Evaluating the structure-coefficient theorem of evolutionary game theory

Alex McAvoy[a,b,c,1] and John Wakeley[a,1]

In order to accommodate the empirical fact that population structures are rarely simple, modern studies of evolutionary dynamics allow for complicated and highly heterogeneous spatial structures. As a result, one of the most difficult obstacles lies in making analytical deductions, either qualitative or quantitative, about the long-term outcomes of evolution. The "structure-coefficient" theorem is a well-known approach to this problem for mutation–selection processes under weak selection, but a general method of evaluating the terms it comprises is lacking. Here, we provide such a method for populations of fixed (but arbitrary) size and structure, using easily interpretable demographic measures. This method encompasses a large family of evolutionary update mechanisms and extends the theorem to allow for asymmetric contests to provide a better understanding of the mutation–selection balance under more realistic circumstances. We apply the method to study social goods produced and distributed among individuals in spatially heterogeneous populations, where asymmetric interactions emerge naturally and the outcome of selection varies dramatically, depending on the nature of the social good, the spatial topology, and the frequency with which mutations arise.

evolutionary dynamics | game theory | mutation | selection

Population structure is nearly ubiquitous in nature and occurs on many levels. Microbes exhibit strong spatial heterogeneity in response to variation in temperature, pH, and vegetation cover in terrestrial soils (1) and to harsh environmental gradients and patchy habitats at deep-sea hydrothermal vents (2). On a very different scale, global human social networks are known to consist of many highly connected clusters, often held together by dense cores (3), which can, in turn, affect both public opinion (4) and disease transmission (5, 6). Evolutionary outcomes depend on population structure because it constrains social interactions and the dispersal of offspring.

Fully a century ago, Hagedoorn and Hagedoorn (7) argued, based on observations of spatial heterogeneity in land snails, that variation due to chance in small local populations is an important factor of evolution. Early theoretical work by Wright (8, 9) and Malécot (10) on the distribution of gene frequencies and genetic identity in regularly subdivided populations stimulated later research in population and evolutionary genetics, focused largely on symmetric population structures with discrete, well-mixed subpopulations connected by migration. These include the island model (8), with symmetric migration between every pair of subpopulations, and the one-dimensional and two-dimensional stepping-stone models (11–13), with migration only between neighboring subpopulations on a lattice. These same models have been used to study social evolution; e.g., see Nakamaru et al. (14).

Decades of subsequent research clarified the empirical facts of population structure, as well as its importance in theoretical models of evolution (15, 16). Increased relatedness—and, thus, greater genetic identity within local populations—is the hallmark of population structure. The cause is local "genetic drift" (hereafter just drift) due to the vagaries of reproduction when the number of individuals in a given region is not large and migration is restricted. Frequencies of traits vary from one location to the next, which, in turn, affects the outcome of selection because it alters the rate of same-type interactions. Using diffusion models, for example, it has been shown that high local identity between individuals can change the effective population size and selection coefficient (17). These effects can dramatically alter the course of evolution [e.g., through increasing the fixation probability of recessive advantageous mutations (18)]. Ideas of nonuniform interaction have also been crucial in evolutionary game theory, even in models of otherwise well-mixed, infinite populations (19–24). That interaction structure and reproductive structure may work together nonindependently has been further emphasized by laboratory experiments (25, 26).

Population structure affects who interacts with whom, but it can also fundamentally change the nature of these interactions. In what follows, we use the example of competition

Significance

The balance between natural selection and mutation in populations with complex spatial structures is complicated to understand. Qualitatively, it is known that when selection acts sufficiently weakly, its effects on a trait's mean frequency can be understood by calculating, at most, three coefficients. Outside of special cases, however, computing these coefficients has remained an open problem. Here, we present a method for calculating these coefficients, which allows one to evaluate the effects of weak selection on populations with arbitrary mutation probabilities and spatial structure. This method includes a qualitative description of how simple demographic quantities, such as those related to birth and death rates, influence a trait's mean frequency.

between producers of a good (a social trait) and nonproducers. Producers generate a good of benefit b and pay a cost, c, to do so. While the production of a good in isolation is straightforward, its expression within a structured population can take different forms. One option is for a producer to generate a separate good for each neighbor, each of benefit b (to the neighbor) and cost c (to the producer). As a producer's group size grows, so, too, does the total good generated by the producer, a quality that has been observed in social grooming among primates (27). For other kinds of divisible goods, such as food shared among vampire bats (28) or the time and effort a researcher allocates across several collaborations (29), a more natural assumption might be that the good is produced once, and the benefit, b, is divided among neighbors (30). How much any one neighbor receives depends on how many share the benefit, which can vary significantly in a spatially heterogeneous population. Allotments of social goods in structured populations lead naturally to asymmetric games, in which payoffs to individuals depend on not just the trait of immediate interest (e.g., producer versus nonproducer), but on other factors as well, including group sizes (31, 32).

In this work, we consider the distribution of social goods in a general model of a graph-structured population, which can include the discrete-subpopulation models mentioned above, but, more importantly, can depict arbitrarily fine-grained heterogeneity (33–36). We are interested in whether a social trait is favored in evolution, with competition determined by a game that may be asymmetric. With the greater flexibility of an arbitrary graph comes greater complexity, which prohibits comprehensive analysis for all but the smallest populations or those with high degrees of symmetry. This trade-off between model complexity and tractability has prompted a variety of approaches, ranging from approximation techniques (37, 38) to exact results on homogeneous population structures (39–43) to the analysis of simulation algorithms (44). We follow a common fruitful way forward—namely, to assume that selection is weak and to focus on the stationary frequency of the social trait, in which case increased local relatedness and its effect on evolution can be captured with relative ease (45–50).

When selection is weak and mutations appear sufficiently infrequently, the problem can be reduced to a consideration of fixation probabilities (46, 47, 51, 52). Recent techniques allow one to compute fixation probabilities in populations with arbitrary spatial structure (53, 54). Under the assumption that a mutation either fixes or is removed from the population prior to the appearance of another mutant, fixation probabilities allow faithful approximations of the mean frequency of a trait (51). This assumption is slightly more restrictive than it first appears, as the amount of time it takes for the population to return to a monomorphic state can depend strongly on the nature of the interactions. For instance, it is known that for some kinds of frequency-dependent interactions, in which coexistence is possible for extended periods of time, the maximum allowable mutation probability decays exponentially in the population size (55, 56). For these kinds of traits, fixation probabilities are no longer necessarily the most meaningful or relevant measures of evolutionary dynamics.

Aside from the question of whether mutations are rare enough for fixation probabilities to be appropriate measures, we note that the mutation probability need not be small at all. In cultural contexts, mutations can reasonably occur with high frequency (57–62). Assuming weak selection, but not weak mutation, Tarnita et al. (58) elegantly boiled down the question of whether a trait is favored to a single quantity called the structure coefficient. To state their theorem, we use the notation of Débarre et al. (41)

and assume that there are two types in the population: S ("social") and NS ("nonsocial"). Payoffs for interactions between types are given by the matrix

$$
\begin{array}{cc}
 & \begin{array}{cc} \text{S} & \text{NS} \end{array} \\
\begin{array}{c} \text{S} \\ \text{NS} \end{array} & \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.
\end{array} \qquad [1]
$$

The social trait, S, is said to be favored by selection (via the game) if its equilibrium frequency is greater than what it would be under neutrality—i.e., if the first-order effect of selection on the mean frequency of the type is positive (46, 47). As this single matrix applies to all pairs of individuals, this is necessarily a symmetric game. Following Tarnita et al. (58), we assume that mutations occur with fixed probability $u > 0$ per reproduction event and that the resulting trait is sampled uniformly at random from the space of possible types. In particular, when there are two traits, S is favored by selection if its equilibrium frequency is above $1/2$. Under some mild conditions on the population structure and the evolutionary update rule, Tarnita et al. (58) show that there exists a "structure coefficient," σ, such that weak selection favors S over NS in the mutation–selection equilibrium whenever

$$
\sigma a_{11} + a_{12} > a_{21} + \sigma a_{22}. \qquad [2]
$$

The structure coefficient depends on the mutation probability, population structure, and update rule, but not on the payoffs of the game. Thus, under weak selection and considering only the mean frequency of S at equilibrium, σ captures all relevant aspects of population structure. It may be noted that σ appears in Eq. **2** as a weight on the payoffs of same-type interactions. If one models the effects of selection on fecundity and survival separately, using different payoff matrices for each, then the structure-coefficient theorem involves two weights, one for each step of the life cycle (41). These weights still retain the property of applying to same-type interactions, and when the two payoff matrices coincide (or one is zero), a condition involving only one structure coefficient is recovered. Our extension of the structure-coefficient theorem includes two coefficients, though we do not distinguish between different steps of a life cycle using different payoff matrices.

Tarnita et al. (58) proved the existence of σ, but did not give a general method for computing it. One of our goals here is to provide a recipe for calculating σ in terms of simple demographic quantities. However, knowing how to calculate σ still does not allow us to study the production of divisible social goods. To see why, we can return to the two kinds of social goods considered previously. Following the terminology of McAvoy et al. (63), we refer to goods in which each neighbor gets b at a cost of c as "pp-goods" since both the total benefit and the total cost are proportional ("p") to the number of interaction partners. The other kind of good, in which a single benefit, b, is divided among all neighbors, is termed an "ff-good" to indicate that the total benefits and costs are both fixed ("f"), independent of the number of neighbors. In both cases the interaction between any pair of neighboring individuals is a kind of "donation game" (64), but the combined outcome of all such pairwise games across the population differs for pp-goods versus ff-goods.

The case of pp-goods falls within the scope of the model of Tarnita et al. (58) since the benefits and costs are the same for every pair of neighbors in the population. In terms of Eq. **1**, pp-goods satisfy $a_{11} = b - c$, $a_{12} = -c$, $a_{21} = b$, and $a_{22} = 0$. Here, games are symmetric, and population structure only affects σ. In contrast, with ff-goods, population structure usually results in asymmetric games and cannot be summarized by a single

coefficient. The payoff to an individual of type x at location i in the game against an individual of type y at location j depends on their numbers of neighbors. We write a_{xy}^{ij} to denote the dependence on their locations and suppose that w_i and w_j are their numbers of neighbors. For brevity in much of what follows, we refer to the individuals at locations i and j simply as "i" and "j." If i is a producer, then j receives a benefit b/w_i, and i pays a cost c/w_i for having j as a neighbor. Similarly, if j is a producer, then j pays c/w_j to provide a benefit b/w_j to i. Thus, for ff-goods, $a_{11}^{ij} = (b/w_j - c/w_i)$, $a_{12}^{ij} = -c/w_i$, $a_{21}^{ij} = b/w_j$, and $a_{22}^{ij} = 0$. Games involving ff-goods in structured populations are symmetric only when every individual has the same number of neighbors, such that a_{xy}^{ij} is independent of i and j.

Here, we provide a method of analyzing mutation–selection dynamics in populations with finite size, N, and fixed (but arbitrary) structure. Specifically, we obtain an extension of the structure-coefficient theorem to games written in terms of a_{xy}^{ij}. We prove a result analogous to Eq. **2**, but that involves more than a single structure coefficient. This allows us to quantify how much a strategy is favored or disfavored, not just whether its mean frequency is above or below the value expected under neutral drift. We provide a general method of computing these structure coefficients in terms of simple demographic quantities and give a complexity bound on the calculations in terms of the population size. We also consider games with any finite number of strategies, providing a method for calculating the coefficients appearing in the generalization of the structure-coefficient theorem due to Tarnita et al. (59).

We use this method to investigate pp-goods versus ff-goods in graph-structured populations, in order to understand how the differences between these goods may be amplified by spatial heterogeneity, as well as how the rate of mutation affects the extent to which they are each favored (or disfavored) by selection. The results show strong effects of both population structure and mutation probability, in terms of which traits are favored and by how much. Trivially, as the mutation probability approaches one, selection becomes ineffective, independent of the type of good. But the approach to this can be nonmonotonic. The extent to which a trait is favored, and even whether it is favored or disfavored, depends on the mutation probability. Further, the relative advantage or disadvantage of pp-goods versus ff-goods depends on mutation. Finally, we use simulations to assess the sensitivity of these results beyond the limits of weak selection, and we find them to be qualitatively robust.

Theory and Results

We begin with an overview of the theory developed in detail in *SI Appendix*. We discuss the key parameters involved in assessing the effects selection on the mean frequencies of strategies or traits, and we describe how these quantities are used to compute structure coefficients. Table 1 provides a glossary of all the notation we use.

Following refs. 46 and 47, we consider the first-order effect of selection on the mean frequency of a trait when the population is at stationarity. We use the general model of population structure of Allen and McAvoy (65), in which a scalar parameter δ captures the intensity or strength of selection. By stationarity, we mean statistical equilibrium, with respect to the processes of mutation, selection via a game in which traits are strategies, and reproduction in a finite population. If $\langle x_S \rangle$ is the mean frequency of the social trait S in this stationary distribution, then S is said to be favored when $\langle x_S \rangle$ is greater than the corresponding prediction under neutrality. This is the framework of the structure-coefficient

Table 1. Glossary of notation

Symbol	Description
b	Benefit of a social good
c	Cost of a social good
S	The mutant (social) trait
NS	The resident (nonsocial) trait
a_{xy}	Payoff to x against y in a symmetric game
u	Probability of a mutation on reproduction
σ	Structure coefficient of Tarnita et al. (58)
w_i	Number of neighbors (degree) of i
a_{xy}^{ij}	Payoff to x at i against y at j in an asymmetric game
N	Population size
δ	Selection intensity
$\langle x_S \rangle$	Equilibrium mean frequency of S
$K_1^{k\ell}, K_2^{k\ell}$	Structure coefficients for asymmetric games
R	Set of individuals to be replaced
α	Offspring-to-parent map
n	Number of traits
$\mathbf{x}$ (x_i)	State of the population of (i)
$p_{(R,\alpha)}(\mathbf{x})$	Probability of replacement event (R, α) in state $\mathbf{x}$
$e_{ij}(\mathbf{x})$	Probability that i replaces j in state $\mathbf{x}$
$d_i(\mathbf{x})$	Death probability of i in state $\mathbf{x}$
A_{ij}	Probability of moving from i to j in the ancestral walk
z_i	Equilibrium probability that an ancestral line is at i
π_i	Reproductive value of i
z_i^{mut}	Mean probability that first ancestral mutation is at i
π_i^{mut}	Mutation-weighted reproductive value of i
ϕ_{ij}	Probability that i and j are identical by state
$U_i(\mathbf{x})$	Payoff to i in state $\mathbf{x}$
$\mathbf{F}(\mathbf{x})$ $(F_i(\mathbf{x}))$	Fecundity of the population (i) in state $\mathbf{x}$
m_k^{ij}	Marginal effect of k's fecundity on i replacing j
$p_{(x_i, x_j)}^{ij}$	Neutral probability that i is x_i and j is x_j
$p_{(x_i, x_j, x_k)}^{ijk}$	Neutral probability that i is x_i, j is x_j, and k is x_k
Ω_{ij}	Interaction structure of the population
w_{ij}	Adjacency matrix of the population
$\circ$	Indicates neutral drift ($\delta = 0$)

theorem, which follows from considering $\frac{d}{d\delta}\langle x_S \rangle$ evaluated at $\delta = 0$.

We take the approach in McAvoy et al. (63) and consider the effects of a deviation from neutrality on the fecundities of individuals ($\mathbf{F}$), as well as the effects of the resulting deviations in fecundities on the frequency of the trait. Fecundities depend on the game payoffs and the structure of the population. Thus, we compute the first-order effects of selection using the relationship

$$\frac{d\langle x_S \rangle}{d\delta} = \frac{d\langle x_S \rangle}{d\mathbf{F}}\frac{d\mathbf{F}}{d\delta}. \qquad [3]$$

Our solution involves three types of measures of population heterogeneity, which, in turn, depend on fundamental demographic parameters. The first measure, π^{mut}, is a type of reproductive value that summarizes patterns of identity by descent across the population. The second, ϕ, is a measure of identity by state, for all pairs of individuals in the population, which determines expected payoffs to individuals. The third, m_k^{ji}, is the effect of a deviation from neutral fecundity in individual k on the probability that j replaces i in one time step. While the first two measures are evaluated under neutrality ($\delta = 0$), the third measure links these quantities to the effects of selection. Note again that, for brevity, by "j replaces i," we mean the event that the individual at location j replaces the individual at location i and, similarly, that i, j, and k are technically locations within the population, not individuals.

Our main result, which is derived in *SI Appendix*, says that

$$\langle x_S \rangle = \frac{1}{2} + \delta \frac{1}{2} \sum_{k,\ell=1}^{N} \left(\begin{array}{c} K_1^{k\ell}\left(a_{11}^{k\ell} - a_{22}^{k\ell}\right) \\ + K_2^{k\ell}\left(a_{12}^{k\ell} - a_{21}^{k\ell}\right) \end{array} \right) + O\left(\delta^2\right), \qquad \textbf{[4]}$$

where

$$K_1^{k\ell} = \frac{1}{2u} \sum_{i,j=1}^{N} \pi_i^{\mathrm{mut}} m_k^{ji} \left(\begin{array}{c} -(\phi_{ik} + \phi_{i\ell}) \\ +(1-u)(\phi_{jk} + \phi_{j\ell}) + u \end{array} \right); \qquad \textbf{[5a]}$$

$$K_2^{k\ell} = \frac{1}{2u} \sum_{i,j=1}^{N} \pi_i^{\mathrm{mut}} m_k^{ji} \left(\begin{array}{c} -(\phi_{ik} - \phi_{i\ell}) \\ +(1-u)(\phi_{jk} - \phi_{j\ell}) \end{array} \right). \qquad \textbf{[5b]}$$

Again, we assume that $u > 0$, but we note that Eq. **5** does not diverge as $u \to 0$. Both $-(\phi_{ik} + \phi_{i\ell}) + (1-u)(\phi_{jk} + \phi_{j\ell}) + u$ and $-(\phi_{ik} - \phi_{i\ell}) + (1-u)(\phi_{jk} - \phi_{j\ell})$ in Eq. **5** vanish in the limit $u \to 0$ because all individuals must eventually be identical by state in the absence of mutation. These terms divided by u then each have finite limits as $u \to 0$. We refer the reader to refs. 54 and 65 for discussions of the rare-mutation limit in this class of models.

Modeling Population Dynamics. We consider a discrete-time population model with arbitrary spatial structure. The population evolves through a sequence of replacement events, which consist of pairs, (R, α), where $R \subseteq \{1, \ldots, N\}$ is the set of individuals to be replaced and $\alpha : R \to \{1, \ldots, N\}$ is the parentage map (65, 66). In such an event, the individual at location $i \in R$ is replaced by the offspring of the individual at location $\alpha(i)$. If $i \notin R$, then i simply lives through the time step. We assume that there are no "empty" sites.

Each individual has one of n traits. We use **x** to denote the current state of the population, which is a list of the trait values of all N individuals. For example, with just $n = 2$ possible trait values, S and NS, we would have $x_i \in \{S, NS\}$ for $i \in \{1, \ldots, N\}$. Here, we consider $n = 2$, but our results in *SI Appendix* allow for $n > 2$. We use $p_{(R,\alpha)}(\mathbf{x})$ to denote the probability that replacement event (R, α) occurs. This "replacement rule" (65, 66) depends on the state of the population because individual fecundities are affected by game payoffs.

We allow symmetric, parent-independent mutation with probability u per individual offspring in a replacement event; that is, with probability $1 - u$, the offspring of $\alpha(i)$ that replaces i has the same strategy as its parent. With probability u, the offspring mutates and acquires a strategy uniformly-at-random— for example, from $\{S, NS\}$ when $n = 2$. Thus, offspring and parent might still be identical by state after a mutation occurs, but they will not be identical by descent. We assume that the mutation probability is positive and is the same for all individuals.

If $u > 0$ and a mild assumption on the replacement rule $p_{(R,\alpha)}(\mathbf{x})$ holds–namely, that at least one individual can propagate its offspring throughout the entire population (*SI Appendix*, section 1.1)–then this process has a unique stationary distribution with $0 < \langle x_S \rangle < 1$. We study this distribution under the assumption of weak selection inherent to the structure-coefficient theorem, with selection intensity $\delta \geqslant 0$ determining the importance of the game to fecundity. Note that $p_{(R,\alpha)}(\mathbf{x})$ does not depend on **x** when $\delta = 0$ because trait values do not affect fecundity under neutrality. We use $p_{(R,\alpha)}^{\circ}$ to denote the corresponding neutral replacement rule.

The fundamental demographic parameters of individuals are marginal properties of $p_{(R,\alpha)}(\mathbf{x})$. The first and most important is the probability that i replaces j in one time step,

$$e_{ij}(\mathbf{x}) := \sum_{\substack{(R,\alpha) \\ j \in R, \, \alpha(j)=i}} p_{(R,\alpha)}(\mathbf{x}). \qquad \textbf{[6]}$$

Note that this replacement involves the death of the previous individual, j. Other demographic quantities are functions of the $e_{ij}(\mathbf{x})$. The second demographic parameter of which we make extensive use here is the probability that i dies in one time step, $d_i(\mathbf{x}) := \sum_{j=1}^{N} e_{ji}(\mathbf{x})$. Weak selection causes deviations from the neutral values of these two quantities. We call these neutral values e_{ij}° and d_i° and note that they are independent of the population state, **x**. Deviations in e_{ij} (and all other demographic quantities) due to selection via the game do depend on **x**, but only through changes in the fecundities (**F**), as in Eq. **3**. We use $\mathbf{F}^{\circ}$ to denote fecundities under neutral drift.

In the following subsections, we discuss the three types of measures of population heterogeneity introduced briefly in Eq. **5**, which derive from the analysis in *SI Appendix*.

Reproductive value and identity by descent.. The coefficient π_i^{mut} in Eq. **5** is a mutation-weighted reproductive value, which summarizes patterns of identity by descent between individuals and their direct ancestors at location i. To illustrate what π_i^{mut} represents here, we first recall the classical notion of reproductive value.

Under neutral drift, probabilities of birth, death, and replacement still vary among individuals due to their locations in the population structure. Reproductive values measure the contributions of individuals to distant future generations or, equivalently, the contributions of long-ago ancestral individuals to the present generation (67–73). If π_i is the reproductive value of i, then with probability $1 - d_i^{\circ}$, individual i survives and brings its value π_i forth into the next time step; with probability e_{ij}°, individual i replaces j and, in doing so, acquires the value π_j. At stationarity, the reproductive values satisfy the balance equation $\pi_i = (1 - d_i^{\circ})\pi_i + \sum_{j=1}^{N} e_{ij}^{\circ}\pi_j$.

The solution to this recurrence is unique up to the value of $\sum_{i=1}^{N} \pi_i$. A standard normalization is $\sum_{i=1}^{N} \pi_i = 1$. However, this is not the only relevant normalization, and here it will be useful to work with something different. Reproductive values are closely related to the stationary distribution of a random walk on the population. Let $A_{ij} := e_{ji}^{\circ}/d_i^{\circ}$ be the probability that the parent of i was located at j, conditioned on the death of i. Thus, j replacing i forward in time equates to i moving to j backward in time. This ancestral random walk tracks only parentage and ignores lifespans (65). It has a unique stationary distribution, z, which satisfies $z_i = \sum_{j=1}^{N} z_j A_{ji}$ and represents the equilibrium probability of an ancestral line being at i after many lifetimes. Weighting this probability by the expected lifespan at i gives the reproductive value, $\pi_i = z_i/d_i^{\circ}$, of i.

The distinction between π_i and z_i aids in the interpretation of π_i^{mut} and its relationship to patterns of identity by descent. Consider the ancestral random walk defined by the matrix A. Starting with an individual at location j, each step backward in time is associated to a mutation with probability u and to a faithful transmission of genetic material with probability $1 - u$. We consider the endpoint, i, to be the location of the most ancient individual that is identical by descent with j. Thus, the first mutation in the ancestral line from j occurs between i and its parent, $\alpha(i)$. Since a mutation must eventually arise in this walk, the endpoint defines a distribution over $\{1, \ldots, N\}$.

For the first mutation to occur at i after t steps in the past, the first t reproduction events in this lineage must be mutation-free, which happens with probability $(1-u)^t$. The probability

that the lineage is at i is then $\left(A^t\right)_{ji}$. Finally, the parent of i must produce a mutated offspring, which happens with probability u. Summing these probabilities over all t gives a total probability of

$$\sum_{t=0}^{\infty} u\left(1-u\right)^t \left(A^t\right)_{ji} = u\left(\left(I-\left(1-u\right)A\right)^{-1}\right)_{ji}, \quad \text{[7]}$$

that the first mutation happens at i, given a starting position of j. Because $\langle x_S \rangle$ is the mean frequency of the trait taken uniformly over the population, our calculations depend on the average

$$z_i^{\text{mut}} := \frac{1}{N}\sum_{j=1}^{N} u\left(\left(I-\left(1-u\right)A\right)^{-1}\right)_{ji}. \quad \text{[8]}$$

z_i^{mut} is the analog of z_i, but weighted by identity by descent (i.e., with a time scale dependent on the mutation probability). In the limit $u \to 0$, z_i^{mut} converges to z_i. At the other extreme, as $u \to 1$, z_i^{mut} becomes uniform on $\{1, \ldots, N\}$, owing to the fact that the mutation will then occur when $t=0$, so Eq. 7 will tend to one for $i=j$ and zero for $i \neq j$. As a distribution, $\sum_{i=1}^{N} z_i^{\text{mut}} = 1$ for any u.

The mutation-weighted reproductive value π_i^{mut} in Eq. 5 is defined by $\pi_i^{\text{mut}} := z_i^{\text{mut}}/d_i^{\circ}$. Analogous to the expression $\pi_i = z_i/d_i^{\circ}$, the mutation-weighted reproductive value captures both the contribution of i to the total population, through the factor z_i^{mut}, and the time scale of reproduction, through the factor $1/d_i^{\circ}$. In the limits $u \to 0$ and $u \to 1$, π_i^{mut} converges to π_i and $1/\left(Nd_i^{\circ}\right)$, respectively. Further details about how π^{mut} arises may be found in *SI Appendix*, section SI.1.2.

Identity by state.. When selection is weak, neutral probabilities of identity by state play a key role in the outcome of frequency-dependent selection (46, 74). Let ϕ_{ij} denote the stationary probability that two individuals at locations i and j are identical by state (have the same type or strategy) under neutral drift. Of course, $\phi_{ii} = 1$ for $i = 1, \ldots, N$. The other ϕ_{ij} for $i \neq j$ can be computed by using standard techniques, recursively over a single time step. We present the details in *SI Appendix*, section SI.2, where we account for the possible replacement of both i and j and for the possible occurrence of zero, one, or two mutations in *SI Appendix*, Eq. **SI.23**. Solving for these $N\left(N-1\right)/2$ probabilities, ϕ_{ij} for $i \neq j$, is the main computational burden of our method. In our application to producers of social goods below, we discuss a simple intuitive example of this approach to identities by state.

Marginal effects of selection on replacement.. Up to this point, our focus has been on neutrality ($\delta = 0$). Since we are interested in the perturbative effects of selection on the neutral process, we also need to understand how changes in individual fecundity affect the transmission of traits. In evolutionary games, the influence of the state of the population on replacement can be modeled by using a fecundity function, which assigns to each state a positive quantity representing an individual's competitive abilities (including, but not limited to, the propensity to reproduce). If $U_i\left(\mathbf{x}\right)$ is the payoff of individual i in state $\mathbf{x}$, then a typical conversion of payoff to fecundity is of the form $F_i\left(\mathbf{x}\right) = e^{\delta U_i\left(\mathbf{x}\right)}$ or $F_i\left(\mathbf{x}\right) = 1 + \delta U_i\left(\mathbf{x}\right)$, where $\delta \geqslant 0$ represents the selection intensity. These two forms of F_i are equivalent under weak selection, so we do not need to explicitly choose one (75). In either case, weak selection is modeled as a small deviation, proportional to δ, from a neutral baseline fecundity F_i° for all $i = 1, \ldots, N$. We note that F_i represents a general measure of the competitive abilities of i, which can pertain to reproduction, survival, or a combination of the two;

we use the term "fecundity" instead of "fitness" to avoid confusion because the latter term is defined in many ways in the literature.

Our analysis of weak selection uses the fact that the replacement rule $p_{\left(R,\alpha\right)}\left(\mathbf{x}\right)$ depends on $\mathbf{x}$ because $\mathbf{x}$ determines game payoffs, which then determine fecundities. The demographic parameters of individuals, which are marginal properties of $p_{\left(R,\alpha\right)}\left(\mathbf{x}\right)$, also depend implicitly on fecundities. The population's fecundities are given, in general, by $\mathbf{F} \in \left(0,\infty\right)^N$, with $\mathbf{F} = \mathbf{F}^{\circ}$ when $\delta = 0$. We employ $\mathbf{F}$ as a set of intermediate variables between δ and the main object of our analysis, the mean trait frequency $\langle x_S \rangle$. In our model, selection is directly mediated by one demographic parameter in particular, the replacement probability $e_{ij}\left(\mathbf{x}\right)$. As in ref. 63, we define $m_k^{ij} := \left.\frac{\partial e_{ij}}{\partial F_k}\right|_{\mathbf{F}=\mathbf{F}^{\circ}}$ as a measure of how individual payoffs affect competition in an evolutionary game. These terms are independent of the game itself, and they allow us to express the marginal effects of selection on the probability that i replaces j conveniently in terms of the payoffs of the game,

$$e_{ij}\left(\mathbf{x}\right) = e_{ij}^{\circ} + \delta \sum_{k=1}^{N} m_k^{ij} U_k\left(\mathbf{x}\right) + O\left(\delta^2\right). \quad \text{[9]}$$

Note that we implicitly assume that $p_{\left(R,\alpha\right)}\left(\mathbf{x}\right)$ is a smooth function of δ (*SI Appendix*, section SI.1.1).

Mean Trait Frequencies. To see how our main result may be written succinctly in terms of probabilities of identity in state (ϕ_{ij}), as well as the quantities π_i^{mut} and m_k^{ji}, it is helpful to begin with the trait values of individuals. Fundamentally, the first-order effect of selection on the mean frequency of the social trait S depends on expected payoffs, which, in turn, depend on traits. In *SI Appendix*, we show that

$$\langle x_S \rangle = \frac{1}{2} + \delta \frac{1}{u} \sum_{i,j,k=1}^{N} \pi_i^{\text{mut}} m_k^{ji} \begin{pmatrix} \left(1-u\right)\frac{1}{2}\langle U_k|j\sim S\rangle^{\circ} \\ +u\frac{1}{2}\langle U_k\rangle^{\circ} \\ -\frac{1}{2}\langle U_k|i\sim S\rangle^{\circ} \end{pmatrix} + O\left(\delta^2\right), \quad \text{[10]}$$

where $\langle \cdot \rangle^{\circ}$ is the expectation taken over the stationary distribution of trait values across the population under neutrality. Thus, selective events involving i are scaled by the mutation-weighted reproductive value, π_i^{mut}. Then, for each j whose offspring can replace i, m_k^{ji} quantifies how the performance of each individual k influences the probability that j replaces i. Finally, accounting for the possibility of mutation, the influx of S to i (via the reproduction of j) and the outflux of S from i (via the death of i) depend on three kinds of expected payoffs to individual k: $\langle U_k \mid j \sim S \rangle^{\circ}$ and $\langle U_k \mid i \sim S \rangle^{\circ}$ are the expected payoffs to k conditioned on j and i having type S, respectively, while $\langle U_k \rangle^{\circ}$ is the unconditional expected payoff to k.

Two kinds of marginal probabilities are required to describe the expected payoffs in Eq. **10**. Let $p_{\left(x_i,x_j\right)}^{ij}$ be the probability that i and j have types $x_i, x_j \in \{S, NS\}$. Similarly, define $p_{\left(x_i,x_j,x_k\right)}^{ijk}$ for the triplet i, j, k. Using this notation, the expected payoff to k when j has type S satisfies

$$\frac{1}{2}\langle U_k \mid j \sim S \rangle^{\circ} = \sum_{\ell=1}^{N} \begin{pmatrix} p_{\left(S,S,S\right)}^{jk\ell} a_{11}^{k\ell} + p_{\left(S,S,NS\right)}^{jk\ell} a_{12}^{k\ell} \\ +p_{\left(S,NS,S\right)}^{jk\ell} a_{21}^{k\ell} + p_{\left(S,NS,NS\right)}^{jk\ell} a_{22}^{k\ell} \end{pmatrix}. \quad \text{[11]}$$

The prefactor of $1/2$ is the probability that j has type S under neutral drift. The left-hand side of Eq. **11** is multiplied by $1-u$ in Eq. **10** because if there is no mutation, then the offspring of j that replaces i has type S. Since we care only about the influx of S

to i from j, we do not consider expected payoffs conditioned on $x_j = \text{NS}$. Similarly, regardless of the type of j, the offspring of j will mutate and end up type S with probability $u/2$. In this case, the expected payoff to k is

$$\langle U_k \rangle^\circ = \sum_{\ell=1}^{N} \left(\begin{array}{c} p_{(\text{S,S})}^{k\ell} a_{11}^{k\ell} + p_{(\text{S,NS})}^{k\ell} a_{12}^{k\ell} \\ + p_{(\text{NS,S})}^{k\ell} a_{21}^{k\ell} + p_{(\text{NS,NS})}^{k\ell} a_{22}^{k\ell} \end{array} \right), \qquad \textbf{[12]}$$

which does not depend on the trait of j. Both Eqs. **11** and **12** correspond to the influx of S to i, so they have positive weights in Eq. **10**. In contrast, the expected payoff to k when $x_i = \text{S}$ satisfies

$$\frac{1}{2} \langle U_k \mid i \sim \text{S} \rangle^\circ = \sum_{\ell=1}^{N} \left(\begin{array}{c} p_{(\text{S,S,S})}^{ik\ell} a_{11}^{k\ell} + p_{(\text{S,S,NS})}^{ik\ell} a_{12}^{k\ell} \\ + p_{(\text{S,NS,S})}^{ik\ell} a_{21}^{k\ell} + p_{(\text{S,NS,NS})}^{ik\ell} a_{22}^{k\ell} \end{array} \right), \qquad \textbf{[13]}$$

which is subtracted because the replacement of i by j results in the loss of this S at i. The prefactor of $1/2$ is the probability that i has type S under neutral drift. The left-hand side of Eq. **13** corresponds to the outflux of S from i through death. As such, we do not need to consider terms of the form $\langle U_k \mid i \sim \text{NS} \rangle^\circ$ because these correspond to zero outflux of S at i through death.

To present our main result in the form suggested by Eqs. **4** and **5**, we express $p_{(x_i, x_j)}^{ij}$ and $p_{(x_i, x_j, x_k)}^{ijk}$ in terms of probabilities of identity by state. Simplifications are available when there are just two possible trait values. We have $p_{(\text{S,S})}^{ij} = p_{(\text{NS,NS})}^{ij} = \phi_{ij}/2$ and $p_{(\text{S,NS})}^{ij} = p_{(\text{NS,S})}^{ij} = (1 - \phi_{ij})/2$. When there are two possible trait values (and only in this case; *SI Appendix*), $p_{(x_i, x_j, x_k)}^{ijk}$ can be calculated by using just pairwise identity probabilities, as outlined in Table 2. These expressions, together with Eqs. **11–13**, reduce Eq. **10** to our main result, given by Eqs. **4** and **5**. An overview of the steps involved in evaluating this result in practice is given in Box 1.

Eq. **4** has a form similar to the structure-coefficient theorem of Tarnita et al. (58). In fact, our result includes this structure-coefficient theorem as a special case. Note that a_{xy}^{ij} implicitly encodes which individuals interact and how payoffs are aggregated through the dependence on i and j. For example, we would define the payoffs to be zero for noninteracting pairs of individuals. In the case of a symmetric game described by Eq. **1**, an interaction structure for the population needs to be specified. Let $(\Omega_{ij})_{i,j=1}^{N}$ be a matrix representing this interaction structure. Then, if the individuals at locations i and j have types S and NS,

Table 2. Calculating the probability that *i*, *j*, and *k* have types x_i, x_j, and x_k, respectively, under neutral drift when there are just two possible trait values, S and NS

Triplet, (x_i, x_j, x_k)	Probability, $p_{(x_i, x_j, x_k)}^{ijk}$
(S, S, S)	$\frac{1}{4}(\phi_{ij} + \phi_{ik} + \phi_{jk} - 1)$
(S, S, NS)	$\frac{1}{4}(\phi_{ij} - \phi_{ik} - \phi_{jk} + 1)$
(S, NS, S)	$\frac{1}{4}(-\phi_{ij} + \phi_{ik} - \phi_{jk} + 1)$
(S, NS, NS)	$\frac{1}{4}(-\phi_{ij} - \phi_{ik} + \phi_{jk} + 1)$
(NS, S, S)	$\frac{1}{4}(-\phi_{ij} - \phi_{ik} + \phi_{jk} + 1)$
(NS, S, NS)	$\frac{1}{4}(-\phi_{ij} + \phi_{ik} - \phi_{jk} + 1)$
(NS, NS, S)	$\frac{1}{4}(\phi_{ij} - \phi_{ik} - \phi_{jk} + 1)$
(NS, NS, NS)	$\frac{1}{4}(\phi_{ij} + \phi_{ik} + \phi_{jk} - 1)$

For each triplet of traits, (x_i, x_j, x_k), the probability $p_{(x_i x_j x_k)}^{ijk}$ is a function of the identity-by-state probabilities ϕ_{ij}, ϕ_{ik}, and ϕ_{jk}

respectively, i gets payoff $\Omega_{ij} a_{12}$ and j gets payoff $\Omega_{ji} a_{21}$ from this interaction. If i and j do not interact, then $\Omega_{ij} = 0$. For every individual, these scaled payoffs are summed over all interaction partners.

A graph-structured population can be described by using an adjacency matrix $(w_{ij})_{i,j=1}^{N}$, with $w_{ij} = 1$ if i and j are neighbors and $w_{ij} = 0$ otherwise. In our model, this would then constrain the replacement rule, with (for example) $e_{ij}(\mathbf{x}) = 0$ whenever $w_{ij} = 0$. A common approach is to set the interaction matrix $\Omega_{ij} = w_{ij}$ to indicate that individuals interact only with neighbors and accumulate the resulting payoffs, e.g., additively. Another approach is to let $\Omega_{ij} = w_{ij}/w_i$, where w_i is the total number of neighbors of i, yielding average payoffs. Both kinds of payoffs are commonly used in evolutionary game theory (75). Whatever the interaction structure is for a particular population, the structure-coefficient theorem for symmetric games may be obtained from the result above by setting $a_{11}^{ij} = \Omega_{ij} a_{11}$, $a_{12}^{ij} = \Omega_{ij} a_{12}$, $a_{21}^{ij} = \Omega_{ij} a_{21}$, and $a_{22}^{ij} = \Omega_{ij} a_{22}$ in Eq. **4**.

Our results can be extended to games with $n > 2$ strategies. For general matrix games with two strategies, as considered above, or for additive matrix games with $n \geqslant 3$ strategies, the complexity of evaluating the structure-coefficient theorem is bounded by solving a linear system of size $O(N^2)$. For nonadditive matrix games with $n \geqslant 3$ strategies, this linear system is of size $O(N^3)$. We give a complete description of these extensions to n strategies in *SI Appendix*.

Applications to Producers of Social Goods. We conclude with an application of our theoretical results to the evolution of producers of social goods. We are interested in how the evolutionary advantages/disadvantages of making ff-goods versus pp-goods are affected by mutation—for example, whether the noise introduced by mutation always favors one kind of good over another. We investigate how these advantages/disadvantages change as mutation

goes from weak to strong, and we use simulations to verify the accuracy of our numerical predictions for the mean frequency of the social trait at equilibrium.

We assume that, at each point in time, neighbors on a graph interact and receive a total payoff according to the nature of the goods generated by producers. Let $(w_{ij})_{i,j=1}^N$ be the adjacency matrix of the population structure, which here is identical to the interaction structure for the game. A producer at location i pays C_{ij} in order to provide j with a benefit of B_{ij}. For pp-goods, $B_{ij} = bw_{ij}$ and $C_{ij} = cw_{ij}$; for ff-goods, $B_{ij} = bw_{ij}/w_i$ and $C_{ij} = cw_{ij}/w_i$, where w_i is the number of neighbors of i. In other words, we have $a_{11}^{ij} = B_{ji} - C_{ij}$, $a_{12}^{ij} = -C_{ij}$, $a_{21}^{ij} = B_{ji}$, and $a_{22}^{ij} = 0$.

An individual's total payoff, $U_i(\mathbf{x})$, obtained by summing over all interactions, is converted to fecundity, $F_i(\mathbf{x}) = e^{\delta U_i(\mathbf{x})}$, where $\delta \geqslant 0$ is the intensity of selection and $\mathbf{x}$ is the current state of the population. The population then updates according to the death–Birth (dB) rule, in which an individual is first selected uniformly at random for death, and then the neighbors compete, based on their relative fecundities, to fill the vacancy. The neighbors in this reproduction step are also determined by $(w_{ij})_{i,j=1}^N$. In this case, the probability that i replaces j when the state is $\mathbf{x}$ is

$$ e_{ij}(\mathbf{x}) = \frac{1}{N} \frac{F_i(\mathbf{x}) w_{ij}}{\sum_{k=1}^N F_k(\mathbf{x}) w_{kj}}. \qquad [14] $$

Under neutral drift, the probability that i replaces j is $e_{ij}^\circ = (w_{ji}/w_j)/N$, and the probability that i dies in one step of the process is $d_i^\circ = 1/N$. The probability of moving from i to j in one step of the ancestral random walk is then $A_{ij} = w_{ij}/w_i$. The stationary distribution for this random walk satisfies $z_i = w_i/\sum_{j=1}^N w_j$, which gives i a reproductive value of $\pi_i = Nw_i/\sum_{j=1}^N w_j$. For $u > 0$, the relevant notion of reproductive value, π_i^{mut}, has a more complicated expression, but can easily be found by calculating z_i^{mut} from $(A_{ij})_{i,j=1}^N$ (using Eq. **8**) and letting $\pi_i^{\text{mut}} = z_i^{\text{mut}}/d_i^\circ$.

The pairwise identity-by-state probabilities, ϕ, satisfy $\phi_{ii} = 1$ for every $i = 1, \ldots, N$. For $i \neq j$, since exactly one individual is replaced in each time step under dB updating, *SI Appendix*, Eq. **SI.23** gives

$$ \phi_{ij} = u\frac{1}{2} + (1-u)\frac{1}{2}\sum_{k=1}^N A_{ik}\phi_{kj} + (1-u)\frac{1}{2}\sum_{k=1}^N A_{jk}\phi_{ik}. \qquad [15] $$

In other words, to determine the probability that i is identical by state to $j \neq i$, we wait until either i or j is replaced by the offspring of some parent k. The first term on the right-hand side of Eq. **15** represents a mutation from the parent, which happens with probability u and leads to i and j being identical by state with probability $1/2$. Otherwise, with probability $1 - u$, there is no mutation, i is the one replaced with probability $1/2$, k is the parent with probability A_{ik}, and k is identical by state to j with probability ϕ_{kj}. The same logic applies to the last term, when j is the one replaced.

Finally, the marginal effects of selection on replacement satisfy $m_i^{ij} = A_{ji}(1 - A_{ji})/N$ and $m_k^{ij} = -A_{ji}A_{jk}/N$ if $k \neq i$. These marginal effects can be calculated directly from Eq. **14**.

Putting all of these ingredients together, we see that Eq. **10** (or Eq. **4**) takes the form
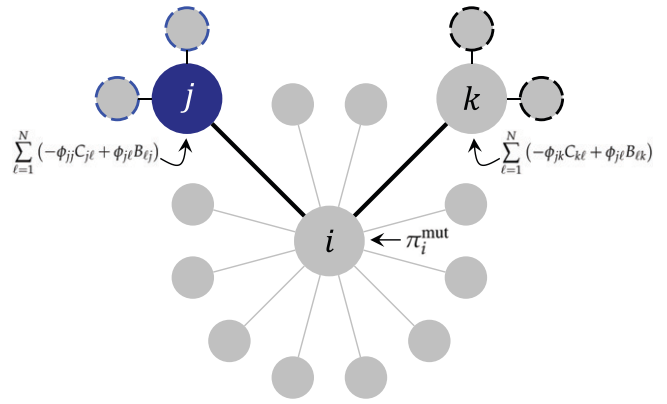
Fig. 1. Intuition for the effects of selection on trait frequency. A location, i, is first chosen with probability proportional to its mutation-weighted reproductive value, π_i^{mut}. Although all individuals are replaced with the same probability under dB updating, π_i^{mut} quantifies how important i is as a location for a social trait to be propagated. If i is replaced, then neighbors compete to fill the vacancy. Since we are concerned with producers of a social good (blue), the relevant comparison is between a neighboring producer and other neighbors. We choose two neighbors, j and k, with probability A_{ij} and A_{ik}, respectively, which represent the probabilities of j and k replacing i under neutral drift (conditioned on i being replaced). If a producer is placed at j, then the probability that another node, ℓ, is a producer is equal to the probability that j and ℓ are identical by state, $\phi_{j\ell}$. Using this distribution, we can calculate the expected payoffs for the two neighbors, j and k, which are depicted next to each of the corresponding nodes. It is the difference between the payoff to the producer, j, and the other neighbor, k, that determines the effects of selection on the mean frequency of producers. The larger this difference is, the more strongly favored producers are.

$$ \langle x_S \rangle = \frac{1}{2} + \delta\frac{1-u}{2Nu}\sum_{i,j,k=1}^N \pi_i^{\text{mut}} A_{ij} A_{ik} \Bigg($$

$$ \underbrace{\sum_{\ell=1}^N (-\phi_{jj}C_{j\ell} + \phi_{j\ell}B_{\ell j})}_{\text{expected payoff to } j, \text{ given } j \text{ is S}} $$

$$ - \underbrace{\sum_{\ell=1}^N (-\phi_{jk}C_{k\ell} + \phi_{j\ell}B_{\ell k})}_{\text{expected payoff to } k, \text{ given } j \text{ is S}} \Bigg) + O(\delta^2). \qquad [16] $$

This expression has a simple interpretation: As with Eq. **10**, the consequences of selection at location i are scaled by the mutation-weighted reproductive value, π_i^{mut}. Each pair of other locations j and k affects what happens at i, according first to the probabilities A_{ij} and A_{ik}. Next, as depicted in Fig. 1, the selective effect of a producer at j depends on the expected payoff to j compared to that of the other neighbor at k. Here, i represents the individual chosen for death, and producers are favored when a neighboring producer is more successful than its other competitors, on average.

As Fig. 2*A* shows, producers of ff-goods can be favored, even when $b < c$. This seemingly paradoxical behavior is consistent with what is observed for ff-goods from an analysis of fixation probabilities in the weak-mutation limit (63)—namely, that they benefit from the existence of a small number of highly connected hubs. When producers are rare and are introduced at a well-connected location, they are selected against, but drift can lead to the establishment of a small colony of reciprocating producers among the neighboring nodes. Even when $b < c$, this can result in a successful producer at the hub because they receive multiple benefits (or partial benefits due to division) and pay only a single cost. Having a relatively large payoff, selection then favors the propagation of the hub's offspring (producers) to the periphery,
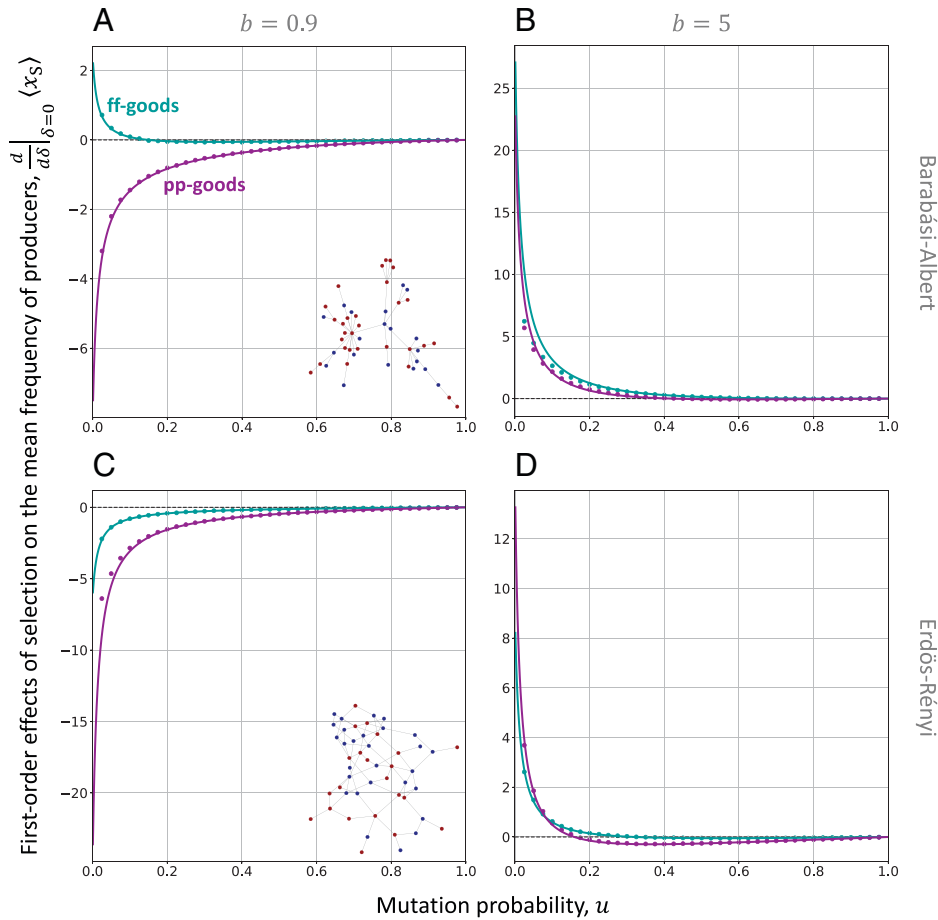
Fig. 2. Effects of selection on the mean frequencies of simple prosocial behaviors. The solid lines indicate calculations performed by using the structure-coefficient theorem (Eqs. **4** and **5**). Dots depict the simulated first-order effects of selection, $(\langle x_S \rangle - 1/2)/\delta$, where $\langle x_S \rangle$ is calculated for a selection intensity of $\delta = 0.05$ by observing the mean frequency of S over 10^8 generations. *A* and *B* show a single Barabási–Albert (preferential attachment) graph of size $N = 50$, with $b < c$ in *A* and $b > c$ in *B*. Even when $b < c$ (here, $b = 0.9$ and $c = 1$), we see that producers of ff-goods can be favored in the mutation–selection equilibrium. *C* and *D* show a single Erdös–Rényi graph of size $N = 50$, with $b < c$ in *C* and $b > c$ in *D*. When $b > c$ (here, $b = 5$ and $c = 1$), whether selection favors producers of ff-goods more than those of pp-goods can depend on the mutation probability. In this example, producers of pp-goods are favored more when the mutation probability is small, but this ranking reverses for slightly larger values of u (even when both are favored relative to neutral drift). The scales of the vertical axes vary to best illustrate the results.

provided that the mutation probability is not too large, which serves to strengthen the hub's selective advantage. Similarly, when producers are widespread, a nonproducer mutant is more likely to arise as a neighbor of this highly connected individual than at the hub itself. Since the hub then has an extremely large payoff, owing to the existence of many separate donations, each of benefit b (or close to b), the hub can easily replace the mutant nonproducer by a producer. Producers therefore thrive by taking over highly connected hubs and populating the neighborhoods with copies of themselves. These neighboring locations are likely to hold producers for long periods of time, which gives them more opportunities to spread their offspring to other locations in the population, even when at a reproductive disadvantage relative to other competitors.

However, this advantage enjoyed by producers of ff-goods diminishes to zero and even becomes (slightly) negative as the mutation probability increases. For one thing, a high mutation probability means that a central producer is more likely to propagate nonproducers to neighboring positions. It also means that when a highly connected producer dies, it is more likely to be replaced by a nonproducer via mutation alone, even when all (or most) neighbors are producers. The detrimental effects of increased mutation probabilities are seen in all examples for which producers of ff-goods or pp-goods are favored when mutation

probabilities are small (Fig. 2 *A*, *B*, and *D*), both when $b < c$ and $b > c$.

When the degree distribution is more tightly concentrated around the mean, the population is less likely to have highly connected hubs surrounded by individuals of small degree. As a result, neighboring producers must share their benefit among a larger number of two-step neighbors. Consequently, the argument we made for why producers can be favored even when $b < c$ on Barabási–Albert graphs breaks down on Erdös–Rényi graphs (Fig. 2*C*). In addition, Fig. 2*D* shows that even when $b > c$ and producers of both kinds of goods are favored (i.e., for mutation probabilities that are not too high), the ranking of which one is favored more can be reversed by changing the mutation probability. Thus, the noise introduced by mutation does not affect pp-goods and ff-goods uniformly. Fig. 2 *A*, *B*, and *D* also demonstrates that for a single kind of good, there need not be any monotonicity in the effects of selection as the mutation probability increases: Increasing u can be harmful to the abundance of producers when u is small, but beneficial to the abundance of producers when u is larger. Finally, and not surprisingly, selection via any kind of game becomes completely ineffective as $u \to 1$.

The differences between these two kinds of population structures, in terms of how they affect mutation–selection dynamics, can be summarized by their associated probabilities of identity by

descent (Eqs. **15** and **16** and Fig. 1). In Fig. 1, if j, a producer, has neighbors (blue dashed circles) who are themselves not highly connected (e.g., for the structure in Fig. 2 *A* and *B*), then j is likely to replace these neighbors when they are selected for death. As a result, j and its immediate neighbors are more likely to be identical by state, provided that u is not too large, which means that the neighbors produce benefits and provide relatively large fractions of their benefits to j. In contrast, the neighbors (black dashed circles) of a competitor, k, are generally less likely to be identical by state to j because paths from j to k (and k's neighbors) go through much more highly connected hubs. Thus, even though k itself is less likely to pay a cost, c, than j, k also receives fewer benefits than j, on average, resulting in a competitive advantage for j in replacing i. For structures like that of Fig. 2 *C* and *D*, the same basic logic of Fig. 1 holds, but now the degree distribution is different, and we see fewer instances of highly connected nodes being surrounded by much-less-connected individuals. As a result, the ratio b/c must generally be larger in order to offset lower probabilities of identity by state, larger net costs paid by j (in the case of pp-goods), and smaller shares of goods being donated to j by neighbors (in the case of ff-goods).

In *SI Appendix*, we include a version of Fig. 2 with even weaker selection ($\delta = 0.01$ instead of $\delta = 0.05$), which, as expected, shows even better agreement with the calculations (*SI Appendix*, Fig. SI.1). We also illustrate the effects of larger values of δ (*SI Appendix*, Figs. SI.2 and SI.3) to assess the departures from these predictions as the strength of selection increases.

Discussion

Since its genesis in modeling animal conflict (31, 76), evolutionary game theory has proven itself to be a tremendous resource in the study of social evolution. But any useful modeling framework necessitates the development of methods and tools for its analysis. Deterministic approaches, such as those based on the replicator equation (77), render the resulting models amenable to standard techniques from the theory of dynamical systems (78), but may neglect important evolutionary phenomena. Stochastic perspectives on modeling population dynamics can incorporate the effects of drift and fine-grained spatial structure, but several practical complications also arise. Outside of the realm of weak selection, there need not be any efficient algorithm for evaluating a population's evolutionary dynamics in general (79), a conclusion that has prompted other kinds of computational techniques, such as the development of more efficient algorithms for numerical simulation (44). Our contribution here is a method to study frequencies of types at stationarity under weak selection.

Our focus on weak selection involves the assumption that replacement probabilities are smooth functions of the selection intensity, δ. One consequence is that the system under small δ is qualitatively similar to the system under neutral drift ($\delta = 0$). For the latter, we assume that replacement probabilities are independent of the state, $\mathbf{x}$, in the absence of selection. We also make a technical assumption that the population evolves as a coherent whole, which ensures that there exists a unique stationary distribution to study in the first place. These assumptions on the model are formally stated in *SI Appendix*. Of course, although they are biologically reasonable, these assumptions also limit the scope of the model and should be taken into consideration when using our results to go beyond the examples illustrated here.

Another limitation is that, although the population can have arbitrary size and structure, both are fixed as the evolutionary process unfolds. Dynamic population structures are both realistic and

recognized (largely through numerical simulations) as having a strong impact on the evolution of prosocial traits like cooperation (30, 80–82). Related to this point, an interesting extension would account for feedback between individuals and the environment as a way to capture ecosystem engineering (83–85). From a computational perspective, it would also be intriguing to explore how these components affect the complexity of quantitative analysis. These considerations are among many that would be necessary for a deeper understanding of social evolution (86).

In our examples, we find that the effects of selection on producers of two different kinds of social goods can be quite sensitive to mutation probabilities. This effect is conceptually similar to the ways in which the infidelity of Mendelian transmission of diploid phenotypes can alter evolutionary game-theoretic predictions relative to those based on simple haploid systems (87, 88). We compare goods that are produced with benefits and costs in proportion to the number of interaction partners (pp-goods) to goods that are produced with fixed benefits and costs then get distributed among interaction partners (ff-goods). We show that increasing the mutation probability can reverse the direction of selection and can change the relative ranking of the two types, even when producers of both kinds of goods are favored by selection. Qualitatively, one thing that the analysis of social goods on heterogeneous graphs yields is the fact that mean frequencies need not be monotonic in the mutation probability. Traulsen et al. (60) also found nonmonotonicity, but only in games with more than two traits. Débarre (61) observed only monotonic effects of mutation probabilities on mean frequencies, but the structures considered there are homogeneous, which can obscure the difference between social goods. Nonmonotonicity may be inferred from the results of Débarre (62) for small to moderate migration rates (changing who interacts with whom) in an island model of population structure.

The importance of high mutation rate has long been recognized in evolutionary theory. If it is high enough, the resulting "quasispecies" crosses an error threshold, beyond which adaptation is impossible (89). This model has been used to study human immunodeficiency viruses, which are highly diverse, owing to rapid mutation (90). The replicator equation has been generalized to include mutation (91), bridging the gap between replicator dynamics (77) for frequency-dependent selection (without mutation) and quasispecies evolution under strong mutation and frequency-independent selection. A notable application is to language learning in populations (92). In simple (genetically single-locus) models, increased mutation may just reduce the efficacy of selection, eventually resulting in neutral evolution, as we find here. But it is important to recognize that empirical situations are likely more complicated because increased mutation rates apply to all loci across the genome and many mutations are deleterious; for one example, see Sprouffske et al. (93).

When the population is heterogeneous, the distribution of social goods naturally gives rise to asymmetric games. Asymmetric contests have an extensive history in theoretical biology, dating back nearly half a century. In a pioneering theoretical work on animal behavior, Maynard Smith and Parker (94) showed that while evolutionarily stable states are frequently mixed in symmetric contests, variability in traits and behaviors is actually the exception in asymmetric contests. Furthermore, they argued that asymmetric interactions should be taken as the norm, either as the result of organismal differences like size or strength or as the result of "uncorrelated" asymmetries, such as those that arise when one individual discovers a resource and another is a latecomer.

One takeaway from earlier studies on asymmetric games, as well as from our own examples of social goods, is that there is

unequivocally both a loss of generality and a loss of realism in approximating interactions by symmetric games. The assumption of homogeneous interactions severely limits the scope of a model. This aspect of social evolution is still relatively underexplored in the literature. For example, the spatially mediated heterogeneity arising in the production of ff-goods underscores, among other things, the need for an asymmetric version of the structure-coefficient theorem. Even though no mathematical model can accurately capture every aspect of a population evolving in the natural world, striving for refinements that are simultaneously realistic and amenable to analysis is necessary for a deeper understanding of population dynamics.

1. R. B. Franklin, A. L. Mills, Multi-scale variation in spatial heterogeneity for microbial community structure in an eastern Virginia agricultural field. *FEMS Microbiol. Ecol.* **44**, 335–346 (2003).
2. H. Harmsen, D. Prieur, C. Jeanthon, Distribution of microorganisms in deep-sea hydrothermal vent chimneys investigated by whole-cell hybridization and enrichment culture of thermophilic subpopulations. *Appl. Environ. Microbiol.* **63**, 2876–2883 (1997).
3. A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, B. Bhattacharjee, "Measurement and analysis of online social networks" in *IMC '07: Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement* (Association for Computing Machinery, New York, 2007), 29–42.
4. J. K. Lee, J. Choi, C. Kim, Y. Kim, Social media, network heterogeneity, and opinion polarization. *J. Commun.* **64**, 702–722 (2014).
5. W. Jo, D. Chang, M. You, G.-H. Ghim, A social network analysis of the spread of COVID-19 in South Korea and policy implications. *Sci. Rep.* **11**, 8581 (2021).
6. A. Nande, B. Adlam, J. Sheen, M. Z. Levy, A. L. Hill, Dynamics of COVID-19 under social distancing measures are driven by transmission network structure. *PLOS Comput. Biol.* **17**, e1008684 (2021).
7. A. L. Hagedoorn, A. C. Hagedoorn, *The Relative Value of the Processes Causing Evolution* (Martinus Nijhoff, The Hague, Netherlands, 1921).
8. S. Wright, Evolution in Mendelian populations. *Genetics* **16**, 97–159 (1931).
9. S. Wright, Isolation by distance. *Genetics* **28**, 114–138 (1943).
10. G. Malécot, *Les Mathématiques de l'Hérédité* (Masson, Paris, 1948). Extended translation: *The Mathematics of Heredity* (W. H. Freeman, San Francisco, 1969).
11. M. Kimura, G. H. Weiss, The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* **49**, 561–576 (1964).
12. T. Maruyama, Analysis of population structure. I. One-dimensional stepping-stone models of finite length. *Ann. Hum. Genet.* **34**, 201–219 (1970).
13. T. Maruyama, Analysis of population structure. II. Two-dimensional stepping stone models of finite length and other geographically structured populations. *Ann. Hum. Genet.* **35**, 179–196 (1971).
14. M. Nakamaru, H. Matsuda, Y. Iwasa, The evolution of cooperation in a lattice-structured population. *J. Theor. Biol.* **184**, 65–81 (1997).
15. M. Slatkin, Gene flow in natural populations. *Annu. Rev. Ecol. Syst.* **16**, 393–430 (1985).
16. M. Slatkin, Gene flow and the geographic structure of natural populations. *Science* **236**, 787–792 (1987).
17. J. L. Cherry, Selection in a subdivided population with dominance or local frequency dependence. *Genetics* **163**, 1511–1518 (2003).
18. D. Roze, F. Rousset, Selection and drift in subdivided populations: A straightforward method for deriving diffusion approximations and applications involving dominance, selfing and local extinctions. *Genetics* **165**, 2153–2166 (2003).
19. W. D. Hamilton, The genetical evolution of social behaviour. I. *J. Theor. Biol.* **7**, 1–16 (1964).
20. W. D. Hamilton, The genetical evolution of social behaviour. II. *J. Theor. Biol.* **7**, 17–52 (1964).
21. A. Grafen, The hawk-dove game played between relatives. *Anim. Behav.* **27**, 905–907 (1979).
22. W. G. S. Hines, J. M. Smith, Games between relatives. *J. Theor. Biol.* **79**, 19–30 (1979).
23. Y. Tao, S. Lessard, Frequency-dependent selection in sexual family-structured populations. *J. Theor. Biol.* **217**, 525–534 (2002).
24. C. Taylor, M. A. Nowak, Evolutionary game dynamics with non-uniform interaction rates. *Theor. Popul. Biol.* **69**, 243–252 (2006).
25. M. J. Wade, An experimental study of kin selection. *Evolution* **34**, 844–855 (1980).
26. B. Kerr, M. A. Riley, M. W. Feldman, B. J. M. Bohannan, Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors. *Nature* **418**, 171–174 (2002).
27. J. Lehmann, A. H. Korstjens, R. I. M. Dunbar, Group size, grooming and social cohesion in primates. *Anim. Behav.* **74**, 1617–1629 (2007).
28. G. S. Wilkinson, Reciprocal food sharing in the vampire bat. *Nature* **308**, 181–184 (1984).
29. M. O. Jackson, A. Wolinsky, A strategic model of social and economic networks. *J. Econ. Theory* **71**, 44–74 (1996).
30. E. Akçay, Collapse and rescue of cooperation in evolving dynamic networks. *Nat. Commun.* **9**, 2692 (2018).
31. J. Maynard Smith, *Evolution and the Theory of Games* (Cambridge University Press, Cambridge, UK, 1982).
32. M. Broom, J. Rychtář, *Game-Theoretical Models in Biology* (Taylor & Francis, Milton Park, UK, 2013).
33. E. Lieberman, C. Hauert, M. A. Nowak, Evolutionary dynamics on graphs. *Nature* **433**, 312–316 (2005).
34. H. Ohtsuki, M. A. Nowak, J. M. Pacheco, Breaking the symmetry between interaction and replacement in evolutionary dynamics on graphs. *Phys. Rev. Lett.* **98**, 108106 (2007).
35. H. Ohtsuki, J. M. Pacheco, M. A. Nowak, Evolutionary graph theory: Breaking the symmetry between interaction and replacement. *J. Theor. Biol.* **246**, 681–694 (2007).
36. M. Broom, J. Rychtář, An analysis of the fixation probability of a mutant on special classes of non-directed graphs. *Proc.- Royal Soc., Math. Phys. Eng. Sci.* **464**, 2609–2627 (2008).
37. H. Ohtsuki, C. Hauert, E. Lieberman, M. A. Nowak, A simple rule for the evolution of cooperation on graphs and social networks. *Nature* **441**, 502–505 (2006).
38. G. Szabó, G. Fáth, Evolutionary games on graphs. *Phys. Rep.* **446**, 97–216 (2007).
39. P. D. Taylor, T. Day, G. Wild, Evolution of cooperation in a finite homogeneous graph. *Nature* **447**, 469–472 (2007).
40. Y.-T. Chen, Sharp benefit-to-cost rules for the evolution of cooperation on regular graphs. *Ann. Appl. Probab.* **23**, 637–664 (2013).
41. F. Débarre, C. Hauert, M. Doebeli, Social evolution in structured populations. *Nat. Commun.* **5**, 3409 (2014).
42. C. Mullon, L. Lehmann, The robustness of the weak selection approximation for the evolution of altruism against strong selection. *J. Evol. Biol.* **27**, 2272–2282 (2014).
43. Y.-T. Chen, A. McAvoy, M. A. Nowak, Fixation probabilities for any configuration of two strategies on regular graphs. *Sci. Rep.* **6**, 39181 (2016).
44. L. Hindersin, B. Wu, A. Traulsen, J. García, Computation and simulation of evolutionary game dynamics in finite populations. *Sci. Rep.* **9**, 6946 (2019).
45. P. D. Taylor, How to make a kin selection model. *J. Theor. Biol.* **180**, 27–37 (1996).
46. F. Rousset, S. Billiard, A theoretical basis for measures of kin selection in subdivided populations: Finite populations and localized dispersal. *J. Evol. Biol.* **13**, 814–825 (2000).
47. M. A. Nowak, A. Sasaki, C. Taylor, D. Fudenberg, Emergence of cooperation and evolutionary stability in finite populations. *Nature* **428**, 646–650 (2004).
48. S. Lessard, V. Ladret, The probability of fixation of a single mutant in an exchangeable selection model. *J. Math. Biol.* **54**, 721–744 (2007).
49. T. Antal, H. Ohtsuki, J. Wakeley, P. D. Taylor, M. A. Nowak, Evolution of cooperation by phenotypic similarity. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 8597–8600 (2009).
50. R. Durrett, Spatial evolutionary games with small selection coefficients. *Electron. J. Probab.* **19**, 1–64 (2014).
51. D. Fudenberg, L. A. Imhof, Imitation processes with small mutations. *J. Econ. Theory* **131**, 251–262 (2006).
52. J. Van Cleve, Social evolution and genetic interactions in the short and long term. *Theor. Popul. Biol.* **103**, 2–26 (2015).
53. B. Allen *et al.*, Evolutionary dynamics on any population structure. *Nature* **544**, 227–230 (2017).
54. A. McAvoy, B. Allen, Fixation probabilities in evolutionary dynamics under weak selection. *J. Math. Biol.* **82**, 14 (2021).
55. T. Antal, I. Scheuring, Fixation of strategies for an evolutionary game in finite populations. *Bull. Math. Biol.* **68**, 1923–1944 (2006).
56. B. Wu, C. S. Gokhale, L. Wang, A. Traulsen, How small are small mutation rates? *J. Math. Biol.* **64**, 803–827 (2012).
57. S. A. Frank, The Price equation, Fisher's fundamental theorem, kin selection, and causal analysis. *Evolution* **51**, 1712–1729 (1997).
58. C. E. Tarnita, H. Ohtsuki, T. Antal, F. Fu, M. A. Nowak, Strategy selection in structured populations. *J. Theor. Biol.* **259**, 570–581 (2009).
59. C. E. Tarnita, N. Wage, M. A. Nowak, Multiple strategies in structured populations. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 2334–2337 (2011).
60. A. Traulsen, C. Hauert, H. De Silva, M. A. Nowak, K. Sigmund, Exploration dynamics in evolutionary games. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 709–712 (2009).
61. F. Débarre, Fidelity of parent-offspring transmission and the evolution of social behavior in structured populations. *J. Theor. Biol.* **420**, 26–35 (2017).
62. F. Débarre, Imperfect strategy transmission can reverse the role of population viscosity on the evolution of altruism. *Dyn. Games Appl.* **10**, 732–763 (2020).
63. A. McAvoy, B. Allen, M. A. Nowak, Social goods dilemmas in heterogeneous societies. *Nat. Hum. Behav.* **4**, 819–831 (2020).
64. K. Sigmund, *The Calculus of Selfishness* (Princeton University Press, Princeton, NJ, 2010).
65. B. Allen, A. McAvoy, A mathematical formalism for natural selection with arbitrary spatial and genetic structure. *J. Math. Biol.* **78**, 1147–1210 (2019).
66. B. Allen, C. E. Tarnita, Measures of success in a class of evolutionary models with fixed population size and structure. *J. Math. Biol.* **68**, 109–143 (2014).
67. R. A. Fisher, *The Genetical Theory of Natural Selection* (Clarendon Press, Oxford, UK, 1930).
68. G. R. Price, C. A. B. Smith, Fisher's Malthusian parameter and reproductive value. *Ann. Hum. Genet.* **36**, 1–7 (1972).
69. P. D. Taylor, Allele-frequency change in a class-structured population. *Am. Nat.* **135**, 95–106 (1990).
70. B. Derrida, S. C. Manrubia, D. H. Zanette, Distribution of repetitions of ancestors in genealogical trees. *Physica A* **281**, 1–16 (2000).
71. S. Engen, R. Lande, B.-E. Saether, F. S. Dobson, Reproductive value and the stochastic demography of age-structured populations. *Am. Nat.* **174**, 795–804 (2009).
72. N. H. Barton, A. M. Etheridge, The relation between reproductive value and genetic contribution. *Genetics* **188**, 953–973 (2011).
73. W. Maciejewski, Reproductive value in graph-structured populations. *J. Theor. Biol.* **340**, 285–293 (2014).
74. C. Taylor, D. Fudenberg, A. Sasaki, M. A. Nowak, Evolutionary game dynamics in finite populations. *Bull. Math. Biol.* **66**, 1621–1644 (2004).
75. W. Maciejewski, F. Fu, C. Hauert, Evolutionary game dynamics in populations with heterogenous structures. *PLOS Comput. Biol.* **10**, e1003567 (2014).
76. J. Maynard Smith, G. R. Price, The logic of animal conflict. *Nature* **246**, 15–18 (1973).
77. P. D. Taylor, L. B. Jonker, Evolutionary stable strategies and game dynamics. *Math. Biosci.* **40**, 145–156 (1978).
78. J. Hofbauer, K. Sigmund, *Evolutionary Games and Population Dynamics* (Cambridge University Press, Cambridge, UK, 1998).
79. R. Ibsen-Jensen, K. Chatterjee, M. A. Nowak, Computational complexity of ecological and evolutionary spatial dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 15636–15641 (2015).
80. L. Wardil, C. Hauert, Origin and structure of dynamic cooperative networks. *Sci. Rep.* **4**, 5725 (2014).

81. T. Khoo, F. Fu, S. Pauls, Coevolution of cooperation and partner rewiring range in spatial social networks. *Sci. Rep.* **6**, 36293 (2016).

82. H.-W. Lee, N. Malik, P. J. Mucha, Evolutionary prisoner's dilemma games coevolving on adaptive networks. *J. Complex Netw.* **6**, 1–23 (2018).

83. C. Hauert, C. Saade, A. McAvoy, Asymmetric evolutionary games with environmental feedback. *J. Theor. Biol.* **462**, 347–360 (2019).

84. S. Estrela *et al.*, Environmentally mediated social dilemmas. *Trends Ecol. Evol.* **34**, 6–18 (2019).

85. A. R. Tilman, J. B. Plotkin, E. Akçay, Evolutionary games with environmental feedbacks. *Nat. Commun.* **11**, 915 (2020).

86. E. Akçay, Deconstructing evolutionary game theory: Coevolution of social behaviors with their evolutionary setting. *Am. Nat.* **195**, 315–330 (2020).

87. A. Grafen, The formal Darwinism project in outline. *Biol. Philos.* **29**, 155–174 (2014).

88. H. Rubin, The phenotypic gambit: Selective pressures and ESS methodology in evolutionary game theory. *Biol. Philos.* **31**, 551–569 (2016).

89. M. Eigen, J. McCaskill, P. Schuster, Molecular quasi-species. *J. Phys. Chem.* **92**, 6881–6891 (1988).

90. M. A. Nowak, What is a quasispecies? *Trends Ecol. Evol.* **7**, 118–121 (1992).

91. K. M. Page, M. A. Nowak, Unifying evolutionary dynamics. *J. Theor. Biol.* **219**, 93–98 (2002).

92. N. L. Komarova, Replicator-mutator equation, universality property and population dynamics of learning. *J. Theor. Biol.* **230**, 227–239 (2004).

93. K. Sprouffske, J. Aguilar-Rodríguez, P. Sniegowski, A. Wagner, High mutation rates limit evolutionary adaptation in *Escherichia coli*. *PLoS Genet.* **14**, e1007324 (2018).

94. J. Maynard Smith, G. A. Parker, The logic of asymmetric contests. *Anim. Behav.* **24**, 159–175 (1976).

95. A. McAvoy, J. Wakeley, Code from Evaluating the structure-coefficient of evolutionary game theory. GitHub. https://github.com/alexmcavoy/sigma. Deposited 21 July 2021.