n = 2$, this payoff transformation is the one hinted at by Hamilton (1971) and later often discussed in the theoretical literature (Grafen, 1979; Hines and Maynard Smith, 1979; Day and Taylor, 1998), namely

$$\begin{pmatrix} a'_0 & a'_1 \\ b'_0 & b'_1 \end{pmatrix} = \begin{pmatrix} a_0 + \kappa b_1 & (1+\kappa) a_1 \\ (1+\kappa) b_0 & b_1 + \kappa a_0 \end{pmatrix},$$

where the payoff of the focal is augmented by adding κ times the payoff of the co-player.

3 Evolutionary dynamics of three kinds of helping traits

Throughout the following we assume that each A-player incurs a payoff cost $\gamma > 0$ in order for a social good to be produced (e.g., harvested food, nest defense, or help directed to others). The benefits of the social good are accrued by a subset of individuals in the group that we call "recipients". Each recipient obtains a benefit $\beta_j > 0$ when there are $j > 0$ expressers in the group, and no benefit is produced if no individual expresses the social trait ($\beta_0 = 0$). The benefit is increasing in the number of expressers, that is, the "incremental benefit" $\Delta\beta_j = \beta_{j+1} - \beta_j$ is positive ($\Delta\beta_j > 0$).

Synergies of scale are characterized by the properties of the incremental benefits. In the absence of synergies of scale, each additional expresser increases the benefit by the same amount so that $\Delta\beta_j$ is constant, implying that β_j is linear in j. With negative synergies of scale, $\Delta\beta_j$ is decreasing in j, whereas positive synergies of scale arise when $\Delta\beta_j$ is increasing in j. To illustrate the effects of synergies of scale on the evolutionary dynamics of the social trait, we will consider the special case in which incremental benefits are given by the geometric sequence $\Delta\beta_j = \beta\lambda^j$ for some $\beta > 0$ and $\lambda > 0$, so that benefits are

given by

$$\beta_j = \beta \sum_{\ell=0}^{j-1} \lambda^\ell. \tag{13}$$

With geometric benefits, synergies of scale are absent when $\lambda = 1$, negative when $\lambda < 1$, and positive when $\lambda > 1$.

We distinguish three kinds of social traits according to which individuals are recipients and thus benefit from the expression of the social behavior: (i) "whole-group" (benefits accrue to all individuals in the group, fig. 1.a), (ii) "nonexpresser-only" (benefits accrue only to nonexpressers, fig. 1.b), and (iii) "expresser-only" (benefits accrue only to expressers, fig. 1.c). For whole-group traits there are no synergies of kind: benefits accrue to all individuals irrespective of their kind, i.e., whether they are expressers or nonexpressers. In contrast, nonexpresser-only traits feature negative synergies of kind, whereas expresser-only traits feature positive synergies of kind. These differences are reflected in different payoff structures for the corresponding n-player games, resulting in different direct, indirect, and inclusive gains from switching (see table 2).

A classical example of a whole-group trait is the voluntary provision of public goods (Samuelson, 1954). In this case, the expressed social behavior consists in the production of a good available to others and hence exploitable by nonproducing cheats (nonexpressers). Well-known instances of public-goods cooperation are sentinel behavior in animals (Maynard Smith, 1965; Clutton-Brock et al., 1999), and the secretion of extracellular products (Velicer, 2003; West et al., 2007), such as sucrose-digestive enzymes (Greig and Travisano, 2004; Gore et al., 2009), in social bacteria.

The most prominent social behavior matching our definition of a nonexpresser-only trait is altruistic self-sacrifice, which happens when individuals expressing the social behavior sacrifice themselves (or their reproduction) to benefit nonexpressers (Frank, 2006; West et al., 2006). Sterile castes in eusocial insects (Bourke and Franks, 1995), and bacteria lysing while releasing toxins (Fröhlich and Madeo, 2000) or virulence factors (Ackermann et al., 2008) that benefit other bacteria provide some examples of altruistic self-sacrifice in nature.

Expresser-only traits have been discussed under the rubrics of "synergistic" (Queller, 1984, 1985; Leimar and Tuomi, 1998) and "greenbeard" (Guilford, 1985; Gardner and West, 2010; Queller, 2011) effects, and conceptualized as involving "rowing"(Maynard Smith and Szathmáry, 1995, p. 261-262) or "stag hunt" (Skyrms, 2004) games. Often cited examples include collective hunting (Packer and Ruttan, 1988), foundresses cooperating in colony establishment (Bernasconi and Strassmann, 1999), aposematic (warning) coloration (Queller, 1984, 1985; Guilford, 1988), and the *Ti* plasmid in the bacterial pathogen *Agrobacterium tumefaciens*, which induces its plant host to produce opines, a food source that can be exploited only by bacteria bearing the plasmid (Dawkins, 1999, p. 218, White and Winans, 2007). In each of these examples, the social good accrues only to partners expressing the trait, either because of a greater tendency to group and interact or because of the action of an emergent recognition system discriminating expressers from nonexpressers.

For all three kinds of social traits, the indirect gains from switching are always nonnegative ($e_k \geq 0$ for all k) and hence the indirect effect $\mathcal{B}(z)$ is nonnegative for all z. This implies that we deal with helping traits at the level of payoffs and that increasing κ never leads to less selection for expressing the social behavior. Due to their different ways of defining recipients, however, each social trait is characterized by a social dilemma with structurally different payoff, direct gain, and indirect gains from switching. For nonexpresser-only traits, the direct gains from switching are always negative ($d_k < 0$ for all k) and thus expressing the social behavior is also payoff altruistic ($-\mathcal{C}(z) < 0$ and $\mathcal{B}(z) \geq 0$ for all z). For whole-group and expresser-only traits, expressing the social behavior is not necessarily altruistic, depending on how

the cost γ compares to benefits (β_{k+1}, expresser-only traits) or incremental benefits ($\Delta\beta_k$, whole-group traits).

Before turning to the analysis, we note that a fourth class of social traits is sometimes also distinguished in the literature, namely "other-only" traits where the benefits accrue to all other individuals in the group, but not to the focal expresser itself (Pepper, 2000). Other-only traits, as whole-group traits, lack synergies of kind, and hence the effects of relatedness on the evolutionary dynamics are qualitatively similar for whole-group and other-only traits. We discuss this latter case in more detail in section 3.5 and relegate the formal analysis, which is similar to the one for whole-group traits, to appendix D.

3.1 No synergies of scale

To isolate the effects of synergies of kind, we begin our analysis with the case in which synergies of scale are absent, that is, benefits take the linear form $\beta_j = \beta j$ ($\lambda = 1$ in eq. (13)). The resulting expressions for the inclusive gains from switching and the gain functions for the three different social traits are shown in table 3. In each case, the gain function can be written as

$$\mathcal{G}(z) = (n-1)\left[-C + \kappa B + (1+\kappa)Dz\right],$$

where the parameter $C > 0$ may be thought of as the "effective cost" per co-player of expressing the social trait when none of the co-players expresses the social trait. We have $C = \gamma/(n-1)$ when a focal expresser is not among the recipients (nonexpresser-only traits) and $C = (\gamma - \beta)/(n-1)$ otherwise (whole-group and expresser-only traits). The parameter $B \geq 0$ measures the incremental benefit accruing to each co-player of a focal expresser when none of the co-players expresses the social trait. We thus have $B = 0$ for expresser-only traits and $B = \beta$ otherwise. Finally, D measures synergies of kind and is thus null for whole-group traits ($D = 0$), negative for nonexpresser-only traits ($D = -\beta$) and positive for expresser-only traits ($D = \beta$).

In the absence of synergies of kind ($D = 0$, whole-group traits) selection is frequency independent and defection dominates cooperation ($z = 0$ is the only convergence stable strategy) if $-C + \kappa B < 0$ holds, whereas cooperation dominates defection ($z = 1$ is the only convergence stable strategy) if $-C + \kappa B > 0$ holds.

With negative synergies of kind ($D < 0$), there is negative frequency-dependent selection. Defection dominates cooperation if $-C + \kappa B \leq 0$ holds, whereas cooperation dominates defection if $-C + \kappa B + (1+\kappa)D \geq 0$ holds. If $-C + \kappa B + (1+\kappa)D < 0 < -C + \kappa B$ holds, both $z = 0$ and $z = 1$ are unstable and the singular point

$$z^* = \frac{C - \kappa B}{(1+\kappa)D} \tag{14}$$

is stable.

With positive synergies of kind ($D > 0$), there is positive frequency-dependent selection. Defection dominates cooperation if $-C + \kappa B + (1+\kappa)D \leq 0$ holds, whereas cooperation dominates defection if $-C + \kappa B \geq 0$ holds. When $-C + \kappa B < 0 < -C + \kappa B + (1+\kappa)D$, there is bistability: both $z = 0$ and $z = 1$ are stable and z^* is unstable.

This analysis reveals three important points. First, in the absence of synergies of scale the gain function is linear in z, which allows for a straightforward analysis of the evolutionary dynamics for all three kinds of social traits. Second, because of the linearity of the gain function, the evolutionary dynamics of such games fall into one of the four classical dynamical regimes arising from 2×2 games, namely (i) A dominates B, (ii) B dominates A, (iii) coexistence, and (iv) bistability (see, e.g., Cressman, 2003, section

2.2). Third, which of these dynamical regimes arises is determined by the interaction of relatedness with synergies of kind in a straightforward fashion. For all traits, defection dominates cooperation when relatedness is low. For whole-group traits, high values of relatedness imply that cooperation dominates defection. For nonexpresser-only and expresser-only traits, high relatedness also promotes cooperation, leading to either the coexistence of expressers and nonexpressers (nonexpresser-only traits) or to bistability (expresser-only traits).

3.2 Whole-group traits with synergies of scale

For whole-group traits there are no synergies of kind, but either positive or negative synergies of scale may arise. How do such synergies of scale change the evolutionary dynamics of whole-group helping? Substituting the inclusive gains from switching given in table 2 into equation (10) shows that the gain function for whole-group traits is given by

$$\mathcal{G}(z) = \sum_{k=0}^{n-1} \binom{n-1}{k} z^k (1-z)^{n-1-k} \left\{-\gamma + [1+\kappa(n-1)]\,\Delta\beta_k\right\}. \tag{15}$$

Since the incremental benefit satisfies $\Delta\beta_k > 0$ for all k, the gain function (15) is negative for $\kappa \leq -1/(n-1)$. In this case, defection dominates cooperation and $z = 0$ is the only stable point. Hence, we consider the case $\kappa > -1/(n-1)$ throughout the following.

If synergies of scale are negative ($\Delta\beta_k$ decreasing in k), the direct gains (d_k) indirect gains (e_k) and inclusive gains (f_k) from switching are all decreasing in k. This implies that $-\mathcal{C}(z)$, $\mathcal{B}(z)$ and $\mathcal{G}(z)$ are all decreasing in z (cf. Peña et al., 2014, remark 3). Similarly, if synergies of scale are positive ($\Delta\beta_k$ increasing in k), d_k, e_k and f_k are all increasing in k and hence $-\mathcal{C}(z)$, $\mathcal{B}(z)$ and $\mathcal{G}(z)$ are all increasing in z. In both cases the evolutionary dynamics are easily characterized by applying the results for public goods games with constant costs from Peña et al. (2014, section 4.3): with negative synergies of scale, defection dominates cooperation (so that $z = 0$ is the only convergent stable strategy) if $\gamma \geq [1+\kappa(n-1)]\Delta\beta_0$, whereas cooperation dominates defection if $\gamma \leq [1+\kappa(n-1)]\Delta\beta_{n-1}$ holds. If $[1+\kappa(n-1)]\Delta\beta_{n-1} < \gamma < [1+\kappa(n-1)]\Delta\beta_0$ holds, there is coexistence: both $z = 0$ and $z = 1$ are unstable and there is a unique stable interior point z^*. With positive synergies of scale, defection dominates cooperation if $\gamma \geq [1+\kappa(n-1)]\Delta\beta_{n-1}$, whereas cooperation dominates defection if $\gamma \leq [1+\kappa(n-1)]\Delta\beta_0$. If $[1+\kappa(n-1)]\Delta\beta_0 < \gamma < [1+\kappa(n-1)]\Delta\beta_{n-1}$ holds, there is bistability: both $z = 0$ and $z = 1$ are stable and there is a unique, unstable interior point z^* separating the basins of attraction of these two stable strategies. These results resemble those for the cases in which there are no synergies of scale (section 3.1), but negative, resp. positive synergies of kind are present. In particular, it is again the case that the evolutionary dynamics fall into one of the four classical dynamical regimes arising from 2×2 games.

The effect of relatedness on the evolution of whole-group traits can be better grasped by noting that multiplying and dividing (15) by $1+\kappa(n-1)$, we obtain

$$\mathcal{G}(z) = [1+\kappa(n-1)] \sum_{k=0}^{n-1} \binom{n-1}{k} z^k (1-z)^{n-1-k} \left(-\tilde{\gamma} + \Delta\beta_k\right), \tag{16}$$

where $\tilde{\gamma} = \gamma/[1+\kappa(n-1)]$. Equation (16) is (up to multiplication by a positive constant) equivalent to the gain function of a public goods game between unrelated individuals with payoff cost $\tilde{\gamma}$ for producing the public good, which has been analyzed under different assumptions on the shape of the benefit sequence (Motro, 1991; Bach et al., 2006; Hauert et al., 2006; Peña et al., 2014). Hence, relatedness can be conceptualized as affecting only the cost of cooperation, while leaving synergies of scale and patterns of frequency dependence unchanged.

As a concrete example, consider the case of geometric benefits (13) with $\lambda \neq 1$ (see table 4 for a summary of the results and app. C for a derivation). We find that there are two critical cost-to-benefit ratios

$$\varepsilon = \min\left(1+\kappa(n-1), \lambda^{n-1}[1+\kappa(n-1)]\right) \text{ and } \vartheta = \max\left(1+\kappa(n-1), \lambda^{n-1}[1+\kappa(n-1)]\right), \quad (17)$$

such that for small costs ($\gamma/\beta \leq \varepsilon$) cooperation dominates defection ($z = 1$ is the only stable point) and for large costs ($\gamma/\beta \geq \vartheta$) defection dominates cooperation ($z = 0$ is the only stable point). For intermediate costs ($\varepsilon < \gamma/\beta < \vartheta$), there is a singular point given by

$$z^* = \frac{1}{1-\lambda}\left[1 - \left(\frac{\gamma}{\beta\left[1+\kappa(n-1)\right]}\right)^{\frac{1}{n-1}}\right], \quad (18)$$

such that the evolutionary dynamics are characterized by coexistence if synergies of scale are negative ($\lambda < 1$) and by bistability if synergies of scale are positive ($\lambda > 1$). It is clear from equation (17) that, for a given cost-to-benefit ratio γ/β, increasing relatedness makes larger (resp. smaller) the region in the parameter space where cooperation (resp. defection) dominates. Moreover, and from equation (18), z^* is an increasing (resp. decreasing) function of κ when $\lambda < 1$ (resp. $\lambda > 1$), meaning that the proportion of individuals cooperating at a stable interior point (resp. the size of the basin of attraction of the fully cooperative equilibrium) increases as a function of κ (see fig. 2.*a* and 2.*d*).

3.3 Nonexpresser-only traits with synergies of scale

For nonexpresser-only traits, synergies of kind are negative. In the absence of synergies of scale, and as discussed in section 3.1, this implies negative frequency dependence. To investigate how positive or negative synergies of scale change this baseline scenario, we focus on the case in which relatedness is nonnegative ($\kappa \geq 0$).

From the formulas for d_k and e_k given in table 2, it is clear that, independently of any synergies of scale, the direct gains from switching d_k are decreasing in k. Hence, the direct effect $-\mathcal{C}(z)$ is negative frequency-dependent. When synergies of scale are negative, the indirect gains from switching e_k are also decreasing in k, implying that the indirect effect $\mathcal{B}(z)$ is also negative frequency-dependent and that the same is true for the gain function $\mathcal{G}(z) = -\mathcal{C}(z) + \kappa\mathcal{B}(z)$. Hence, negative synergies of scale lead to evolutionary dynamics that are qualitatively identical to those arising when synergies of scale are absent: for low relatedness, defection dominates cooperation, and for sufficiently high relatedness, a unique interior stable equilibrium appears (see app. E.1 and fig. 2.*b*).

When synergies of scale are positive, the indirect gains from switching e_k may still be decreasing in k because the incremental gain $\Delta\beta_k$ accrues to a smaller number of recipients $(n-1-k)$ as k increases. In such a scenario, always applicable when $n = 2$, the evolutionary dynamics are again qualitatively identical to those arising when synergies of scale are absent. A different picture can emerge if $n > 2$ holds and synergies of scale are not only positive, but also sufficiently strong. Then, the indirect gains from switching may be unimodal (first increasing, then decreasing) in k, implying (Peña et al., 2014) that the indirect benefit $\mathcal{B}(z)$ is similarly unimodal, featuring positive frequency dependence for small z and negative frequency dependence for large z. Depending on the value of relatedness, which modulates how the frequency dependence of $\mathcal{B}(z)$ interacts with that of $\mathcal{C}(z)$, this can give rise to evolutionary dynamics different from those possible without synergies of scale, discussed in section 3.1.

For a concrete example of such evolutionary dynamics, consider the case of geometric benefits (13) with $\lambda > 1$ (see table 4 for a summary of results, app. E.2 for their derivation and fig. 2.*e* for an

illustration). In this case, the evolutionary dynamics for $\kappa \geq 0$ and $n > 2$ depend on the critical value

$$\varrho = \frac{1 + \kappa(n-1)}{\kappa(n-2)}, \tag{19}$$

and on the two critical cost-to-benefit ratios

$$\zeta = \kappa(n-1), \qquad \text{and} \qquad \eta = \frac{1}{\lambda - 1}\left[1 + \lambda\kappa\left(\frac{(n-2)\lambda\kappa}{1+\kappa(n-1)}\right)^{n-2}\right], \tag{20}$$

which satisfy $\varrho > 1$ and $\zeta < \eta$.

With these definitions our results can be stated as follows. For $\lambda \leq \varrho$ the dynamical outcome depends on how the cost-to-benefit ratio γ/β compares to ζ. If $\gamma/\beta \geq \zeta$ (high costs), defection dominates cooperation, while if $\gamma/\beta < \zeta$ (low costs), there is coexistence. For $\lambda > \varrho$, the dynamical outcome also depends on how the cost-to-benefit ratio γ/β compares to η. If $\gamma/\beta \geq \eta$ (high costs), defection dominates cooperation. If $\gamma/\beta \leq \zeta$ (low costs), we have coexistence, with the stable singular point z^* satisfying $z^* > \hat{z}$ where

$$\hat{z} = \frac{\kappa\left[(n-2)\lambda - (n-1)\right] - 1}{[1+\kappa(n-1)](\lambda - 1)}. \tag{21}$$

In the remaining case ($\zeta < \gamma/\beta < \eta$, intermediate costs) the dynamics are characterized by bistable coexistence, with $z = 0$ stable, $z = 1$ unstable, and two singular points z_{L} (unstable) and z_{R} (stable) satisfying $0 < z_{\mathrm{L}} < \hat{z} < z_{\mathrm{R}} < 1$. Numerical values for z_{L} (resp. z_{R}) can be obtained by searching for roots of $\mathcal{G}(z)$ in the interval $(0, \hat{z})$ (resp. $(\hat{z}, 1)$), as we illustrate in figure 2.*e*.

It is evident from the dependence of ϱ, ζ, and η on κ that relatedness plays an important role in determining the stable level(s) of expression of helping. As κ increases, the regions of the parameter space where some non-zero level of expression of helping is stable expand at the expense of the region of dominant non-expression. This is so because ζ and η are increasing functions of κ and ϱ is a decreasing function of κ. Moreover, inside these regions the stable non-zero probability of expressing helping increases with κ (see fig. 2.*b* and 2.*e*). Three cases can be however distinguished as for the effects of increasing κ when starting from a point in the parameter space where $z = 0$ is the only stable point. First, $z = 0$ can remain stable irrespective of the value of relatedness, which characterizes high cost-to-benefit ratios. Second, the system can undergo a transcritical bifurcation as κ increases, destabilizing $z = 0$ and leading to the appearance of a unique stable interior point (fig. 2.*b*). This happens when λ and γ/β are relatively small. Third, there is a range of intermediate cost-to-benefit ratios such that, for sufficiently large values of λ, the system undergoes a saddle-node bifurcation, whereby two singular points (z_{L}, unstable, and z_{R}, stable) appear (fig. 2.*e*). In this latter case, positive synergies of scale are strong enough to interact with negative synergies of kind and relatedness in a nontrivial way.

3.4 Expresser-only traits with synergies of scale

For expresser-only traits, and independently of any synergies of scale, the direct gains from switching d_k (cf. table 2) are increasing in k, implying that the direct effect $-\mathcal{C}(z)$ is positive frequency-dependent. When synergies of scale are positive, the indirect gains from switching e_k are also increasing in k, so that the indirect effect $\mathcal{B}(z)$ is also positive frequency-dependent. Focusing on the case of nonnegative relatedness ($\kappa \geq 0$) this ensures that, just as when synergies of scale are absent, the gain function $\mathcal{G}(z)$ is positive frequency-dependent. Hence, the evolutionary dynamics are qualitatively identical to those arising from linear benefits: for low relatedness, defection dominates cooperation, and for high relatedness, there is bistability, with the basins of attraction of the two pure equilibria $z = 0$ and $z = 1$ being separated

by a unique interior unstable point (see app. F.1 and fig. 2.f).

When synergies of scale are negative, the indirect gains from switching e_k may still be increasing in k because the incremental gain $\Delta\beta_k$ accrues to a larger number of recipients as k increases. In such a scenario, always applicable when $n = 2$, the evolutionary dynamics are again qualitatively identical to those arising when synergies of scale are absent. A different picture can emerge if $n > 2$ holds and synergies of scale are not only negative, but also sufficiently strong. In this case, $\mathcal{B}(z)$ can be negative frequency-dependent for some z, and hence (for sufficiently high values of κ) also $\mathcal{G}(z)$. Similarly to the case of nonexpresser-only traits with positive synergies of scale, this can give rise to patterns of frequency dependence that go beyond the scope of helping without synergies of scale.

To illustrate this, consider the case of geometric benefits (13) with $\lambda < 1$, $\kappa \geq 0$, and $n > 2$ (see table 4 for a summary of results, app. F.2 for proofs and fig. 2.c for an illustration). Defining the critical value

$$\xi = \frac{\kappa(n-2)}{1+\kappa(n-1)}, \tag{22}$$

and the two critical cost-to-benefit ratios

$$\varsigma = \frac{1-\lambda^n}{1-\lambda} + \kappa(n-1)\lambda^{n-1}, \qquad \text{and} \qquad \tau = \frac{1}{1-\lambda}\left[1+\lambda\kappa\left(\frac{(n-2)\kappa}{1+\kappa(n-1)}\right)^{n-2}\right], \tag{23}$$

which satisfy $\xi < 1$ and $\varsigma < \tau$, our result can be stated as follows. For $\lambda \geq \xi$ the evolutionary dynamics depends on how the cost-to-benefit ratio γ/β compares to 1 and to ς. If $\gamma/\beta \leq 1$ (low costs), cooperation dominates defection, while if $\gamma/\beta \geq \varsigma$ (high costs), defection dominates cooperation. If $1 < \gamma/\beta < \varsigma$ (intermediate costs), the dynamics are bistable. For $\lambda < \xi$, the classification of possible evolutionary dynamics is as in the case $\lambda \geq \xi$, except that, if $\varsigma < \gamma/\beta < \tau$, the dynamics are characterized by bistable coexistence, with $z = 0$ stable, $z_{\mathrm{L}} \in (0, \hat{z})$ unstable, $z_{\mathrm{R}} \in (\hat{z}, 1)$ stable, and $z = 1$ unstable, where

$$\hat{z} = \frac{1+\kappa}{[1+\kappa(n-1)](1-\lambda)}. \tag{24}$$

For $\kappa \geq 0$, the critical values ξ, ς, and τ are all increasing functions of κ. Hence, as relatedness κ increases, the regions of the parameter space where some level of expression of helping is stable expand at the expense of the region of dominant nonexpression. Moreover, inside these regions the stable positive probability of expressing helping increases with κ (fig. 2.c). When synergies of scale are "sufficiently" negative ($\lambda < \xi$) and for intermediate cost-to-benefit ratios ($\varsigma < \gamma/\beta < \tau$) relatedness and synergies interact in a nontrivial way, leading to saddle-node bifurcations as κ increases (fig. 2.c).

3.5 Connections with previous models

Our model without synergies of scale, for which the $\mathcal{G}(z)$ is linear in z (section 3.1) extends classical two-player matrix games between relatives (e.g. Grafen, 1979, Frank, 1998, ch. 5-6) to the more general case of n-player linear games between relatives. Indeed, for $n = 2$, identifying scaled relatedness κ with relatedness r, and up to normalization of the payoff matrices, equation (14) recovers Grafen (1979, eq. 9) and Frank (1998, eq. 5.6). Interestingly, Frank (1998, p. 98) considers a two-player model of helping with two pure strategies ("nesting" or expressing a queen phenotype, and "helping" or expressing a sterile worker phenotype), which is a particular case of our model of nonexpresser-only traits.

Our results on whole-group traits with geometric returns (section 3.2 and app. C) extend the model studied by Hauert et al. (2006, p. 198) from the particular case of interactions between unrelated individuals ($\kappa = 0$) to the more general case of interactions between relatives ($\kappa \in [-1, 1]$) and recover the result by Archetti (2009, p. 476) in the limit $\lambda \to 0$, in which the game is also called a "volunteer's

dilemma" (Diekmann, 1985). Although we restricted our attention to the cases of constant, decreasing, and increasing incremental benefits, it is clear that equation (16) applies to benefits β_j of any shape. Hence, general results about the stability of equilibria in public goods games (Peña et al., 2014) with sigmoid benefits (Bach et al., 2006; Archetti and Scheuring, 2011) carry over to games between relatives.

For their model of "self-destructive cooperation" in bacteria, Ackermann et al. (2008) assumed a nonexpresser-only trait with no synergies of scale, and a haystack model of population structure implying $\kappa = (N_\mathrm{o} - N)/(N_\mathrm{o}(N-1))$, where $n = N_\mathrm{o} \geq N$ is the number of offspring among which the game is played (see eq. (A.4)). Identifying our γ and β with (respectively) their β with b, the main result of Ackermann et al. (2008) (eq. 7 in their supplementary material) is recovered as a particular case of our result that the unique convergent stable strategy for this case is given by $z^* = [\kappa(n-1)\beta - \gamma]/[(1+\kappa)(n-1)\beta]$ (eq. (14)). The fact that in this example κ is a probability of coalescence within groups shows that social interactions effectively occur between family members, and hence that kin selection is crucial to the understanding of self-destructive cooperation (Gardner and Kümmerli, 2008).

As mentioned before, the analysis of other-only traits follows closely that of whole-group traits (see app. D). The model of altruistic helping in Eshel and Motro (1988) considers such an other-only trait. In their model, one individual in the group needs help, which can be provided (action A) or denied (action B) by its $n-1$ neighbors: a situation Eshel and Motro call the "three brothers' problem" when $n = 3$. Suppose that the cost for each helper is a constant $\varepsilon > 0$ independent on the number of expressers (Eshel and Motro (1988)'s "risk for each volunteer", denoted by c in their paper) and that the benefit for the individual in need when k co-players offer help is given by v_k (Eshel and Motro (1988)'s "gain function", denoted by b_k in their paper). Then, if individuals need help at random, the payoffs for helping (A) and not helping (B) are given by $a_k = -\varepsilon(n-1)/n + v_k/n$ and $b_k = v_k/n$. Defining $\gamma = \varepsilon(n-1)/n$ and $\beta_k = v_k/(n-1)$, we have $a_k = -\gamma + \beta_k$ and $b_k = \beta_k$. Comparing these with the payoffs for whole-group traits in table 2, it is apparent that the key difference between other-only traits and whole-group traits is that an expresser is not among the recipients of its own helping behavior. As we show in appendix D, our results for whole-group traits carry over to such other-only traits. In particular, our results for whole-group traits with geometric benefits can be used to recover results 1,2, and 3 of Eshel and Motro (1988) and to extend them from family-structured to spatially-structured populations.

Finally, Van Cleve and Lehmann (2013) discuss an n-player coordination game. They assume payoffs given by $a_k = 1 + S(R/S)^{k/(n-1)}$ and $b_k = 1 + P(T/P)^{k/(n-1)}$, for positive R, S, T, and P, satisfying $R > T$, $P > S$ and $P > T$. It is easy to see that both the direct effect $-\mathcal{C}(z)$ and the indirect effect $\mathcal{B}(z)$ are strictly increasing functions of z having exactly one sign change. This implies that, for $\kappa \geq 0$, the evolutionary dynamics are characterized by bistability, with the basins of attraction of the two equilibria $z = 0$ and $z = 1$ being divided by the interior unstable equilibrium z^*. Importantly, and in contrast to the social traits analyzed in this article, expressing the payoff dominant action A does not always qualify as a helping trait, as $\mathcal{B}(z)$ is negative for some interval $z \in [0, \hat{z})$. As a result, increasing scaled relatedness κ can have mixed effects on the location of z^*. Both of these predictions are well supported by the numerical results reported by Van Cleve and Lehmann (2013), where increasing κ leads to a steady increase in z^* for $R = 2$, $S = 0.5$, $P = 1.5$, $T = 0.25$, and a steady decrease in z^* for $R = 2$, $S = 0.5$, $P = 1.5$, $T = 1.25$, see their figure 5. This illustrates that relatedness (and thus spatial structure) plays an important role not only in the specific context of helping games but also in the more general context of nonlinear multiplayer games.

4 Discussion

We have shown that, when phenotypic differences are small, the selection gradient on a mixed strategy of a symmetric two-strategy n-player matrix game is proportional to the average inclusive payoff gain to an individual switching strategies, and that this can be written as a polynomial in Bernstein form (eq. (10)). As a result, convergence stability of strategies in spatially structured populations can be determined from the shape of the inclusive gain sequence (eq. (9)) and the mathematical properties of polynomials in Bernstein form (Farouki, 2012; Peña et al., 2014). We applied these results to the evolution of helping under synergies of scale and kind, and unified and extended previous analysis. The most important conclusion we reach is that, although an increase in (scaled) relatedness κ always tempers the social dilemma faced by cooperative individuals in a helping game, how the social dilemma is relaxed crucially depends on the synergies of kind and scale involved.

The simplest case is the one of whole-group traits (fig. 1a). Since there are no synergies of kind, only synergies of scale can introduce frequency dependent selection. For $\kappa \geq 0$, negative (resp. positive) synergies of scale induce negative (resp. positive) frequency-dependent selection. Moreover, increasing relatedness can transform a game in which defection is dominant (Prisoner's Dilemma) into a game in which cooperation and defection coexist (Snowdrift or anti-coordination game) when synergies of scale are negative (fig. 2.a), or into a game in which both cooperation and defection are stable (Stag Hunt or coordination game) when synergies of scale are positive (fig. 2.d).

More complex interactions between relatedness and frequency dependence can arise when there are both synergies of kind and scale. For nonexpresser-only traits (fig. 1.b), synergies of kind are negative and helping is altruistic, so that in the absence of relatedness defection dominates cooperation (as in a Prisoner's Dilemma). When synergies of scale are absent (linear benefits) or negative (diminishing incremental benefits), both the direct and the indirect effect are decreasing in z and selection is negative frequency-dependent. In this case, increasing relatedness might turn the game into a Snowdrift or an anti-coordination game, where the probability of cooperating is an increasing function of relatedness (fig. 2.b). Contrastingly, when synergies of scale are positive (increasing incremental benefits) the indirect effect may become unimodal in z. This paves the way for new patterns of evolutionary dynamics and bifurcations. For the particular case of geometric benefits, we find that there is a range of cost-to-benefit ratios such that, for sufficiently strong positive synergies of scale, increasing relatedness induces a saddle-node bifurcation whereby two internal equilibria appear, the leftmost unstable and the rightmost stable (fig. 2.e). After the bifurcation occurs, the evolutionary dynamics are characterized by bistable coexistence, where the first stable equilibrium is pure defection ($z = 0$) and the second is a mixed equilibrium in which individuals help with a positive probability (z_{R}).

For expresser-only traits (where synergies of kind are positive, fig. 1c) a similar interaction between relatedness and synergies occurs. When synergies of scale are absent or positive, $\mathcal{G}(z)$ is increasing in z for $\kappa \geq 0$. In this case, increasing relatedness might turn a scenario reminiscent of the Prisoner's Dilemma into a Stag Hunt or coordination game, where the size of the basin of attraction of the cooperative equilibrium is an increasing function of relatedness (fig. 2.f). Contrastingly, if synergies of scale are negative, relatedness may interact nontrivially with synergies to produce a dynamical outcome which is qualitatively identical to that arising from nonexpresser-only traits with positive synergies of scale, namely, bistable coexistence (fig. 2.c).

The three kinds of helping traits we considered are also different in the conditions they impose on the origin and the maintenance of helping. To see this, consider a payoff cost γ so large that the direct sequence d_k is negative. For the case of unrelated individuals ($\kappa = 0$) this implies that B dominates A so that $z = 0$ is the only stable strategy. We ask what happens when κ is increased and focus on the stability of the end-points $z = 0$ and $z = 1$.

For whole-group traits, the indirect gains from switching when co-players are all defectors (e_0) and when co-players are all helpers (e_{n-1}) are both positive. This opens up the opportunity for both (i) $z = 0$ to be destabilized if $\kappa > -d_0/e_0$, and (ii) $z = 1$ to be stabilized if $\kappa > -d_{n-1}/e_{n-1}$, which underlies the classical effect that increasing relatedness can destabilize defection and stabilize helping.

In contrast, one of these two scenarios is missing for nonexpresser-only and expresser-only traits. For nonexpresser-only traits, we have $e_0 > 0$ and $e_{n-1} = 0$ irrespectively of the shape of the benefit sequence. Hence, although $z = 0$ can be destabilized by increasing κ (allowing for some level of helping to be evolutionarily accessible from $z = 0$), $z = 1$ can never be stabilized and so full helping is never an evolutionary (convergent) stable point. Exactly the opposite happens for expresser-only traits, where $e_n > 0$ but $e_0 = 0$. As a result, $z = 1$ can become stable (if $\kappa > -d_{n-1}/e_{n-1}$) but $z = 0$ can never be destabilized by increasing κ. This implies that, under our assumptions, an expresser-only trait with high costs ($\gamma > \beta_n$) can never evolve from a monomorphic population of nonexpressers ($z = 0$), and this for any value of κ.

The kind of social trait also has a big impact on the amount of (scaled) relatedness required to make stable some level of helping. This quantitative effect is illustrated in figure 2. When synergies of scale are negative ($\lambda = 0.7$) and the cost-to benefit ratio is relatively low ($\gamma/\beta = 3.5$), for whole-group and nonexpresser-only traits, moderate amounts of relatedness ($\kappa \approx 0.132$ for whole-group traits, $\kappa \approx 0.185$ for nonexpresser-only traits) are sufficient for a non-zero level of expression of helping to be stable (fig. 2.*a* and 2.*b*). In contrast, a comparatively large amount of relatedness ($\kappa \approx 0.699$) is required for some non-zero level of helping to be stable if the trait is expresser-only (fig. 2.*c*). In the case of positive synergies of scale ($\lambda = 1.25$) and relatively high cost-to-benefit ratio ($\gamma/\beta = 15$), full expression of helping is stable already with $\kappa = 0$ for whole-group and expresser-only traits (fig. 2.*d* and 2.*f*). Contrastingly, for nonexpresser-only traits, a positive probability of expressing helping is stable only for large values of relatedness ($\kappa > \kappa^* \approx 0.548$, fig. 2.*e*).

We modeled social interactions by assuming that actions implemented by players are discrete. This is in contrast to many kin-selection models of games between relatives, which assume a continuum of pure actions in the form of continuous amounts of effort devoted to some social activity (e.g., Frank 1994; Johnstone et al. 1999; Reuter and Keller 2001; Wenseleers et al. 2010). Such continuous-action models have the advantage that the "fitness function" or "payoff function" (the counterpart to our eq. (3)) usually takes a simple form that facilitates mathematical analysis. On the other hand, there are situations where individuals can express only a few behavioral alternatives or morphs, such as worker and queen in the eusocial Hymenoptera (Wheeler, 1986), different behavioral tactics in foraging (e.g., "producers" and "scroungers" in house sparrows *Passer domesticus*; Barnard and Sibly, 1981) and hunting (e.g., lionesses positioned as "wings" and others positioned as "centres" in collective hunts; Stander, 1992), or distinct phenotypic states (e.g., capsulated and non-capsulated cells in *Pseudomonas fluorescens*; Beaumont et al., 2009). These situations are more conveniently modeled by means of a discrete-action model like the one presented here, but we expect that our qualitative results about the interaction between synergy and relatedness carry over to continuous-action models.

Synergistic interactions are likely to be much more common in nature than additive interactions where both synergies of scale and kind are absent. Given the local demographic structure of biological populations, interactions between relatives are also likely to be the rule rather than the exception. Empirical work should thus aim at measuring not only the genetic relatedness of interactants and the fitness costs and benefits of particular actions, but also at identifying the occurrences of positive and negative synergies of kind and scale, as it is the interaction between synergies and relatedness which determines the qualitative outcomes of the evolutionary dynamics of helping (fig. 2).

5 Acknowledgements

This work was partly supported by Swiss NSF Grants PBLAP3-145860 (to JP) and PP00P3-123344 (to LL).

A The haystack model

Many models of social interactions have assumed different versions of the haystack model (e.g., Matessi and Jayakar, 1976; Ackermann et al., 2008), where several rounds of unregulated reproduction can occur within groups before a round of complete dispersal (Maynard Smith, 1964) so that competition is effectively global. In these cases, as we will see below, κ takes the simpler interpretation of the coalescence probability of the gene lineage of two interacting individuals in their group. Here, we calculate κ for different variants of the haystack model.

The haystack model can be seen as a special case of the island model where dispersal is complete and where dispersing progeny compete globally. In this context, the fecundity of an adult is the number of its offspring reaching the stage of global density-dependent competition. The conception of offspring may occur in a single or over multiple rounds of reproduction, so that a growth phase within patches is possible. In this context, the number N of "adults" is better thought of as the number of founding individuals (or lineages, or seeds) on a patch.

Two cases need to be distinguished when it comes to social interactions. First, the game can be played between the adult individuals (founders) in which case

$$\kappa = 0, \tag{A.1}$$

since relatedness is zero among founders on a patch and there is no local competition. Alternatively, the game is played between offspring after reproduction and right before their dispersal. In this case two individuals can be related since they can descend from the same founder. Since there is no local competition, κ is directly the relatedness between two interacting offspring and is obtained as the probability that the two ancestral lineages of two randomly sampled offspring coalesce in the same founding individual (relatedness in the island model is defined as the cumulative coalescence probability over several generations, see e.g., Rousset, 2004, but owing to complete dispersal gene lineages can only coalesce in founders).

In order to evaluate κ for the second case, we assume that, after growth, exactly N_{o} offspring are produced and that the game is played between them ($n = N_{\mathrm{o}}$). Founding individuals, however, may contribute a variable number of offspring. Let us denote by O_i the random number of offspring descending from the "adult" individual $i = 1, 2, ..., N$ on a representative patch after reproduction, i.e., O_i is the size of lineage i. Owing to our assumption that the total number of offspring is fixed, we have $N_{\mathrm{o}} = O_1 + O_2 + ... + O_N$, where the O_i's are exchangeable random variables (i.e., neutral process, $\delta = 0$). The coalescence probability κ can then be computed as the expectation of the ratio of the total number of ways of sampling two offspring from the same founding parent to the total number of ways of sampling two offspring:

$$\kappa = \mathrm{E}\left[\sum_{i=1}^{N} \frac{O_i(O_i - 1)}{N_{\mathrm{o}}(N_{\mathrm{o}} - 1)}\right] = N\left(\frac{\sigma^2 + \mu^2 - \mu}{N_{\mathrm{o}}(N_{\mathrm{o}} - 1)}\right), \tag{A.2}$$

where the second equality follows from exchangeability, $\mu = \mathrm{E}[O_i]$ is the expected number of offspring descending from any individual i, and $\sigma^2 = \mathrm{E}\left[(O_i - \mu)^2\right]$ is the corresponding variance. Due to the fact that the total number of offspring is fixed, we also necessarily have $\mu = N_{\mathrm{o}}/N$ (i.e., $N_{\mathrm{o}} = \mathrm{E}[N_{\mathrm{o}}] = \mathrm{E}[O_1 + O_2 + ... + O_N] = N\mu$), whereby

$$\kappa = \frac{N_{\mathrm{o}} - N}{N(N_{\mathrm{o}} - 1)} + \frac{\sigma^2 N}{N_{\mathrm{o}}(N_{\mathrm{o}} - 1)}, \tag{A.3}$$

which holds for any neutral growth process.

We now consider different cases:

(i) Suppose that there is no variation in offspring production between founding individuals, as in the life cycle described by Ackermann et al. (2008). Then $\sigma^2 = 0$, and equation (A.3) simplifies to

$$\kappa = \frac{(N_\mathrm{o} - N)}{N(N_\mathrm{o} - 1)}. \tag{A.4}$$

(ii) Suppose that each of the N_o offspring has an equal chance of descending from any founding individual, so that each offspring is the result of a sampling event (with replacement) from a parent among the N founding individuals. Then, the offspring number distribution is binomial with parameters N_o and $1/N$, whereby $\sigma^2 = (1 - 1/N)N_\mathrm{o}/N$. Substituting into equation (A.3) produces

$$\kappa = \frac{1}{N}. \tag{A.5}$$

In more biological terms, this case results from a situation where individuals produce offspring according to a Poisson process and where exactly N_o individuals are kept for interactions (i.e., the conditional branching process of population genetics; Ewens, 2004).

(iii) Suppose that the offspring distribution follows a beta-binomial distribution, with number of trials N_o and shape parameters $\alpha > 0$ and $\beta = \alpha(N - 1)$. Then, $\mu = N_\mathrm{o}/N$ and

$$\sigma^2 = \frac{N_\mathrm{o}(N-1)(\alpha N + N_\mathrm{o})}{N^2(1 + \alpha N)},$$

which yields

$$\kappa = \frac{1 + \alpha}{1 + \alpha N}. \tag{A.6}$$

In more biological terms, this reproductive scheme results from a situation where individuals produce offspring according to a negative binomial distribution (larger variance than Poisson, which is recovered when $\alpha \to \infty$), and where exactly N_o individuals are kept for interactions.

B Gains from switching and the gain function

In the following we establish the expressions for $\mathcal{C}(z)$ and $\mathcal{B}(z)$ given in equations (7)–(8); equation (10) is then immediate from the definition of f_k (9) and the identity $\mathcal{G}(z) = -\mathcal{C}(z) + \kappa\mathcal{B}(z)$.

Recalling the definitions of $\mathcal{C}(z)$ and $\mathcal{B}(z)$ from equation (4) as well as the definitions of d_k and e_k from equations (5)–(6) we need to show

$$\left.\frac{\partial \pi(z_\bullet, z_\circ)}{\partial z_\bullet}\right|_{z_\bullet = z_\circ = z} = \sum_{k=0}^{n-1} \binom{n-1}{k} z^k (1-z)^{n-1-k} \left[a_k - b_k\right], \tag{B.1}$$

$$\left.\frac{\partial \pi(z_\bullet, z_\circ)}{\partial z_\circ}\right|_{z_\bullet = z_\circ = z} = \sum_{k=0}^{n-1} \binom{n-1}{k} z^k (1-z)^{n-1-k} \left[k\Delta a_{k-1} + (n-1-k)\Delta b_k\right], \tag{B.2}$$

where the function π has been defined in equation (3). Equation (B.1) follows directly by taking the partial derivative of π with respect to $z_\bullet$ and evaluating at $z_\bullet = z_\circ = z$, so it remains to establish equation (B.2).

Our derivation of equation (B.2) uses properties of polynomials in Bernstein form (Farouki, 2012).

Such polynomials, which in general can be written as $\sum_{k=0}^{m}\binom{m}{k}x^k(1-x)^{m-k}c_k$, where $x\in[0,1]$, satisfy

$$\frac{\mathrm{d}}{\mathrm{d}x}\sum_{k=0}^{m}\binom{m}{k}x^k(1-x)^{m-k}c_k = m\sum_{k=0}^{m-1}\binom{m-1}{k}x^k(1-x)^{m-1-k}\Delta c_k.$$

Applying this property to equation (3) and evaluating the resulting partial derivative at $z_\bullet = z_\circ = z$, yields

$$\left.\frac{\partial\pi(z_\bullet,z_\circ)}{\partial z_\circ}\right|_{z_\bullet=z_\circ=z} = (n-1)z\sum_{k=0}^{n-2}\binom{n-2}{k}z^k(1-z)^{n-2-k}\Delta a_k + (n-1)(1-z)\sum_{k=0}^{n-2}\binom{n-2}{k}z^k(1-z)^{n-2-k}\Delta b_k. \tag{B.3}$$

In order to obtain equation (B.2) from equation (B.3) it then suffices to establish

$$x\sum_{k=0}^{m-1}\binom{m-1}{k}x^k(1-x)^{m-1-k}c_k = \sum_{k=0}^{m}\binom{m}{k}x^k(1-x)^{m-k}\frac{kc_{k-1}}{m} \tag{B.4}$$

and

$$(1-x)\sum_{k=0}^{m-1}\binom{m-1}{k}x^k(1-x)^{m-1-k}c_k = \sum_{k=0}^{m}\binom{m}{k}x^k(1-x)^{m-k}\frac{(m-k)c_k}{m}, \tag{B.5}$$

as applying these identities to the terms on the right side of equation (B.3) yields the right side of equation (B.2).

Let us prove equation (B.4) (eq. (B.5) is proven in a similar way). Starting from the left side of equation (B.4), we multiply and divide by $m/(k+1)$ and distribute x to obtain

$$x\sum_{k=0}^{m-1}\binom{m-1}{k}x^k(1-x)^{m-1-k}c_k = \sum_{k=0}^{m-1}\frac{m}{k+1}\binom{m-1}{k}x^{k+1}(1-x)^{m-(k+1)}c_k\frac{k+1}{m}.$$

Applying the identity $\binom{r}{k}=\frac{r}{k}\binom{r-1}{k-1}$ and changing the index of summation to $k=k+1$, we get

$$x\sum_{k=0}^{m-1}\binom{m-1}{k}x^k(1-x)^{m-1-k}c_k = \sum_{k=1}^{m}\binom{m}{k}x^k(1-x)^{m-k}\frac{kc_{k-1}}{m}.$$

Finally, changing the lower index of the sum by noting that the summand is zero when $k=0$ gives equation (B.4).

C Whole-group traits with geometric benefits

With geometric benefits, we have $\Delta\beta_k = \beta\lambda^k$, so that the inclusive gains from switching for whole-group traits are given by $f_k = -\gamma + [1+\kappa(n-1)]\,\beta\lambda^k$. Using the formula for the probability generating function of a binomial random variable, equation (10) can be written as

$$\mathcal{G}(z) = -\gamma + [1+\kappa(n-1)]\,\beta\,(1-z+\lambda z)^{n-1}. \tag{C.1}$$

As $\mathcal{G}(z)$ is either decreasing ($\lambda<1$) or increasing ($\lambda>1$) in z, A (resp. B) is a dominant strategy if and only if $\min[\mathcal{G}(0),\mathcal{G}(1)]\geq 0$ (resp. if and only if $\max[\mathcal{G}(0),\mathcal{G}(1)]\leq 0$). Using equation (C.1) to calculate $\mathcal{G}(0)$ and $\mathcal{G}(1)$ then yields the critical cost-to-benefit ratios $\varepsilon = \min[\mathcal{G}(0),\mathcal{G}(1)]$ and $\vartheta = \max[\mathcal{G}(0),\mathcal{G}(1)]$

given in equation (17). The value of z^* given in equation (18) is obtained by solving $\mathcal{G}(z^*) = 0$.

D Other-only traits

In contrast to what happens in whole-group traits, individuals expressing an other-only trait are automatically excluded from the consumption of the good they create, although they can still reap the benefits of goods created by other expressers in their group. Payoffs for such other-only traits are given by $a_k = -\gamma + \beta_k$ and $b_k = \beta_k$, so that the inclusive gains from switching are given by $f_k = -\gamma + \kappa\left[k\Delta\beta_{k-1} + (n-1-k)\Delta\beta_k\right]$. For this payoff constellation, it is straightforward to obtain the indirect benefits $\mathcal{B}(z)$ from equation (B.3) in appendix B. Observing that $\Delta a_k = \Delta b_k = \Delta\beta_k$ holds for all k, we have

$$\mathcal{B}(z) = \left.\frac{\partial\pi(z_\bullet, z_\circ)}{\partial z_\circ}\right|_{z_\bullet = z_\circ = z} = \sum_{k=0}^{n-2}\binom{n-2}{k} z^k(1-z)^{n-2-k}(n-1)\Delta\beta_k.$$

Using equation (7) and the equality $a_k - b_k = -\gamma$, we have that the direct benefit is given by $-\mathcal{C}(z) = -\gamma$. Substituting these expressions for $\mathcal{C}(z)$ and $\mathcal{B}(z)$ into equation (4), we obtain

$$\mathcal{G}(z) = \sum_{k=0}^{n-2}\binom{n-2}{k} z^k(1-z)^{n-2-k}\left[-\gamma + \kappa(n-1)\Delta\beta_k\right]. \tag{D.1}$$

If $\kappa \leq 0$, our assumption that the benefit sequence is increasing implies that $\mathcal{G}(z)$ is always negative, so that $z = 0$ is the only stable point (defection dominates cooperation).

To analyze the case where $\kappa > 0$, it is convenient to observe that equation (D.1) is of a similar form as equation (15). The only differences are that the summation in equation (D.1) extends from 0 to $n-2$ (rather than $n-1$) and that the term multiplying the incremental benefit $\Delta\beta_k$ is given by $\kappa(n-1)$ (rather than $1 + \kappa(n-1)$). All the results obtained for whole-group traits can thus be easily translated to the case of other-only traits.

Specifically, we have the following characterization of the resulting evolutionary dynamics. In the absence of synergies of scale, selection is frequency-independent with defection dominating cooperation if $\kappa < \gamma/[(n-1)\beta]$ and cooperation dominating defection if $\kappa > \gamma/[(n-1)\beta]$. With negative synergies of scale, the gain function is decreasing in z (negative frequency dependence). Defection dominates cooperation (so that $z = 0$ is the only convergent stable strategy) if $\gamma \geq \kappa(n-1)\Delta\beta_0$, whereas cooperation dominates defection if $\gamma \leq \kappa(n-1)\Delta\beta_{n-2}$ holds. If $\kappa(n-1)\Delta\beta_{n-2} < \gamma < \kappa(n-1)\Delta\beta_0$ holds, the unique convergent stable strategy z^* features coexistence. With positive synergies of scale, the gain function is increasing in z (positive frequency dependence). Defection dominates cooperation if $\gamma \geq \kappa(n-1)\Delta\beta_{n-2}$, whereas cooperation dominates defection if $\gamma \leq \kappa(n-1)\Delta\beta_0$. If $\kappa(n-1)\Delta\beta_0 < \gamma < \kappa(n-1)\Delta\beta_{n-2}$, there is bistability: both $z = 0$ and $z = 1$ are stable and there is a unique, unstable interior point z^* separating the basins of attraction of these two stable strategies.

When benefits are geometric (13), the gain function is given by

$$\mathcal{G}(z) = -\gamma + \kappa(n-1)\beta(1 - z + \lambda z)^{n-2},$$

so that, for $\lambda \neq 1$, the evolutionary dynamics are similar to the case of whole-group traits after redefining the critical cost-to-benefit ratios as

$$\varepsilon = \min\left(\kappa(n-1), \lambda^{n-2}\kappa(n-1)\right) \qquad \text{and} \qquad \vartheta = \max\left(\kappa(n-1), \lambda^{n-2}\kappa(n-1)\right)$$

and letting

$$z^* = \frac{1}{1-\lambda}\left[1 - \left(\frac{\gamma}{\beta\kappa(n-1)}\right)^{\frac{1}{n-2}}\right].$$

E Nonexpresser-only traits

For nonexpresser-only traits, the inclusive gains from switching are given by

$$f_k = -\gamma - \beta_k + \kappa(n-1-k)\Delta\beta_k. \tag{E.1}$$

E.1 Negative synergies of scale

When synergies of scale are negative, we have the following general result.

Result 1 (Nonexpresser-only traits with negative synergies of scale)**.** *Let f_k be given by equation* (E.1) *with $\beta_0 = 0$, β_k increasing and $\Delta\beta_k$ decreasing in k and let $\kappa \geq 0$ (the case $\kappa < 0$ is trivial). Then*

1. *If $\gamma \geq \kappa(n-1)\Delta\beta_0$, $z = 0$ is the only stable point (B dominates A).*
2. *If $\gamma < \kappa(n-1)\Delta\beta_0$, both $z = 0$ and $z = 1$ are unstable and there is a unique internal stable point $z^* \in (0,1)$ (coexistence).*

To prove this result, we start by observing that the assumptions in the statement imply that f_k is decreasing in k. In particular, we have $f_{n-1} < f_0$. Consequently, if $f_0 \leq 0$ (which holds if and only if $\gamma \geq \kappa(n-1)\Delta\beta_0$) the inclusive gain sequence has no sign changes and its initial sign is negative. Observing that $f_{n-1} = -\gamma - \beta_{n-1} < 0$ always holds true, the inequality $f_0 > 0$ (which holds if and only if $\gamma < \kappa(n-1)\Delta\beta_0$) implies that the decreasing sequence f_k has one sign change and that its initial sign is positive. Result 1 is then obtained by an application of Peña et al. (2014, result 3).

E.2 Geometric benefits

For geometric benefits, we obtain the following result.

Result 2 (Nonexpresser-only traits with geometric benefits)**.** *Let f_k be given by equation* (E.1) *with β_k given by equation* (13) *and let $\kappa \geq 0$ and $n > 2$ (the cases $\kappa < 0$ or $n = 2$ are trivial). Moreover, let ϱ, ζ and η be defined by equations* (19) *and* (20)*. Then*

1. *If $\lambda \leq \varrho$, $\mathcal{G}(z)$ is nonincreasing in z. Furthermore:*
   - *(a) If $\gamma/\beta < \zeta$, both $z = 0$ and $z = 1$ are unstable and there is a unique internal stable point $z^* \in (0,1)$ (coexistence).*
   - *(b) If $\gamma/\beta \geq \zeta$, $z = 0$ is the only stable point (B dominates A).*
2. *If $\lambda > \varrho$, $\mathcal{G}(z)$ is unimodal in z with mode given by $\hat{z} = \frac{\kappa[(n-2)\lambda-(n-1)]-1}{[1+\kappa(n-1)](\lambda-1)}$. Furthermore:*
   - *(a) If $\gamma/\beta \leq \zeta$, both $z = 0$ and $z = 1$ are unstable and there is a unique internal stable point $z^* > \hat{z}$ (coexistence).*
   - *(b) If $\zeta < \gamma/\beta < \eta$, there are two interior singular points z_{L} and z_{R} satisfying $z_{\mathrm{L}} < \hat{z} < z_{\mathrm{R}}$. The points $z = 0$ and z_{R} are stable, whereas z_{L} and $z = 1$ are unstable (bistable coexistence).*
   - *(c) If $\gamma/\beta \geq \eta$, then $z = 0$ is the only stable point (B dominates A).*

Observing that $\varrho > 1$ holds for $\kappa \geq 0$ and that the case $\lambda = 1$ (no synergies of scale) is trivial, we can prove this result by considering three cases: (i) $\lambda < 1$, (ii) $1 < \lambda \leq \varrho$, and (iii) $\varrho < \lambda$.

For $\lambda < 1$, we have negative synergies of scale and hence result 1 applies with $\Delta\beta_0 = \beta$. Recalling the definition of $\zeta = \kappa(n-1)$ from equation (20) and rearranging, this yields result 2.1 for the case $\lambda \leq 1 < \varrho$.

To obtain the result for the remaining two cases, we calculate the first and second forward differences of the benefit sequence (13) and substitute them into

$$\Delta f_k = -(1+\kappa)\Delta\beta_k + \kappa(n-2-k)\Delta^2\beta_k,\ k = 0, 1, \ldots, n-2.$$

to obtain

$$\Delta f_k = \beta\lambda^k \left\{\kappa\left[(n-2)\lambda - (n-1)\right] - 1 + \kappa(1-\lambda)k\right\},\ k = 0, 1, \ldots, n-2.$$

For $\lambda > 1$, the sequence Δf_k is decreasing in k and hence can have at most one sign change. Moreover, since $\Delta f_{n-2} = -\beta\lambda^{n-2}(1+\kappa) < 0$ always holds true, the sign pattern of Δf_k depends exclusively on how $\Delta f_0 = \beta\left\{\kappa\left[(n-2)\lambda - (n-1)\right] - 1\right\}$ compares to zero. Observe, too, that $f_{n-1} < 0$ always holds true and that the sign of f_0 is identical to the sign of $\zeta - \gamma/\beta$.

Consider the case $1 < \lambda \leq \varrho$. Recalling the definition of ϱ (eq. (19)) we then have $\Delta f_0 \leq 0$, implying that Δf_k has no sign changes and that its initial sign is negative, i.e., f_k is nonincreasing. Hence, if $f_0 \leq 0$ (which holds if and only if $\gamma/\beta \geq \zeta$), the inclusive gain sequence has no sign changes and its initial sign is negative. Otherwise, that is, if $\gamma/\beta < \zeta$ holds, we have $f_0 > 0 > f_{n-1}$ so that the inclusive gain sequence has one sign change and its initial sign is positive. Result 2.1 then follows from Peña et al. (2014, result 3).

For $\lambda > \varrho$ we have $\Delta f_0 > 0$, implying that Δf_k has one sign change from $+$ to $-$, i.e., f_k is unimodal. This implies that the gain function $\mathcal{G}(z)$ is also unimodal with its mode $\hat{z}$ being determined by $\mathcal{G}'(\hat{z}) = 0$ (Peña et al., 2014, section 3.4.3). Using the assumption of geometric benefits, we can express $\mathcal{G}(z)$ is closed form as

$$\mathcal{G}(z) = -\gamma + \frac{\beta}{\lambda-1} + \beta\left\{\kappa(n-1) - \frac{1}{\lambda-1} - \left[1+\kappa(n-1)\right]z\right\}(1-z+\lambda z)^{n-2}$$

with corresponding derivative

$$\mathcal{G}'(z) = (n-1)\beta(\lambda-1)(1-z+\lambda z)^{n-3}\left\{\kappa(n-2) - \frac{1+\kappa}{\lambda-1} - \left[1+\kappa(n-1)\right]z\right\}.$$

Solving $\mathcal{G}'(\hat{z}) = 0$ then yields $\hat{z}$ as given in equation (21). The corresponding maximal value of the gain function is

$$\mathcal{G}(\hat{z}) = -\gamma + \frac{\beta}{\lambda-1}\left[1 + \kappa\lambda\left(\frac{(n-2)\kappa\lambda}{1+\kappa(n-1)}\right)^{n-2}\right].$$

Result 2.2 follows from an application of Peña et al. (2014, result 5) upon noticing that $f_0 \geq 0$ (precluding the stability of $z = 0$ and ensuring $\mathcal{G}(\hat{z}) > 0$) holds if and only if $\gamma/\beta \leq \zeta$ and that $\mathcal{G}(\hat{z}) \leq 0$ (ensuring that B dominates A) is satisfied if and only if $\gamma/\beta \geq \eta$. (We note that the range of cost-to-benefit ratios γ/β for which bistable coexistence occurs is nonempty, that is $\eta > \zeta$ holds. Otherwise there would exist γ/β satisfying both $\gamma/\beta \leq \zeta$ and $\gamma/\beta \geq \eta$ which in light of result 2.2.(a) and result 2.2.(c) is clearly impossible.)

F Expresser-only traits

For expresser-only traits, the inclusive gains from switching are given by

$$f_k = -\gamma + \beta_{k+1} + \kappa k \Delta\beta_k. \tag{F.1}$$

F.1 Positive synergies of scale

In the case of positive synergies of scale, we have the following general result.

Result 3 (Expresser-only traits with positive synergies of scale)**.** *Let f_k be given by equation* (F.1) *with β_k and $\Delta\beta_k$ increasing in k and let $\kappa \geq 0$. Then*

1. *If $\gamma \leq \beta_1$, $z = 1$ is the only stable point (A dominates B).*
2. *If $\beta_1 < \gamma < \beta_n + \kappa(n-1)\Delta\beta_{n-1}$, both $z = 0$ and $z = 1$ are stable and there is a unique internal unstable point $z^* \in (0, 1)$ (bistability).*
3. *If $\gamma \geq \beta_n + \kappa(n-1)\Delta\beta_{n-1}$, $z = 0$ is the only stable point (B dominates A).*

The arguments used for deriving this result are analogous to those used for deriving the results for the case of nonexpresser-only traits and negative synergies of scale (result 1 in app. E). The assumptions in the statement of the result imply that f_k is increasing in k. In particular, we have $f_0 < f_{n-1}$. The sign pattern of the inclusive gain sequence thus depends on the values of its endpoints in the following way. If $f_0 \geq 0$ (which holds if and only if $\gamma \leq \beta_1$), f_k has no sign changes and a positive initial sign. If $f_{n-1} \leq 0$ (which holds if and only if $\gamma \geq \beta_n + \kappa(n-1)\Delta\beta_{n-1}$), f_k has no sign changes and a negative initial sign. If $f_0 < 0 < f_{n-1}$ (which holds if and only if $\beta_1 < \gamma < \beta_n + \kappa(n-1)\Delta\beta_{n-1}$) f_k has one sign change and a negative initial sign. Result 3 follows from these observations upon applying Peña et al. (2014, result 3).

F.2 Geometric benefits

For geometric benefits, we obtain the following result.

Result 4 (Expresser-only traits with geometric benefits)**.** *Let f_k be given by equation* (F.1) *with β_k given by equation* (13) *and let $\kappa \geq 0$ and $n > 2$ (the cases $\kappa < 0$ or $n = 2$ are trivial). Moreover, let ξ, ς and τ be defined by equations* (22) *and* (23)*. Then*

1. *If $\lambda \geq \xi$, $\mathcal{G}(z)$ is nondecreasing in z. Furthermore*
   - *(a) If $\gamma/\beta \leq 1$, $z = 1$ is the only stable point (A dominates B).*
   - *(b) If $1 < \gamma/\beta < \varsigma$, both $z = 0$ and $z = 1$ are stable and there is a unique internal unstable point $z^* \in (0, 1)$ (bistability).*
   - *(c) If $\gamma/\beta \geq \varsigma$, $z = 0$ is the only stable point (B dominates A).*
2. *If $\lambda < \xi$, $\mathcal{G}(z)$ is unimodal in z, with mode given by $\hat{z} = \frac{1+\kappa}{[1+\kappa(n-1)](1-\lambda)}$. Furthermore*
   - *(a) If $\gamma/\beta \leq 1$, $z = 1$ is the only stable point (A dominates B).*
   - *(b) If $1 < \gamma/\beta \leq \varsigma$, both $z = 0$ and $z = 1$ are stable and there is a unique internal unstable point $z^* \in (0, \hat{z})$ (bistability).*
   - *(c) If $\varsigma < \gamma/\beta < \tau$, there are two interior singular points z_{L} and z_{R} satisfying $z_{\mathrm{L}} < \hat{z} < z_{\mathrm{R}}$. The points $z = 0$ and z_{R} are stable, whereas z_{L} and $z = 1$ are unstable (bistable coexistence).*
   - *(d) If $\gamma/\beta \geq \tau$, $z = 0$ is the only stable point (B dominates A).*

The arguments used for deriving this result are analogous to those used for deriving the results for nonexpresser-only traits with geometric benefits (result 2 in app. E). Observing that $\xi < 1$ holds and ignoring the trivial case $\lambda = 1$, there are three cases to consider: (i) $\lambda > 1$, (ii) $1 > \lambda \geq \xi$, and (iii) $\xi > \lambda$.

For $\lambda > 1$ synergies of scale are positive and hence result 3 applies with $\beta_1 = \beta$ and $\beta_n + \kappa(n-1)\Delta\beta_{n-1} = \beta\varsigma$. This yields result 4.1 for the case $\lambda > 1$.

To obtain the results for the remaining two cases, we calculate the first and second forward differences of the benefit sequence (13) and substitute them into

$$\Delta f_k = \Delta\beta_{k+1} + \kappa\left\{(k+1)\Delta^2\beta_k + \Delta\beta_k\right\},\ k = 0, 1, \ldots, n-2,$$

to obtain

$$\Delta f_k = \beta\lambda^k\left[\lambda(1+\kappa) + \kappa(\lambda-1)k\right],\ k = 0, 1, \ldots, n-2.$$

For $\lambda < 1$, the sequence Δf_k is decreasing in k and hence can have at most one sign change. Moreover, as $\Delta f_0 = \beta\lambda(1+\kappa) > 0$ always holds true, the initial sign of Δf_k is positive and whether or not the sequence Δf_k has a sign change depends solely on how Δf_{n-2} compares to zero. Observe, too, that for $\lambda < 1$ we have $\varsigma > 1$ as $\lambda^n < \lambda$ holds.

Consider the case $\xi \leq \lambda < 1$. By the definition of ξ (eq. (22)) this implies $\Delta f_{n-2} \geq 0$. In this case Δf_k has no sign changes and f_k is nondecreasing. The sign pattern of the inclusive gain sequence can then be determined by looking at how the signs of its endpoints depend on the cost-to-benefit ratio γ/β. If $\gamma/\beta \leq 1$, then $f_0 \geq 0$, implying that f_k has no sign changes and its initial sign is positive. If $\gamma/\beta \geq \varsigma$, then $f_n \leq 0$ and hence f_k has no sign changes and its initial sign is negative. If $1 < \gamma/\beta < \varsigma$, then $f_0 < 0 < f_n$, i.e., f_k has one sign change and its initial sign is negative. Result 4.1 then follows from an application of Peña et al. (2014, result 3).

For $\lambda < \xi$ we have $\Delta f_{n-2} < 0$, implying that Δf_k has one sign change from $+$ to $-$, i.e., f_k is unimodal. Hence, the gain function $\mathcal{G}(z)$ is also unimodal (Peña et al., 2014, section 3.4.3) with mode $\hat{z}$ determined by $\mathcal{G}'(\hat{z}) = 0$. Using the assumption of geometric benefits, we can express $\mathcal{G}(z)$ is closed form as

$$\mathcal{G}(z) = -\gamma + \frac{\beta}{1-\lambda} + \beta\lambda\left\{\left[1+\kappa(n-1)\right]z - \frac{1}{1-\lambda}\right\}(1-z+\lambda z)^{n-2},$$

with corresponding derivative

$$\mathcal{G}'(z) = (n-1)\beta\lambda\left\{1+\kappa-(1-\lambda)\left[1+\kappa(n-1)\right]z\right\}(1-z+\lambda z)^{n-3}.$$

Solving $\mathcal{G}'(\hat{z}) = 0$ then yields $\hat{z}$ as given in equation (24). The corresponding maximal value of the gain function is

$$\mathcal{G}(\hat{z}) = -\gamma + \frac{\beta}{1-\lambda}\left[1 + \lambda\kappa\left(\frac{(n-2)\kappa}{1+\kappa(n-1)}\right)^{n-2}\right].$$

Result 4.2 then follows from applying Peña et al. (2014, result 5). In particular, if $\gamma/\beta \leq 1$, we also have $\gamma/\beta < \varsigma$, ensuring that $f_0 \geq 0$ and $f_{n-1} > 0$ hold (with unimodality then implying that the gain function is positive throughout). If $1 < \gamma/\beta \leq \varsigma$, we have $f_0 < 0$ and $f_{n-1} \geq 0$ (with unimodality then implying $\mathcal{G}(\hat{z}) > 0$). If $\varsigma < \gamma/\beta$, we have $f_0 < 0$ and $f_{n-1} < 0$. Upon noticing that $\mathcal{G}(\hat{z}) \leq 0$ is satisfied if and only if $\gamma/\beta \geq \tau$ holds, this yields the final two cases in result 4.2.

References

Ackermann, M., Stecher, B., Freed, N. E., Songhet, P., Hardt, W.-D., Doebeli, M., 2008. Self-destructive cooperation mediated by phenotypic noise. Nature 454 (7207), 987–990.

Akçay, E., Van Cleve, J., 2012. Behavioral responses in structured populations pave the way to group optimality. American Naturalist 179 (2), 257–269.

Archetti, M., 2009. The volunteer's dilemma and the optimal size of a social group. Journal of Theoretical Biology 261 (3), 475–480.

Archetti, M., Scheuring, I., 2011. Coexistence of cooperation and defection in public goods games. Evolution 65 (4), 1140–1148.

Axelrod, R., Hamilton, W., 1981. The evolution of cooperation. Science 211 (4489), 1390–1396.

Bach, L., Helvik, T., Christiansen, F., 2006. The evolution of n-player cooperation–threshold games and ESS bifurcations. Journal of Theoretical Biology 238 (2), 426–434.

Barnard, C., Sibly, R., 1981. Producers and scroungers: a general model and its application to captive flocks of house sparrows. Animal Behaviour 29 (2), 543–550.

Beaumont, H. J. E., Gallie, J., Kost, C., Ferguson, G. C., Rainey, P. B., 2009. Experimental evolution of bet hedging. Nature 462 (7269), 90–93.

Bernasconi, G., Strassmann, J. E., 1999. Cooperation among unrelated individuals: the ant foundress case. Trends in Ecology & Evolution 14 (12), 477–482.

Bourke, A., Franks, N., 1995. Social Evolution in Ants. Princeton University Press, Princeton, NJ.

Boyd, R., Richerson, P. J., 1988. The evolution of reciprocity in sizable groups. Journal of Theoretical Biology 132 (3), 337–356.

Champagnat, N., Ferrière, R., Méléard, S., 2006. Unifying evolutionary dynamics: from individual stochastic processes to macroscopic models. Theoretical Population Biology 69 (3), 297–321.

Christiansen, F. B., 1991. On conditions for evolutionary stability for a continuously varying character. American Naturalist 138 (1), 37–50.

Clobert, J., Danchin, E., Dhondt, A. A., Nichols, J. D. (Eds.), 2001. Dispersal. Oxford University Press.

Clutton-Brock, T. H., O'Riain, M. J., Brotherton, P. N. M., Gaynor, D., Kansky, R., Griffin, A. S., Manser, M., 1999. Selfish sentinels in cooperative mammals. Science 284 (5420), 1640–1644.

Corning, P. A., 2002. The re-emergence of "emergence": a venerable concept in search of a theory. Complexity 7 (6), 18–30.

Cressman, R., 2003. Evolutionary dynamics and extensive form games. MIT Press, Cambridge, MA.

Dawkins, R., 1999. The extended phenotype: the long reach of the gene. Oxford University Press, Oxford, UK.

Day, T., Taylor, P. D., May 1997. Hamilton's rule meets the Hamiltonian: kin selection on dynamic characters. Proceedings of the Royal Society B: Biological Sciences 264 (1382), 639–644.

Day, T., Taylor, P. D., 1998. Unifying genetic and game theoretic models of kin selection for continuous traits. Journal of Theoretical Biology 194 (3), 391–407.

Diekmann, A., 1985. Volunteer's dilemma. Journal of Conflict Resolution 29 (4), 605–610.

Eshel, I., 1983. Evolutionary and continuous stability. Journal of Theoretical Biology 103 (1), 99–111.

Eshel, I., Motro, U., 1981. Kin selection and strong evolutionary stability of mutual help. Theoretical Population Biology 19 (3), 420–433.

Eshel, I., Motro, U., 1988. The three brothers' problem: kin selection with more than one potential helper. 1. The case of immediate help. American Naturalist 132 (4), 550–566.

Ewens, W. J., 2004. Mathematical Population Genetics. Springer-Verlag, New York, NY.

Farouki, R. T., 2012. The Bernstein polynomial basis: a centennial retrospective. Computer Aided Geometric Design 29 (6), 379–419.

Fletcher, J. A., Zwick, M., 2006. Unifying the theories of inclusive fitness and reciprocal altruism. American Naturalist 168 (2), 252–262.

Frank, S. A., 1994. Kin selection and virulence in the evolution of protocells and parasites. Proceedings of the Royal Society B: Biological Sciences 258 (1352), 153–161.

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

Frank, S. A., 2006. Social selection. Oxford University Press, Oxford, UK, Ch. 23, pp. 350–363.

Fröhlich, K.-U., Madeo, F., 2000. Apoptosis in yeast – a monocellular organism exhibits altruistic behaviour. FEBS Letters 473 (1), 6–9.

Gardner, A., Kümmerli, R., 2008. Social evolution: this microbe will self-destruct. Current Biology 18 (21), R1021–R1023.

Gardner, A., West, S. A., 2006. Demography, altruism, and the benefits of budding. Journal of Evolutionary Biology 19, 1707–1716.

Gardner, A., West, S. A., 2010. Greenbeards. Evolution 64 (1), 25–38.

Gardner, A., West, S. A., Wild, G., 2011. The genetical theory of kin selection. Journal of Evolutionary Biology 24 (5), 1020–1043.

Geritz, S. A. H., Kisdi, E., Meszéna, G., Metz, J. A. J., 1998. Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. Evolutionary Ecology 12, 35–57.

Gillespie, J. H., 1991. The causes of molecular evolution. Oxford University P, New York, NY.

Godfrey-Smith, P., Kerr, B., 2009. Selection in ephemeral networks. American Naturalist 174 (6), 906–911.

Gokhale, C. S., Traulsen, A., 2010. Evolutionary games in the multiverse. Proceedings of the National Academy of Sciences 107 (12), 5500–5504.

Gore, J., Youk, H., van Oudenaarden, A., 2009. Snowdrift game dynamics and facultative cheating in yeast. Nature 459 (7244), 253–256.

Grafen, A., 1979. The hawk-dove game played between relatives. Animal Behaviour 27, Part 3 (0), 905–907.

Greig, D., Travisano, M., 2004. The prisoner's dilemma and polymorphism in yeast suc genes. Proceedings of the Royal Society B: Biological Sciences 271 (Suppl 3), S25–S26.

Guilford, T., 1985. Is kin selection involved in the evolution of warning coloration? Oikos 45 (1), 31–36.

Guilford, T., 1988. The evolution of conspicuous coloration. American Naturalist 131, S7–S21.

Hamilton, W., 1964a. The genetical evolution of social behaviour. II. Journal of Theoretical Biology 7 (1), 17–52.

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

Hamilton, W. D., 1970. Selfish and spiteful behaviour in an evolutionary model. Nature 228 (5277), 1218–1220.

Hamilton, W. D., 1971. Selection of selfish and altruistic behavior in some extreme models. In: Eisenberg, J. F., Dillon, W. S. (Eds.), Man and Beast: Comparative Social Behavior. Smithsonian Press, Washington DC, pp. 57–91.

Hauert, C., Michor, F., Nowak, M. A., Doebeli, M., 2006. Synergy and discounting of cooperation in social dilemmas. Journal of Theoretical Biology 239 (2), 195–202.

Hines, W., Maynard Smith, J., 1979. Games between relatives. Journal of Theoretical Biology 79 (1), 19–30.

Johnstone, R. A., Woodroffe, R., Cant, M., Wright, J., 1999. Reproductive skew in multimember groups. American Naturalist 153, 315–331.

Kurokawa, S., Ihara, Y., 2009. Emergence of cooperation in public goods games. Proceedings of the Royal Society B: Biological Sciences 276 (1660), 1379–1384.

Lehmann, L., Keller, L., 2006a. The evolution of cooperation and altruism – a general framework and a classification of models. Journal of Evolutionary Biology 19, 1365–1376.

Lehmann, L., Keller, L., 2006b. Synergy, partner choice and frequency dependence: their integration into inclusive fitness theory and their interpretation in terms of direct and indirect fitness effects. Journal of Evolutionary Biology 19 (5), 1426–1436.

Lehmann, L., Perrin, N., Rousset, F., 2006. Population demography and the evolution of helping behaviors. Evolution 60, 1137–1151.

Lehmann, L., Rousset, F., 2010. How life history and demography promote or inhibit the evolution of helping behaviours. Philosophical Transactions of the Royal Society B: Biological Sciences 365 (1553), 2599–2617.

Lehmann, L., Rousset, F., 2012. The evolution of social discounting in hierarchically clustered populations. Molecular Ecology 21 (3), 447–471.

Leimar, O., Tuomi, J., 1998. Synergistic selection and graded traits. Evolutionary Ecology 12 (1), 59–71.

Malécot, G., 1975. Heterozygosity and relationship in regularly subdivided populations. Theoretical Population Biology 8 (2), 212–241.

Marshall, J. A. R., 2014. Generalisations of hamilton's rule applied to non-additive public goods games with random group size. Frontiers in Ecology and Evolution 2, 40.

Matessi, C., Jayakar, S. D., 1976. Conditions for the evolution of altruism under darwinian selection. Theoretical Population Biology 9 (3), 360–387.

Maynard Smith, J., 1964. Group selection and kin selection. Nature 201, 1145–1147.

Maynard Smith, J., 1965. The evolution of alarm calls. American Naturalist 99 (904), 59–63.

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

Metz, J., Geritz, S., Meszéna, G., Jacobs, F., van Heerwarden, J., 1996. Stochastic and spatial structures of dynamical systems. North Holland, Amsterdam, Netherlands, Ch. Adaptive dynamics: a geometrical study of the consequences of nearly faithful replication, pp. 183–231.

Motro, U., 1991. Co-operation and defection: playing the field and the ESS. Journal of Theoretical Biology 151 (2), 145–154.

Ohtsuki, H., 2010. Evolutionary games in Wright's island model: kin selection meets evolutionary game theory. Evolution 64 (12), 3344–3353.

Ohtsuki, H., 2012. Does synergy rescue the evolution of cooperation? An analysis for homogeneous populations with non-overlapping generations. Journal of Theoretical Biology 307 (0), 20–28.

Packer, C., Ruttan, L., 1988. The evolution of cooperative hunting. American Naturalist 132 (2), 159–198.

Peña, J., Lehmann, L., Nöldeke, G., 2014. Gains from switching and evolutionary stability in multi-player matrix games. Journal of Theoretical Biology 346 (0), 23–33.

Pepper, J. W., 2000. Relatedness in trait group models of social evolution. Journal of Theoretical Biology 206 (3), 355–368.

Queller, D. C., 1984. Kin selection and frequency dependence: a game theoretic approach. Biological Journal of the Linnean Society 23 (2-3), 133–143.

Queller, D. C., 1985. Kinship, reciprocity and synergism in the evolution of social behaviour. Nature 318 (6044), 366–367.

Queller, D. C., 1992. A general model for kin selection. Evolution 46 (2), 376–380.

Queller, D. C., 1994. Genetic relatedness in viscous populations. Evolutionary Ecology 8 (1), 70–73.

Queller, D. C., 2011. Expanded social fitness and Hamilton's rule for kin, kith, and kind. Proceedings of the National Academy of Sciences 108 (Supplement 2), 10792–10799.

Reuter, M., Keller, L., 2001. Sex ratio conflict and worker production in eusocial Hymenoptera. American Naturalist 158 (2), 166–177.

Rousset, F., 2004. Genetic Structure and Selection in Subdivided Populations. Princeton University Press, Princeton, NJ.

Rousset, F., Billiard, S., 2000. A theoretical basis for measures of kin selection in subdivided populations: finite populations and localized dispersal. Journal of Evolutionary Biology 13, 814–825.

Sachs, J. L., Mueller, U. G., Wilcox, T. P., Bull, J. J., 2004. The evolution of cooperation. The Quarterly Review of Biology 79 (2), 135–160.

Samuelson, P. A., 1954. The pure theory of public expenditure. The Review of Economics and Statistics 36 (4), 387–389.

Sawyer, S., Felsenstein, J., 1983. Isolation by distance in a hierarchically clustered population. Journal of Applied Probability 20 (1), 1–10.

Skyrms, B., 2004. The Stag Hunt and the Evolution of Social Structure. Cambridge University Press, Cambridge, UK.

Stander, P., 1992. Cooperative hunting in lions: the role of the individual. Behavioral Ecology and Sociobiology 29 (6), 445–454.

Sumpter, D. J. T., 2010. Collective Animal Behavior. Princeton University Press, Princeton, NJ.

Taylor, P., 1992. Altruism in viscous populations — an inclusive fitness model. Evolutionary Ecology 6 (4), 352–356.

Taylor, P., Maciejewski, W., 2012. An inclusive fitness analysis of synergistic interactions in structured populations. Proceedings of the Royal Society B: Biological Sciences 279, 45964603.

Taylor, P. D., 1989. Evolutionary stability in one-parameter models under weak selection. Theoretical Population Biology 36 (2), 125–143.

Taylor, P. D., Frank, S. A., 1996. How to make a kin selection model. Journal of Theoretical Biology 180 (1), 27–37.

Taylor, P. D., Irwin, A. J., 2000. Overlapping generations can promote altruistic behavior. Evolution 54 (4), 1135–1141.

Traulsen, A., Nowak, M. A., 2006. Evolution of cooperation by multilevel selection. Proceedings of the National Academy of Sciences 103 (29), 10952–10955.

Trivers, R. L., 1971. The evolution of reciprocal altruism. The Quarterly Review of Biology 46 (1), 35–57.

Van Cleve, J., 2014. Social evolution and genetic interactions in the short and long term. bioRxiv.

Van Cleve, J., Akçay, E., 2014. Pathways to social evolution: reciprocity, relatedness, and synergy. Evolution 68 (8), 2245–2258.

Van Cleve, J., Lehmann, L., 2013. Stochastic stability and the evolution of coordination in spatially structured populations. Theoretical Population Biology 89 (0), 75–87.

Velicer, G. J., 2003. Social strife in the microbial world. Trends in Microbiology 11 (7), 330–337.

Wenseleers, T., Gardner, A., Foster, K. R., 2010. Social evolution theory: a review of methods and approaches. In: Szekely, T., Moore, A., Komdeur, J. (Eds.), Social Behaviour: Genes, Ecology and Evolution. Cambridge University Press, Cambridge, pp. 132–158.

West, S. A., Diggle, S. P., Buckling, A., Gardner, A., Griffin, A. S., 2007. The social lives of microbes. Annu. Rev. Ecol. Evol. Syst. 38, 53–77.

West, S. A., Griffin, A. S., Gardner, A., Diggle, S. P., 2006. Social evolution theory for microorganisms. Nat Rev Micro 4 (8), 597–607.

Wheeler, D. E., 1986. Developmental and physiological determinants of caste in social hymenoptera: Evolutionary implications. American Naturalist 128 (1), 13–34.

White, C. E., Winans, S. C., 2007. Cell–cell communication in the plant pathogen agrobacterium tumefaciens. Philosophical Transactions of the Royal Society B: Biological Sciences 362 (1483), 1135–1148.

Wright, S., 1931. Evolution in Mendelian populations. Genetics 16, 97–159.

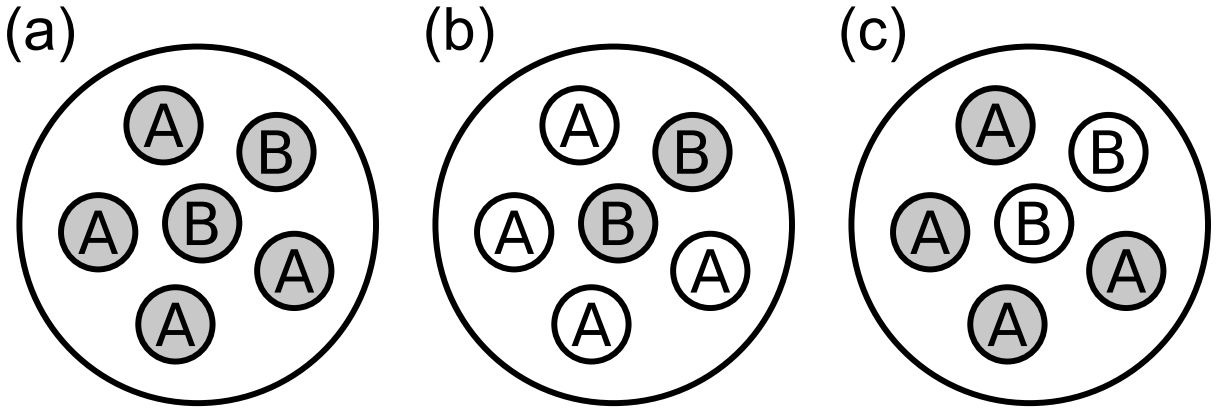

Figure 1: Three kinds of social traits. Expressers (As) provide a social good at a personal cost, nonexpressers (Bs) do not. The set of recipients (filled circles) of the social good depends on the particular kind of social interaction. *a*, Whole-group traits. *b*, Nonexpresser-only traits. *c*, Expresser-only traits.

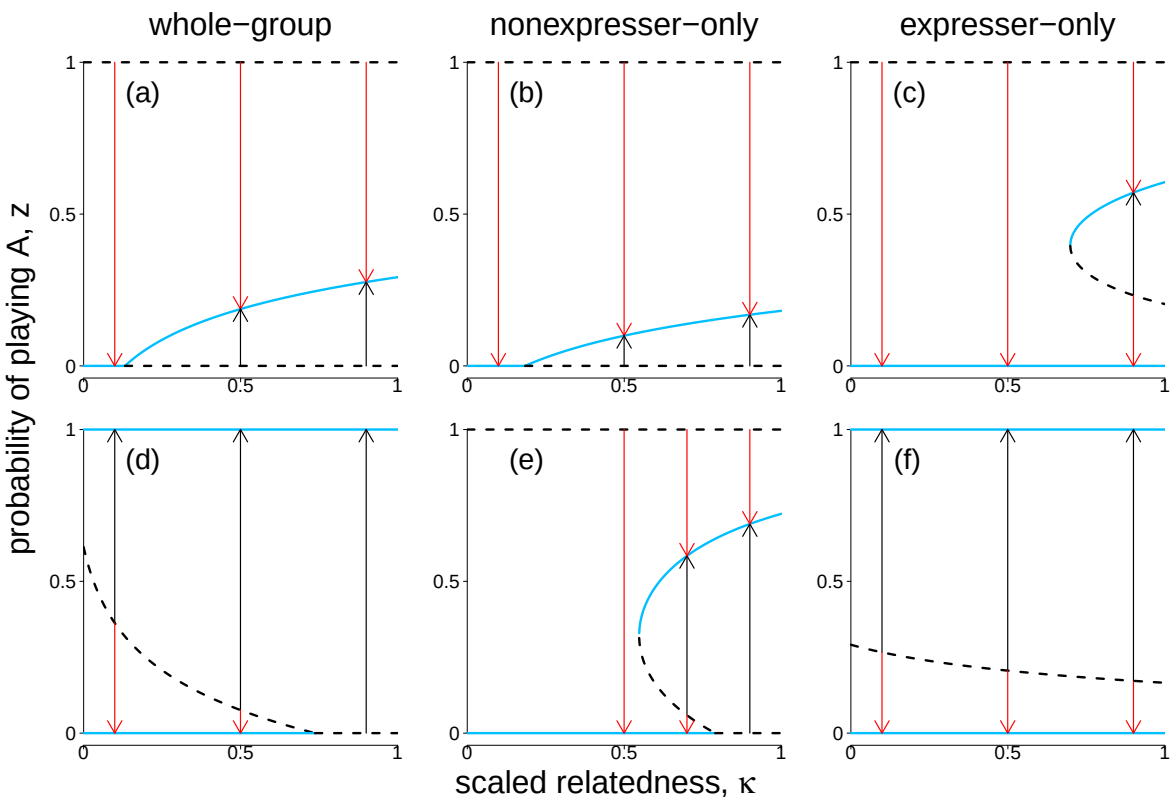

Figure 2: Bifurcation diagrams for whole-group (a, d), nonexpresser-only (b, e), and expresser-only (c, f) traits with geometric benefits. The scaled relatedness coefficient $\kappa \geq 0$ serves as a control parameter. Arrows indicate the direction of evolution for the probability of expressing helping. Solid lines stand for stable equilibria; dashed lines for unstable equilibria. a, b, c, Negative synergies of scale ($\lambda = 0.7$) and low cost-to-benefit ratio ($\gamma/\beta = 3.5$). d, e, f, Positive synergies of scale ($\lambda = 1.25$) and high cost-to-benefit ratio ($\gamma/\beta = 15$). In all plots, $n = 20$.

Symbol	Definition
A	first of two pure strategies (e.g., "help")
a_k	payoff to an A-player matched with k A-players and $n-1-k$ B-players
B	second of two pure strategies (e.g., "do not help")
B	payoff benefit parameter in games without synergies of scale
$\mathcal{B}(z)$	(fecundity) indirect effect
b	fitness benefit of carrying a social allele
b_k	payoff to a B-player matched with k A-players and $n-1-k$ B-players
C	payoff cost parameter in games without synergies of scale
c	fitness cost of carrying a social allele
$-\mathcal{C}(z)$	(fecundity) direct effect
d_k	direct gain from switching to a focal matched with k A-players and $n-1-k$ B-players
D	synergy parameter in games without synergies of scale
e_k	indirect gain from switching to a focal matched with k A-players and $n-1-k$ B-players
f_k	inclusive gain from switching to a focal matched with k A-players and $n-1-k$ B-players
$\mathcal{G}(z)$	gain function ($= -\mathcal{C}(z) + \kappa\mathcal{B}(z)$)
j	total number of A-players in a group
k	number of A co-players of a focal individual
N	number of adult individuals in a group
n	number of players (usually $= N$)
r	relatedness coefficient
β_j	benefit from the social good when j individuals express helping (play A)
β	parameter of the geometric benefits
z	resident strategy (phenotype)
$z_\bullet$	strategy (phenotype) of a focal individual
$z_{\ell(\bullet)}$	strategy (phenotype) of the ℓ-th co-player of the focal individual
$z_\circ$	average strategy (phenotype) of the neighbors of a focal individual
Δ	first forward difference operator ($\Delta c_k = c_{k+1} - c_k$)
γ	payoff cost of expressing helping
λ	parameter of the geometric benefits
κ	scaled relatedness coefficient
π	average payoff to a focal individual

Table 1: Symbols used in this article.

Social trait	a_k	b_k	d_k	e_k	f_k
whole-group	$-\gamma + \beta_{k+1}$	β_k	$-\gamma+\Delta\beta_k$	$(n-1)\Delta\beta_k$	$-\gamma+(1+\kappa(n-1))\Delta\beta_k$
nonexpresser-only	$-\gamma$	β_k	$-\gamma-\beta_k$	$(n-1-k)\Delta\beta_k$	$-\gamma-\beta_k+\kappa(n-1-k)\Delta\beta_k$
expresser-only	$-\gamma + \beta_{k+1}$	0	$-\gamma+\beta_{k+1}$	$k\Delta\beta_k$	$-\gamma+\beta_{k+1}+\kappa k\Delta\beta_k$

Table 2: Payoff structures and gains from switching for three kinds of social traits. In each case expressers (A-players) incur a cost $\gamma > 0$ and recipients obtain a benefit $\beta_j \geq 0$ which depends on the number of expressers j they experience. The number of expressers experienced by a focal is $j = k$ if the focal is a nonexpresser, otherwise it is $j = k + 1$. Direct gains (d_k) and indirect gains (e_k) are calculated by substituting the expressions for a_k and b_k into equations (5) and (6). Inclusive gains from switching (f_k) are then obtained from equation (9).

Social trait	f_k	$\mathcal{G}(z)$
whole-group	$-\gamma+\beta+\kappa(n-1)\beta$	$-\gamma+\beta+\kappa(n-1)\beta$
nonexpresser-only	$-\gamma+\kappa(n-1)\beta-(1+\kappa)\beta k$	$-\gamma+\kappa(n-1)\beta-(1+\kappa)\beta(n-1)z$
expresser-only	$-\gamma+\beta+(1+\kappa)\beta k$	$-\gamma+\beta+(1+\kappa)\beta(n-1)z$

Table 3: Inclusive gains from switching (f_k) and gain functions ($\mathcal{G}(z)$) for the case of linear benefits (no synergies of scale). The inclusive gains from switching are obtained by replacing $\beta_j = \beta j$ and $\Delta\beta_j = \beta$ into the corresponding expression from table 2. The gain function is then obtained from equation (10).

<table>
<tr><td rowspan="4">whole-group</td><td colspan="2">$\lambda < 1$</td><td colspan="2">$\lambda > 1$</td></tr>
<tr><td>$\gamma/\beta \leq \varepsilon$</td><td>$z = 1$</td><td>$\gamma/\beta \leq \varepsilon$</td><td>$z = 1$</td></tr>
<tr><td>$\varepsilon < \gamma/\beta < \vartheta$</td><td>$z^* \in (0, 1)$</td><td>$\varepsilon < \gamma/\beta < \vartheta$</td><td>$z = 0$, $z = 1$</td></tr>
<tr><td>$\gamma/\beta \geq \vartheta$</td><td>$z = 0$</td><td>$\gamma/\beta \geq \vartheta$</td><td>$z = 0$</td></tr>
<tr><td rowspan="4">nonexpresser-only</td><td colspan="2">$\lambda \leq \varrho$</td><td colspan="2">$\lambda > \varrho$</td></tr>
<tr><td>$\gamma/\beta < \zeta$</td><td>$z^* \in (0, 1)$</td><td>$\gamma/\beta < \zeta$</td><td>$z^* \in (\hat{z}, 1)$</td></tr>
<tr><td rowspan="2">$\gamma/\beta \geq \zeta$</td><td rowspan="2">$z = 0$</td><td>$\zeta \leq \gamma/\beta < \eta$</td><td>$z = 0$, $z_{\mathrm{R}} \in (\hat{z}, 1)$</td></tr>
<tr><td>$\gamma/\beta \geq \eta$</td><td>$z = 0$</td></tr>
<tr><td rowspan="5">expresser-only</td><td colspan="2">$\lambda < \xi$</td><td colspan="2">$\lambda \geq \xi$</td></tr>
<tr><td>$\gamma/\beta \leq 1$</td><td>$z = 1$</td><td>$\gamma/\beta \leq 1$</td><td>$z = 1$</td></tr>
<tr><td>$1 < \gamma/\beta < \varsigma$</td><td>$z = 0$, $z = 1$</td><td>$1 < \gamma/\beta < \varsigma$</td><td>$z = 0$, $z = 1$</td></tr>
<tr><td>$\varsigma \leq \gamma/\beta < \tau$</td><td>$z = 0$, $z_{\mathrm{R}} \in (\hat{z}, 1)$</td><td rowspan="2">$\gamma/\beta \geq \varsigma$</td><td rowspan="2">$z = 0$</td></tr>
<tr><td>$\gamma/\beta \geq \tau$</td><td>$z = 0$</td></tr>
</table>

Table 4: Convergence stable strategies for the three kinds of social traits with geometric benefits. The results hold for whole-group traits if $\kappa > -1/(n-1)$ and for nonexpresser-only and expresser-only traits if $\kappa \geq 0$. For whole-group traits, ε and ϑ are given by equation (17). For nonexpresser-only traits, ϱ, ζ, η, and $\hat{z}$ are functions of κ, n, and λ (see eq. (19), (20), and (21)). For expresser-only traits, ξ, ς, τ, and $\hat{z}$ are functions of κ, n, and λ (see eq. (22), (23) and (24)).