Social goods dilemmas in heterogeneous societies

Alex McAvoy[1,*]   Benjamin Allen[2]   Martin A. Nowak[1,3,*]

[1]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138 USA;
[2]Department of Mathematics, Emmanuel College, Boston, MA 02115 USA;
[3]Department of Mathematics, Harvard University, Cambridge, MA 02138 USA

Abstract

Prosocial behaviors are encountered in the donation game, the prisoner's dilemma, relaxed social dilemmas, and public goods games. Many studies assume that the population structure is homogeneous, meaning all individuals have the same number of interaction partners, or that the social good is of one particular type. Here, we explore general evolutionary dynamics for arbitrary spatial structures and social goods. We find that heterogeneous networks, wherein some individuals have many more interaction partners than others, can enhance the evolution of prosocial behaviors. However, they often accumulate most of the benefits in the hands of a few highly-connected individuals, while many others receive low or negative payoff. Surprisingly, selection can favor producers of social goods even if the total costs exceed the total benefits. In summary, heterogeneous structures have the ability to strongly promote the emergence of prosocial behaviors, but they also create the possibility of generating large inequality.

1   Introduction

Prosocial behaviors are often studied using two-player or many-player games. In the first case, we encounter the donation game (Sigmund, 2010, Radzvilavicius et al., 2019), prisoners dilemma (Axelrod, 1984, Szabó and Tőke, 1998, Abramson and Kuperman, 2001, Broom and Rychtář, 2013), or relaxed social dilemmas (Maynard Smith, 1982, Hauert and Doebeli, 2004, Doebeli et al., 2004). In the second case, we are typically in the world of public goods games (Lloyd, 1833, Hardin, 1968, Szabó and Hauert, 2002, Pinheiro et al., 2014, Peña et al., 2016a, Zhong et al., 2017). In both kinds of games, it is usually assumed that players are in identical positions and affect all others equally. This homogeneity can be a consequence of the spatial structure of the population; for example, all individuals might have the same number of neighbors. However, even within spatially-heterogeneous populations, it is often assumed that every group (or pair)

---

*Corresponding authors: Alex McAvoy (`alexmcavoy@g.harvard.edu`) and Martin A. Nowak (`martin_nowak@harvard.edu`).

plays the same game. In this study, we consider "social goods dilemmas" in which individuals may pay a cost to produce a good that benefits their neighbors. In social goods dilemmas, the distribution of benefits and costs can depend on the population structure as well as on the nature of the good itself. If some individuals are central and well-connected within a group, while others are peripheral, social goods dilemmas lead to heterogeneous game structures with surprising evolutionary dynamics.

For social goods produced within an interaction structure, two questions become immediately apparent: *(i)* is the benefit of receiving the social good from a specific donor independent of the number of recipients (non-rival), or does one neighbor's access to the good decrease that of another (rival)? and *(ii)* is the cost of producing a social good a function of the number of recipients, or is it fixed? In the traditional setting of homogeneous population structures (Nowak and May, 1992, Nakamaru et al., 1997, Lieberman et al., 2005, Ohtsuki et al., 2006, Taylor et al., 2007, Chen, 2013, Débarre et al., 2014), there is no reason to consider these cases separately since the differences between them amount to a simple rescaling of the benefits and/or costs. However, important distinctions among those social goods arise in heterogeneous societies (Santos and Pacheco, 2005, Antal et al., 2006, Gómez-Gardenes et al., 2007, Sood et al., 2008, Cao et al., 2010, Maciejewski et al., 2014, Fan et al., 2017, Allen et al., 2017). In fact, distinguishing among various kinds of social goods is a common practice in economics Goldfarb and Tucker (2019), one that has not fully permeated evolutionary game theory (Santos et al., 2008, Li et al., 2013). The simplest dichotomy is between benefits that are proportional ("p") to the number of recipients and those that are fixed ("f"). The same two options for the cost of a good gives four types of social goods, representing the combinations of benefits and costs: pp (proportional benefits, proportional costs), ff (fixed benefits, fixed costs), pf (proportional benefits, fixed costs), and fp (fixed benefits, proportional costs). Our primary focus here is on pp-, ff-, and pf-goods, which are summarized in Figure 1.

As an example, consider the prosocial act of donating blood (Stutzer et al., 2011). One recipient's use of blood decreases that of another, so blood is a rival good. Blood is also divisible, and a fixed volume of it can be distributed among several individuals in need. However, the nature of this donation as a social good depends not only on the good itself (blood) but also on how the behavior is expressed within the population. A donor might attempt to give each individual in need as much blood as possible, potentially incurring a huge cost for doing so. But they might also decide on a more modest, fixed donation, to be divided evenly among those in need (and possibly supplemented by donations from others). The former case is modeled better as a pp-good, while the latter could be viewed as an ff-good. Similar arguments can be made for other kinds of social behaviors in human communities, such as helping out coworkers, volunteering at a charity, and donating money. In nonhuman societies, relevant examples include social grooming among primates (Dunbar, 1991), food delivery among magpies (Horn et al., 2016), and blood donation (as food) among vampire bats (Wilkinson, 1984).

All neighbors might also be able to benefit from a good in its entirety, even if production of the good entails a fixed cost. Volunteering to maintain a public space, such as a park, is one such example. The cost for doing so can be quantified in terms of time, effort, or money (e.g. purchasing supplies or hiring a groundskeeper). Barring extenuating circumstances, the benefit of having a clean park is not necessarily reduced by another person's use of the space; this can therefore be seen as a pf-good. Such is also the case for information transmission within a social
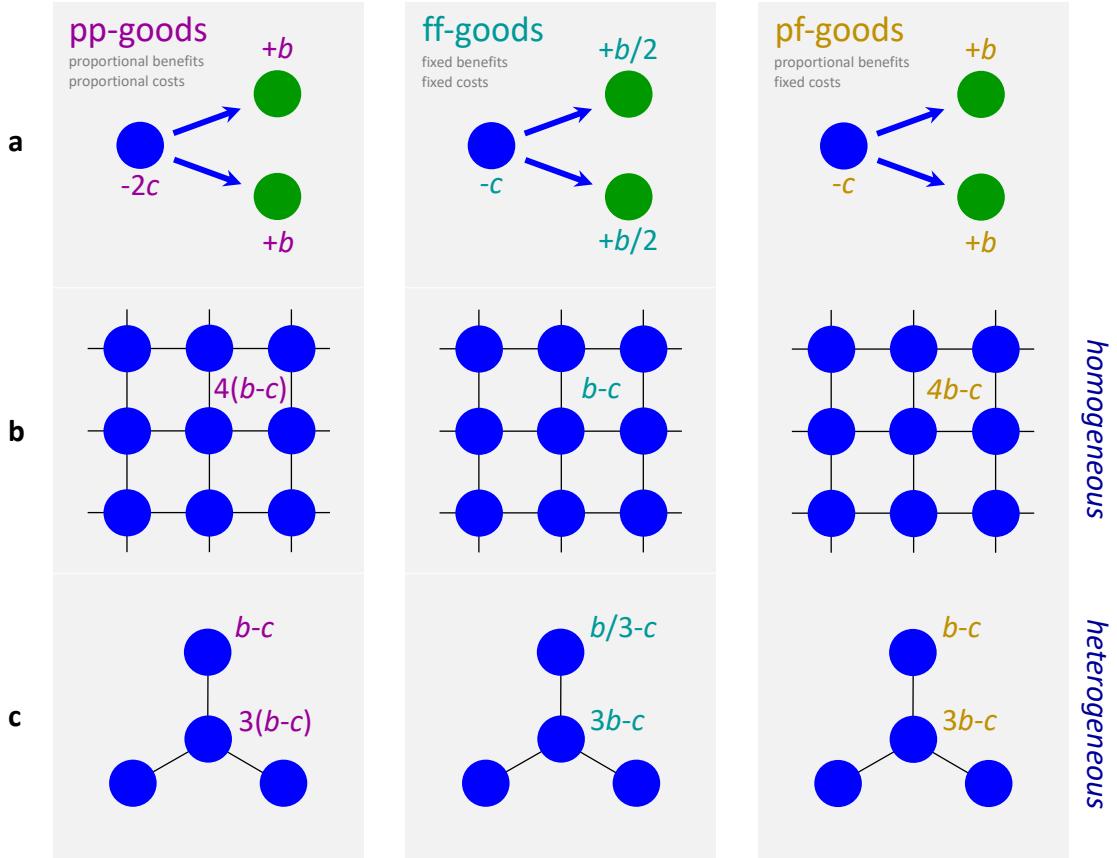
Figure 1: Social goods and prosocial behaviors. a, For pp-goods, a producer pays cost c for each neighbor to receive benefit b. For ff-goods, a producer pays a fixed cost c irrespective of the number of neighbors, k; each neighbor receives benefit b/k. For pf-goods, the cost, c, is again independent of the number of neighbors, but now each neighbor gets to enjoy the benefit, b, in its entirety. **b**, On a regular graph, such as a two-dimensional grid, all individuals have the same number of neighbors; here $k = 4$. For pp-goods, each individual receives payoff $4(b - c)$. For ff-goods, each individual receives payoff $b - c$. Therefore, on regular graphs the payoffs arising for pp-goods and ff-goods are equivalent up to rescaling both b and c. Scaling only b (resp. c) gives an equivalence between pf-goods and ff-goods (resp. pf-goods and pp-goods). **c**, On heterogeneous population structures, such as the star, the three kinds of social goods lead to distinct payoff distributions, and one cannot be obtained from another by rescaling b and/or c. Therefore, heterogeneous graphs highlight important differences among social goods.

3

network, whose initial acquisition could entail a cost but whose full benefit can be enjoyed by more than just a single individual. Entertainment, such as podcasts, radio programs, and video streaming, can also be non-rival (and, in fact, pf-) goods. Publication of a novel scientific finding or method and development of open-source software ordinarily represent pf-goods as well.

The remaining class, fp-goods, is somewhat less natural than the other three because the per-capita benefit decreases with the number of recipients, while the overall cost grows. One way in which such a cost structure might arise is via the production of a divisible, rival good that involves a cost associated to its transmission to a recipient. That being said, we focus our examples primarily on pp-, ff-, and pf-goods since all of the interesting behavior we observe can be illustrated using these kinds of social goods. Our theoretical results cover a much broader class of social goods, however, and we discuss how they can be used to understand the effects of general functional dependencies, asymmetric games, and stochastic payoffs on selection.

2 Results

In our model, the interaction structure of the game is given by a graph (or social network) of size N, in which individuals occupy nodes. The adjacency matrix of this graph satisfies $w_{ij} = 1$ if i and j are neighbors and $w_{ij} = 0$ if i and j are not neighbors. The links specify interactions between individuals. Each individual can choose between two strategies. An individual with strategy A (a "producer") generates goods to distribute among neighbors and pays costs for doing so. An individual with strategy B (a "non-producer") provides no benefits to others and incurs no costs.

Let $w_i = \sum_{j=1}^{N} w_{ij}$ denote the number of neighbors of i (the "degree" of i). For a pp-good, a producer at location i pays total cost, cw_i, and the total benefit bw_i is split among the w_i neighbors; thus, each neighbor receives b. For an ff-good, the producer pays cost, c, and the total benefit b is split among the w_i neighbors; thus, each neighbor receives b/w_i. A pf-good is a hybrid of these two goods; the total cost is c and each neighbor gets b. The behavior is prosocial if both b and c are positive, which we assume throughout this study. However, we make no assumptions regarding the ranking of b and c; we allow $b > c$, $b = c$, and $b < c$.

The first question that needs to be explored is: when is a prosocial behavior wealth producing? A natural measure for total wealth is simply the sum over all benefits minus all costs, assuming everyone is a producer. Using this approach, we find that the answer for both pp- and ff-goods is immediate: on any graph, the prosocial good is wealth producing if and only if $b > c$. In contrast, pf-goods can be wealth producing even when $b < c$ since the total benefit, $b \sum_{i,j=1}^{N} w_{ij}$, is based on the number of edges and the total cost, cN, is based on the number of nodes. More specifically, a pf-good is wealth producing on a graph if and only if $b/c > N / \sum_{i,j=1}^{N} w_{ij}$.

The second question concerns inequality and possible social harm. On a heterogeneous graph, it is clear that even if everyone produces the social good, highly-connected individuals can accumulate a much higher payoff than others. Depending on the graph structure, a small number of individuals could hold the large majority of the wealth that is being produced. The poorest individuals can also end up with negative payoffs, which we call "harmful prosociality." In this case, the poorest members of the population would be better off in the all-B state than in the all-A state. For pp- and pf-goods, harmful prosociality can arise only if $b < c$, but for ff-goods it can arise even if $b > c$.

The third question is: under which conditions do producers evolve in a structured population? Since there is neither mutation nor migration in our model, we use the notion of "fixation probability" to quantify the effects of selection on a population. Let ρ_A be the probability that trait A, held initially by just a single individual within the population, eventually fixes and replaces the resident B-population. Similarly, we denote by ρ_B the fixation probability of type B, defined in the same way as ρ_A but with A and B swapped. We say that selection favors A relative to B if $\rho_A > \rho_B$. Intuitively, in a process with negligible mutation rates between the types, this condition means that the population spends more time in the all-A state than in the all-B state (Fudenberg and Imhof, 2006).

We first derive a general result that applies to almost any evolutionary update mechanism as long as some natural properties hold. Suppose that a producer at location i donates B_{ij} to j at a cost of C_{ij} (Figure 2). Let π_i be the fixation probability of a neutral trait starting in location i (also known the "reproductive value" (Fisher, 1930, Maciejewski, 2014) of i). In the SI, we define a natural distribution over the non-monomorphic states (meaning states with both types, A and B) under neutral drift, and we let x_{ij} be the probability that i and j have the same type in this distribution. Finally, let m_k^{ij} be the marginal effect of k's fecundity on the probability that i replaces j. Using these quantities, which are described in detail in the SI, we show that producers (A) are favored over non-producers (B) under weak selection (Wild and Traulsen, 2007, Fu et al., 2009, Wu et al., 2010, 2013a, Mullon and Lehmann, 2014) whenever

$$\sum_{i,j,k,\ell=1}^{N} \pi_i m_k^{ji} \left(-x_{jk} C_{k\ell} + x_{j\ell} B_{\ell k} \right) > \sum_{i,j,k,\ell=1}^{N} \pi_i m_k^{ji} \left(-x_{ik} C_{k\ell} + x_{i\ell} B_{\ell k} \right). \tag{1}$$

This condition can be evaluated by solving a linear system of $O\left(N^2\right)$ equations, giving an overall complexity of $O\left(N^6\right)$ (since solving a linear system of n equations requires $O\left(n^3\right)$ operations). We give examples and interpretations of this condition for specific update rules in Methods.

The social goods we consider here have the property that $B_{ij} = b\beta_{ij}$ and $C_{ij} = c\gamma_{ij}$ for some $b, c > 0$, where β_{ij} and γ_{ij} are independent of b and c. As a consequence, Equation 1 can be written as $\gamma b > \beta c$, where β and γ are independent of b and c. When $\gamma > 0$, this condition implies that $\rho_A > \rho_B$ in the limit of weak selection whenever $b/c > (b/c)^* = \beta/\gamma$. $(b/c)^*$ is known as the "critical benefit-to-cost ratio" for producers to evolve. If $\gamma < 0$, then the condition for producers to evolve is $b/c < (b/c)^* = \beta/\gamma$. Thus, a negative critical ratio implies that prosocial behaviors, i.e. those with $b, c > 0$, cannot evolve; instead, selection can favor spiteful behaviors with $b < 0$ and $c > 0$ ("costly harm") (Iwasa et al., 1998, Forber and Smead, 2014). This property is one nuance of critical ratios, namely that they are lower bounds on b/c when $\gamma > 0$ and upper bounds when $\gamma < 0$.

2.1 Evolutionary outcomes on heterogeneous structures

We consider several natural update rules that drive evolution through imitation. Under pairwise-comparison (PC) updating (Szabó and Tőke, 1998), a random individual is chosen to update its strategy. It compares its own payoff with that of a single, randomly chosen neighbor. If the neighbor has a higher payoff, then the focal individual adopts the neighbor's strategy. If the neighbor has a lower payoff, then the focal individual retains its current strategy. The payoff
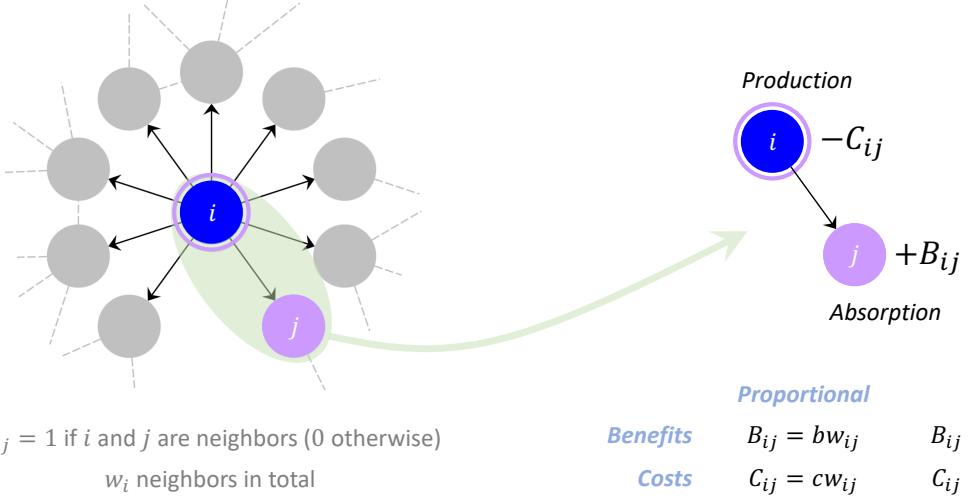
Figure 2: Production and absorption of social goods. A producer at location i donates B_{ij} to j, at a cost of C_{ij}. We give formulas for B_{ij} and C_{ij}, for each of the social goods we analyze, in terms of the adjacency matrix of the population structure, which is defined by $w_{ij} = 1$ if i and j share an edge and $w_{ij} = 0$ otherwise. For ff-goods, with fixed benefits and fixed costs, a producer at location i pays c to donate a total benefit of b, giving b/w_i to each of i's w_i neighbors. The effective cost attributable to a neighbor, j, is $C_{ij} = c/w_i$. Our main examples of benefits and costs all have the property that $B_{ij} = b\beta_{ij}$ and $C_{ij} = c\gamma_{ij}$ for some β_{ij} and γ_{ij} (which are both independent of b and c). b and c quantify the "magnitude" of the social good, while β_{ij} and γ_{ij} quantify both the nature of the good and how it is distributed by a producer.

comparison is subject to noise. Death-birth (DB) and imitation (IM) updating are similar, but they differ in the number of neighbors chosen for comparison and/or whether imitating some neighbor is compulsory (see Figure 3).

These update rules are highly idealized, but they capture important qualitative features of behavior imitation (Fudenberg and Imhof, 2008, Roca et al., 2009, Traulsen et al., 2010). For one thing, a learner is more likely to imitate a model individual's behavior as the model's payoff increases. If a learner cannot compare his or her payoff to all neighbors at once (for instance, if he or she encounters neighbors only occasionally), then PC updating is relevant. When information is more readily available, e.g. within scientific collaboration networks, then IM updating could serve as a better model. In both cases, a learner is not compelled to imitate a behavior. DB updating, which requires imitating some neighbor, could be interpreted in terms of personnel turnover within an organization, for example. An individual in the network might be replaced by a newcomer, who then copies a behavior of someone nearby. We use these (well-studied) update rules to illustrate interesting evolutionary dynamics of social goods, but we emphasize that Equation 1 can readily be applied to a wide variety of update mechanisms.

For PC updating, we prove that producers of pp-, ff-, and pf-goods are never favored on homogeneous graphs. On heterogeneous graphs, it is possible that producers evolve if the benefit-to-cost ratio exceeds a critical value, $(b/c)^*$. For pp-goods, we find that $(b/c)^*$ can never be between zero and one, which means that $b > c$ is a necessary condition for producers to evolve. Thus, for pp-goods, producers can evolve only if they improve the overall wealth of the population. Moreover, they can evolve only if they lead to a positive payoff for even the poorest individuals. In the SI, we also establish this result for pp-goods on heterogeneous graphs under
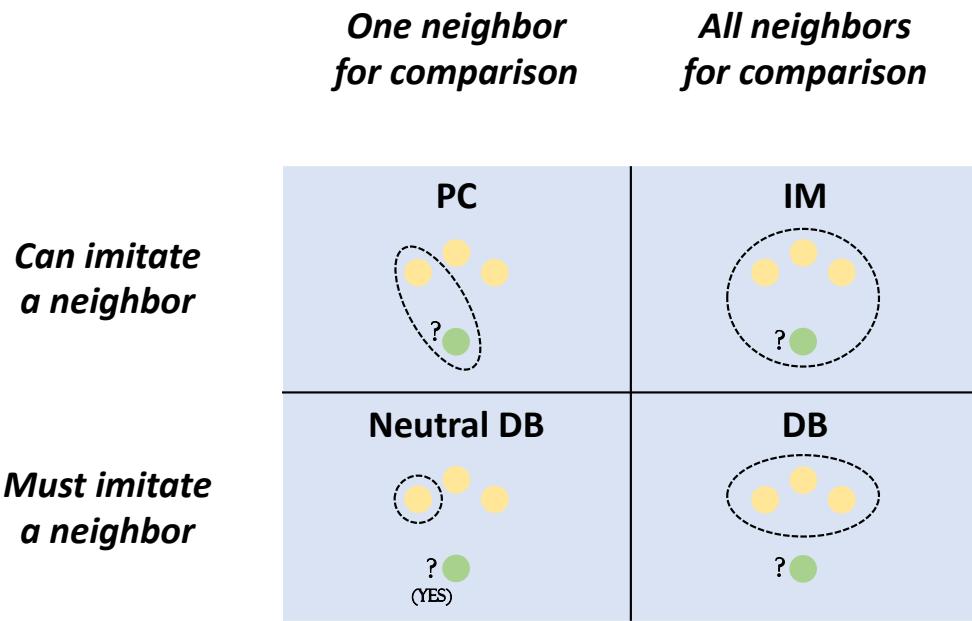
Figure 3: Four update rules driving evolutionary dynamics through imitation. When considering how to imitate a neighbor's action (*A* or *B*) based on payoff, four natural update rules arise, which have each been considered extensively in the literature. Pairwise-comparison (PC) updating involves an individual choosing a random neighbor (yellow) with whom he or she compares payoffs. There is an option to imitate the neighbor, but the focal individual (green) may also choose to retain their existing action. Imitation (IM) updating is similar to this rule, except that the payoff comparison involves a focal individual and all neighbors. Again, this individual can imitate a neighbor but does not have to do so. If one insists that one of these neighbors must be imitated, then we have death-birth (DB) updating. In this case, the green individual is effectively chosen for death because retaining its current behavior is not an option. The final logical case is when, like in PC updating, only a single neighbor is chosen for comparison. This time, however, the focal individual must imitate this neighbor. This model turns out to be equivalent to DB updating when there is no game (i.e. neutral drift) and is therefore not relevant to studying the effects of selection. Our general result (Equation 1) can also account for many update rules beyond these simple (but important) examples.

DB and IM updating. In contrast to PC updating, DB and IM updating are known to support the evolution of producers of pp-goods on homogeneous graphs provided the benefit-to-cost ratio is sufficiently large (Ohtsuki et al., 2006, Chen, 2013).

Consider, for example, a "rich club" network (Zhou and Mondragon, 2004, Colizza et al., 2006, McAuley et al., 2007, Fotouhi et al., 2018), which is defined by a central clique of m individuals, who are connected to each other as well as to n individuals at the periphery. The peripheral individuals are connected to only those in the central clique. This network provides an abstraction of an oligarchy, which is defined by three primary components: "the elite are tightly interconnected among themselves, forming an 'inner circle'; the masses are organized through the intermediation of this inner circle; and the masses are poorly interconnected among themselves" (Ansell et al., 2016). Structures like the rich club arise within corporate hierarchies (Dong et al., 2015), among students in a classroom (based on academic performance) (Vaquero and Cebrian, 2013), and among academic institutions (based on funding) (Ma et al., 2015, Szell and Sinatra, 2015). Surprisingly, $(b/c)^*$ can fall between zero and one for both ff- and pf-goods in such populations (see Figure 4 and Extended Data Figure 5). In particular, producers of ff-goods can evolve even when the total cost of a good exceeds its total benefit.

Note that the critical ratios for pp- and pf-goods in Figure 4 are either both positive or both negative. This example alludes to a more general finding, for all update rules: a graph can support producers of pp-goods for sufficiently large b/c if and only if the same is true for pf-goods. A similar pairing occurs between ff- and fp-goods. In fact, whether there exists some $b, c > 0$ for which selection favors producers depends on only the benefits of the social good; the costs influence the magnitude of $(b/c)^*$ but not its sign (see Methods and SI for details). Statistically speaking, we find that there are more population structures on which producers of ff-goods can evolve than there are for producers of pp- or pf-goods (see Extended Data Figures 6–7). We also observe critical ratios of strictly less than one for ff- and pf-goods on other kinds of graphs, including random graphs (Extended Data Figure 8) and other rich club structures (Jiang and Zhou, 2008) like dense clusters of stars (Extended Data Figure 9). Further comparisons of critical ratios for different kinds of social goods are shown in Extended Data Figure 10 (small graphs) and Extended Data Figure 11 (division of a group into factions).
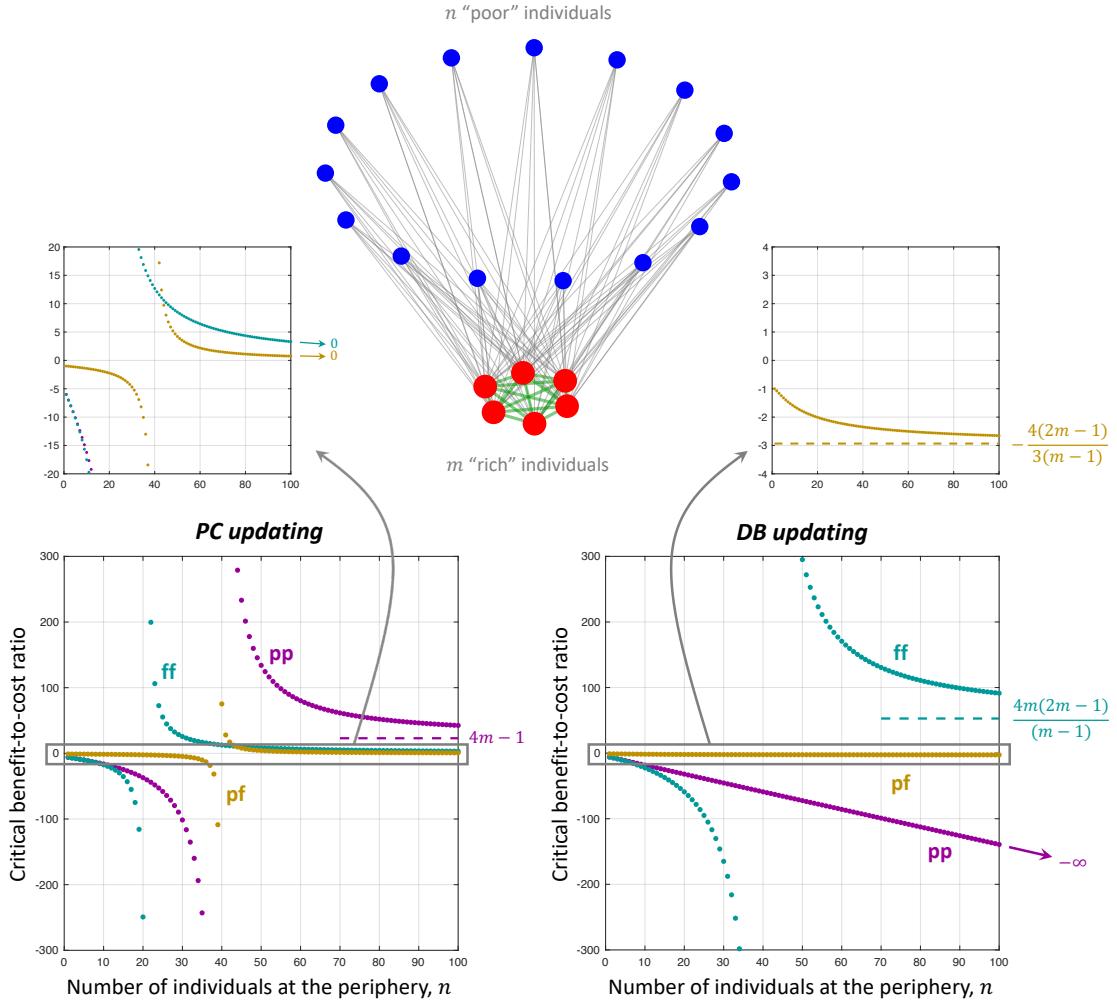
Figure 4: Heterogeneous graphs allow for efficient evolution of prosocial behavior. The "rich club" network consists of a central clique of m well-connected individuals surrounded by n individuals at the periphery Fotouhi et al. (2018). Each of the m "rich" individuals is connected to every other member of the population, but the n "poor" individuals are connected to only the central clique. For $m = 6$, we illustrate the effects of increasing n under PC and DB updating. Under PC updating, the critical ratio for pp-goods approaches $4m - 1$ while those of ff- and pf-goods, remarkably, approach 0. Therefore, producers of ff-goods can evolve even if the total costs exceed the total benefits. For DB updating with $m > 1$, only ff-goods have a positive critical ratio when n is large. This ratio is negative for both pp- and pf-goods, which means that selection can favor "spiteful" pp- and pf-goods provided these goods are sufficiently harmful to others (relative to their cost). For an ff-good with benefit b and cost c, as n grows large the all-producer state results in a payoff that approaches ∞ to each of the m individuals and a payoff that approaches $-c$ to each of the n individuals. Thus, the rich club can hold more than 100% of the total wealth. We give explicit formulas for $(b/c)^*$ for any m and n in the SI.

9
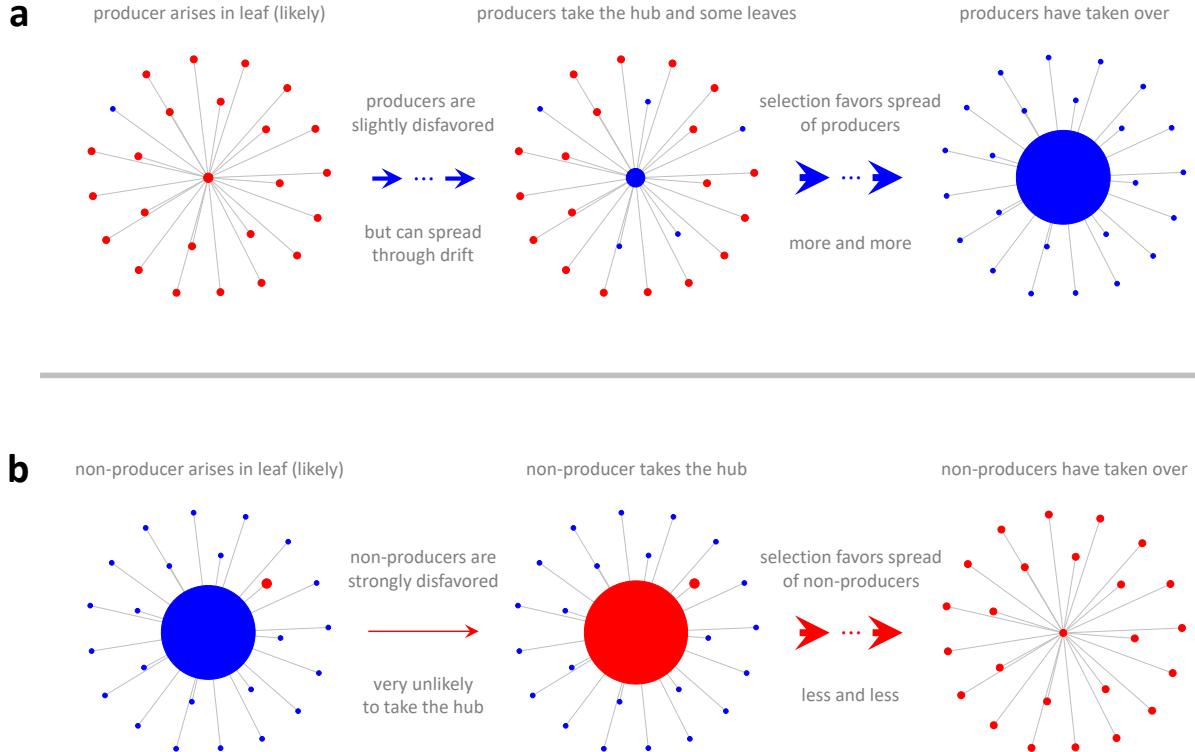
Figure 5: Evolution of producers of ff-goods with $0 < b \leqslant c$ **on the star (PC updating).** The star may be viewed as a special case of the rich club, in which there is just a single "rich" individual ($m = 1$). **a**, invasion and fixation of a mutant producer arising in a leaf under PC updating. This producer has a payoff of $-c$, and the non-producer at the hub gets b. Through drift, this producer can take the hub and propagate a small portion of producers to the leaves. Once there are $k > c/b + 1/(N-1)$ producers at the periphery, a central producer's payoff exceeds that of everyone else in the population and selection favors the further spread of producers. **b**, invasion and fixation of a mutant non-producer arising in a leaf. As soon as a non-producer captures the hub, selection favors the proliferation of non-producers. However, when there is just a single non-producer in the population, a producer at the hub has a much greater payoff than everyone else in the population (even when $0 < b \leqslant c$). Thus, relative to the initial invasion of a producer in **a**, selection acts much more strongly against the initial invasion of a non-producer in **b**. For any fixed $b, c > 0$, these effects become strong enough as N grows that we find $\rho_A > \rho_B$.

10
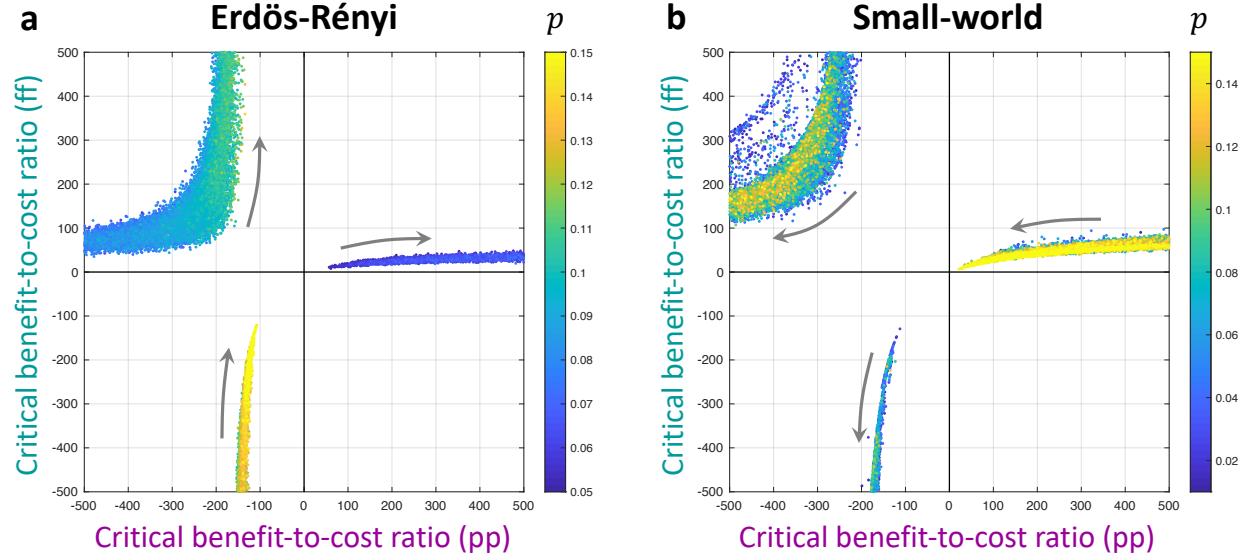
Figure 6: Evolution of producers can be possible for ff-goods but not for pp-goods. a, PC updating on Erdös-Rényi graphs of size $N = 100$ for various edge-inclusion probabilities, p. If p is sufficiently small, the critical benefit-to-cost ratio is positive for both pp- and ff-goods, but for slightly larger p values this ratio can be positive for ff-goods and negative for pp-goods. In the latter case, producers cannot evolve under any b/c ratio for pp-goods, but they can evolve for ff-goods as long as b/c is sufficiently large. **b**, PC updating on small-world networks with different rewiring probabilities, p. Again, there are many examples for which the critical benefit-to-cost ratio is positive for ff-goods but negative for pp-goods.

Even when producers improve the total wealth of a population, their evolution can leave the poorest individuals worse off. In Figure 12, we illustrate this phenomenon using IM updating on an empirical coauthorship network of 379 scientists (Newman, 2006). Producers of ff-goods can evolve only when $b/c \gtrapprox 2.5703$, in which case the total wealth of the population increases. But the proliferation of producers on this graph can leave nearly 20% of the population with negative payoffs, meaning these individuals would prefer the all-B state (non-producers) to the all-A state (producers). The benefit-to-cost ratio must be at least 34 before everyone is better off in all-A. Therefore, wealth-producing goods are not necessarily optimal for everyone in the population (see Extended Data Figure 13 for a summary of possible evolutionary outcomes).

2.2 Asymmetric games

Although our focus so far has been on simple kinds of social goods, our model covers much more complicated asymmetric games. In particular, B_{ij} and C_{ij} can each be any number, and these values need not come from a social good with certain properties shared among all producers. For example, the cost of scientific collaboration, in terms of effort and time, can be less for those a supervisory role than it is for more junior authors. The benefits to the authors might be comparable across roles, e.g. in terms of recognition. Of course, the nature of this difference (both in terms of costs and benefits) is highly dependent on the discipline, with some disciplines being more egalitarian than others (Laurance, 2006, Venkatraman, 2010, Bošnjak and Marušić, 2012, McNutt et al., 2018).

In addition to benefits and costs, the nature of the social goods themselves might vary from
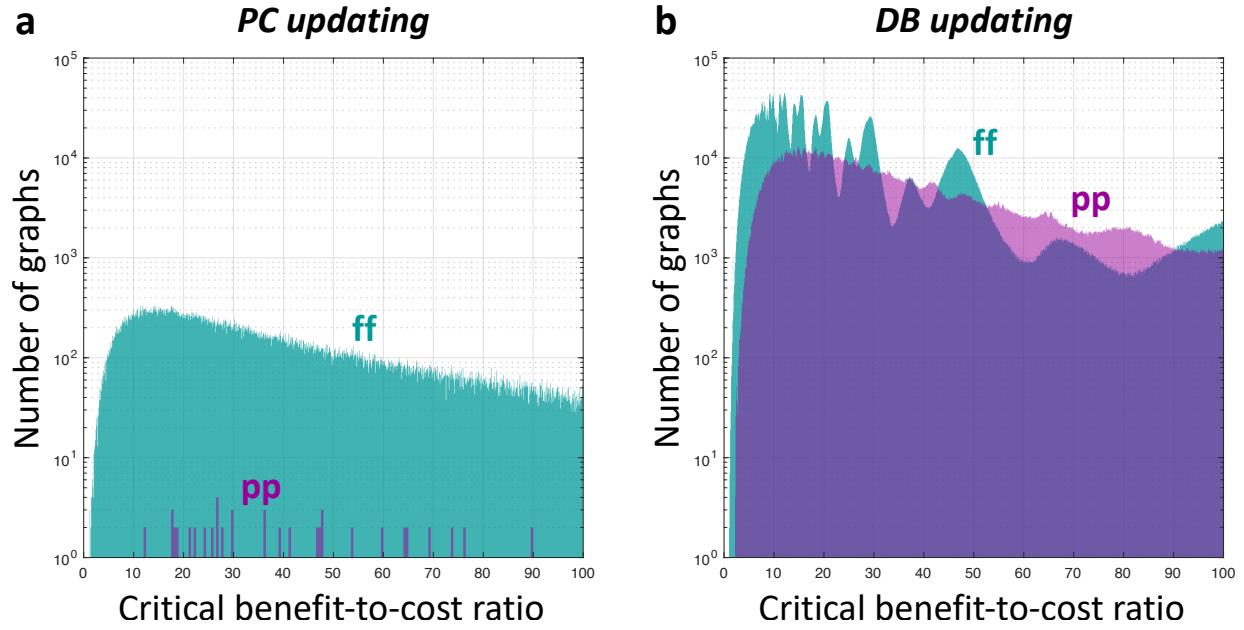
Figure 7: Distributions of critical ratios on small graphs. There are 11,989,763 undirected, unweighted graphs of size at most $N = 10$. Of those that can support the evolution of prosocial behaviors, the critical benefit-to-cost ratios are given for PC updating, **a**, and DB updating, **b**.
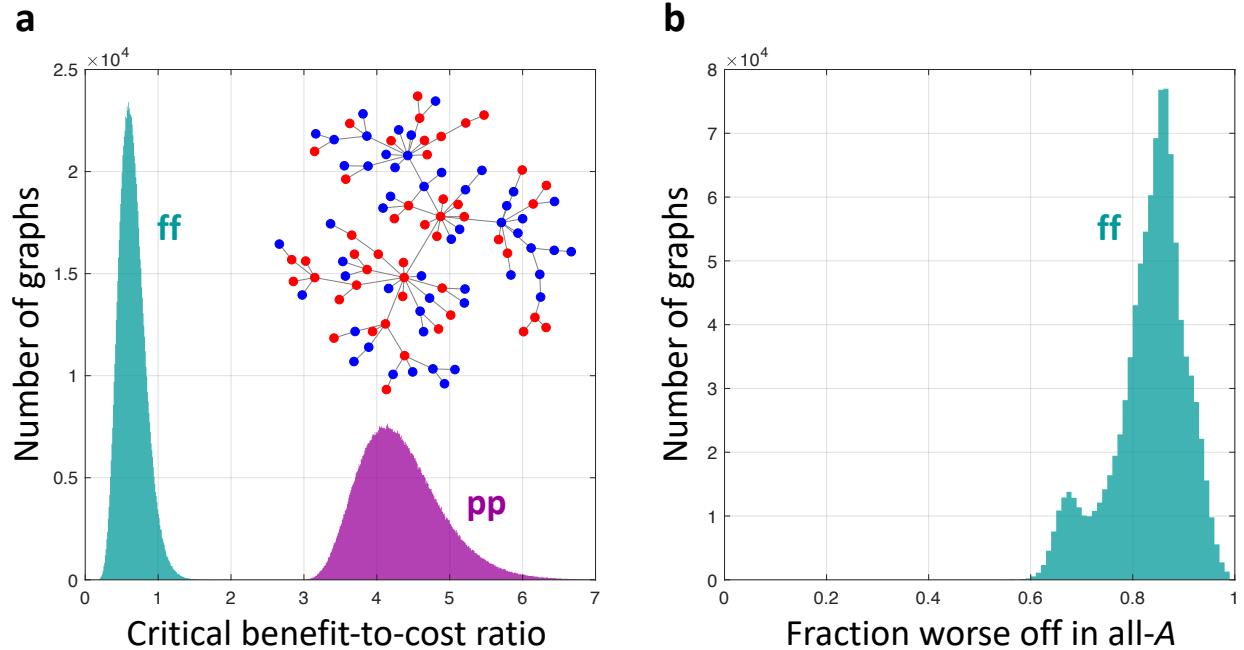
Figure 8: Heterogeneous graphs allow efficient evolution of prosocial behavior. a, The distribution of the critical benefit-to-cost ratio under PC updating for 10^6 preferential-attachment graphs of size $N = 100$ (see Methods). For ff-goods, these structures often have critical benefit-to-cost ratios that are less than one. However, the critical ratio for pp-goods is always greater than one. **b**, When $b/c = (b/c)^*$, these graphs result in a majority (but not all) of the population being worse-off in the all-producer state than in the all-non-producer state.

12
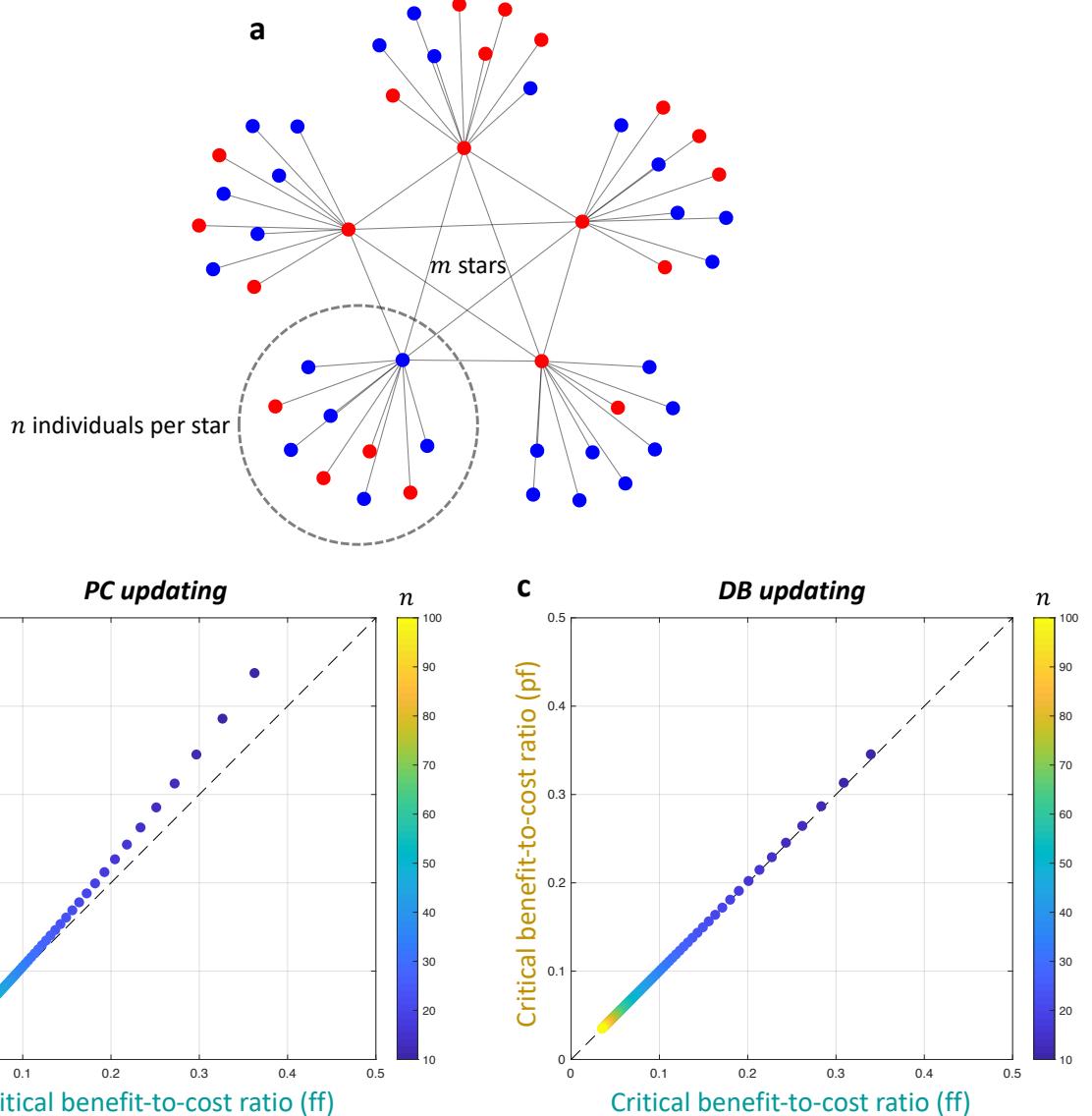
Figure 9: Fixed costs on a dense cluster of stars. Consider a population consisting of m stars, each of size n, connected by a complete graph at their hubs. Provided $m > 1$, this structure results in extremely low critical ratios under both PC and DB updating when n is large. Illustrated here is the case in which $m = 5$. This structure has the interesting property that ff-goods result in lower critical thresholds than pf-goods (both of which are lower than that of pp-goods, which is not depicted here). Qualitatively, the results are similar for both PC and DB updating, with the exception that producers can never be favored by selection on the star ($m = 1$) under DB updating but can be favored under PC updating. We derive explicit formulas for $(b/c)^*$ for any m and n in the SI.

13
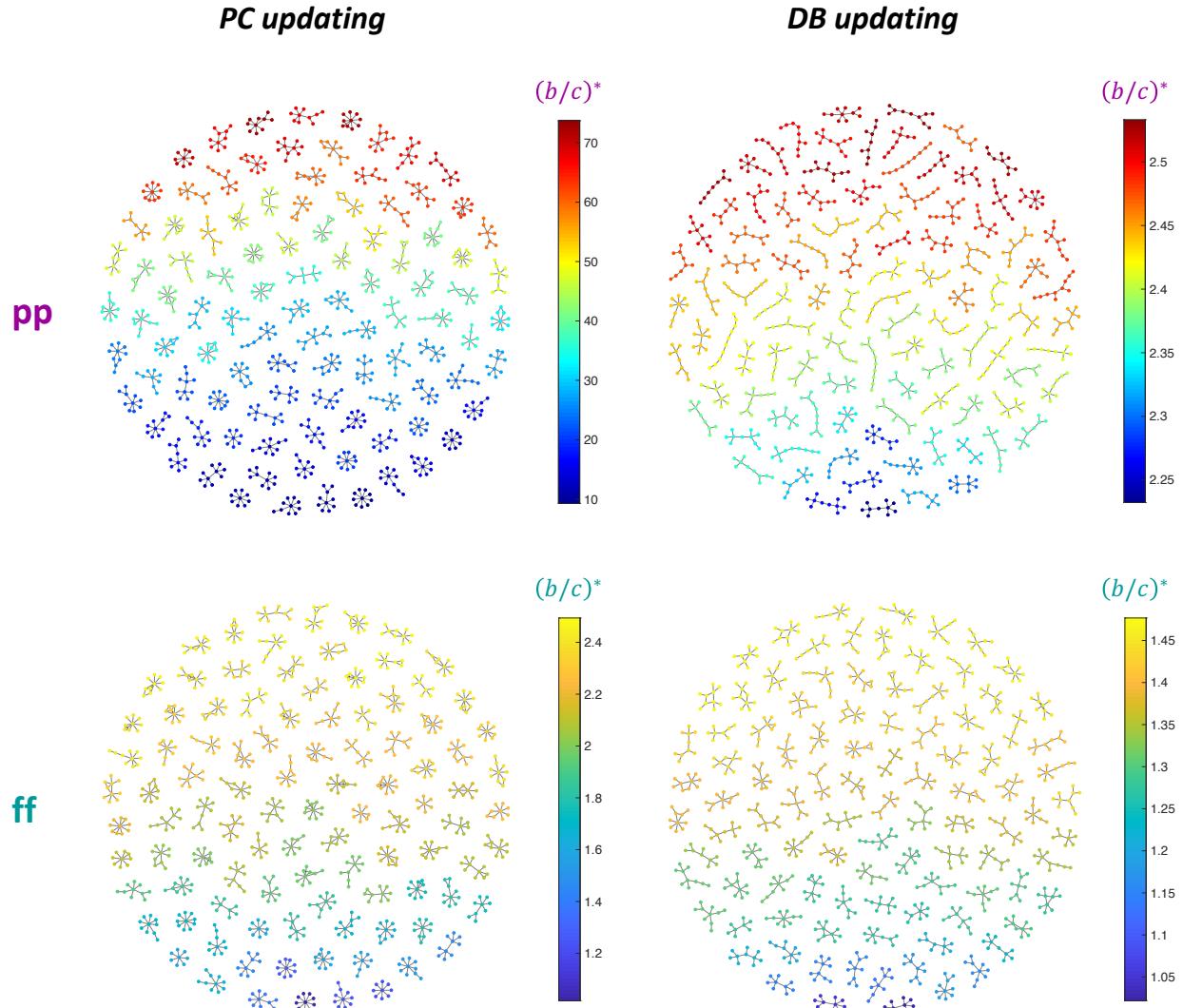
Figure 10: Graphs of size $N \leqslant 10$ **that most easily support prosocial behavior.** For PC and DB updating, we illustrate the 100 graphs with the lowest positive critical ratios for pp- and ff-goods. In each case, the graphs are colored according to their critical ratios. In these examples, ff-goods result in lower critical ratios than pp-goods, and DB updating tends to give lower ratios than PC updating.

|	$(b/c)^*$	$(b/c)_*$			$(b/c)^*$	$(b/c)_*$			$(b/c)^*$	$(b/c)_*$
pp	29.0068	1			126.5400	1			−	1
ff	4.0353	16			4.3247	14			5.0870	15
pf	3.5930	1			19.0310	1			−	1
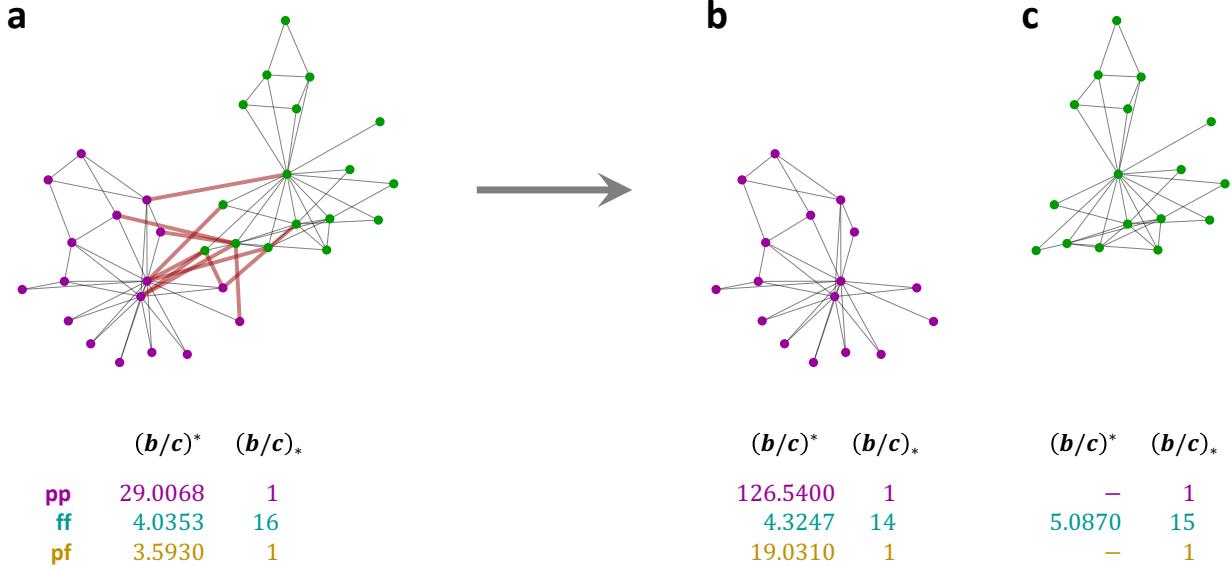
Figure 11: Division of a society into two factions. To illustrate the effects of the division of real-world interaction topologies on evolutionary dynamics, we consider Zachary's karate club (Zachary, 1977), **a**, and the subsequent split of the karate club into two disjoint groups (Girvan and Newman, 2002), **b** and **c**. Under PC updating, producers can evolve on all three networks only in the case of ff-goods. Moreover, even for the two populations (**a** and **b**) in which both ff- and pf-goods can evolve, this split swaps the rankings of the two. In particular, the critical ratio for pf-goods is lower in **a** but that of ff-goods is lower in **b**. The threshold for all individuals to be better off in the all-A state than in the all-B state, $(b/c)_*$, is lowered by the split.

location to location. Returning to the rich club (Figure 4), the central clique might represent a network of content producers (e.g. radio content), while those at the periphery are consumers. If listeners donate money, then this scenario could be reasonably modeled using pf-goods in the center and ff- or pp-goods on the periphery. When those on the periphery produce pp- or ff-goods and those in the central clique produce pf-goods, a sufficient condition for all individuals to be better off is $b > c$ (provided $n \gg m$). It is perhaps unsurprising that in the context of Figure 4, relative to when everyone produces a pp-good (resp. ff-good), it is generally easier (resp. harder) for producers to evolve when those in the center produce pf-goods with the same b and c.

For ff-goods on a rich club, another natural question is whether the individuals in the central clique can scale up their contributions in order to create better outcomes for those at the periphery. In Methods, we show that if wealthy producers scale up their contributions in a way that ensures everyone in the population benefits, it is much harder for producers to evolve at all. Such a population leads to a trade-off between the following two scenarios: *(i)* All individuals in the population produce the same total benefit at the same total cost. Producers easily evolve, leaving well-connected individuals wealthy at the expense of everyone else. *(ii)* Well-connected producers ensure that each neighbor gets back what they contributed. Selection now opposes the spread of producers, leaving the population more often in the asocial (non-producer) state. In the SI, we also discuss how an institution can mitigate the harmful effects of certain prosocial behaviors (like the production of ff-goods), e.g. through a "tax" (Extended Data Figure 14).

Another form of asymmetry arises from stochasticity in the recipient of a donation. Instead of either producing a social good for each neighbor or dividing it up among the neighborhood, a
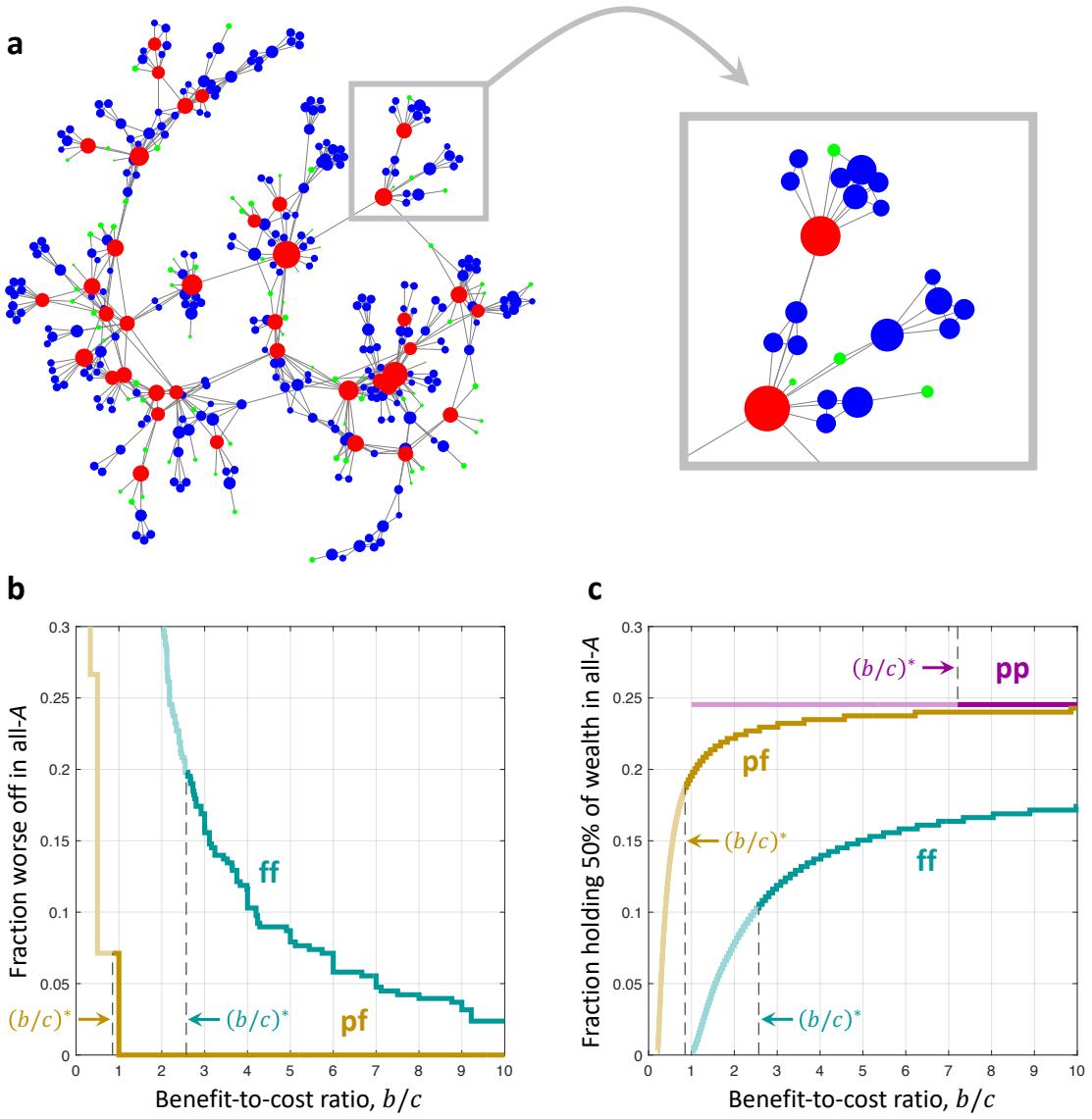
Figure 12: Evolution of prosocial behavior can result in widespread inequality. a, Prosocial inequality arising in a coauthorship network of 379 scientists (Newman, 2006). For IM updating, ff- and pf-goods have lower barriers for selection to favor producers than pp-goods ($(b/c)^* \approx 7.2104$ for pp-goods, $(b/c)^* \approx 2.5703$ for ff-goods, and $(b/c)^* \approx 0.8506$ for pf-goods). In the all-B (non-producer) state, all individuals have the same payoff. But for ff-goods, at the critical ratio where producers evolve, 75 out of 379 individuals (depicted in green on the graph) are worse off in all-A than they would be in all-B. **b**, This fraction decreases as the benefit-to-cost ratio increases. The benefit-to-cost ratio must be at least $(b/c)_* = 34$, which is much larger than $(b/c)^* = 2.5703$, for all-A to be better than all-B for everyone. For pf-goods, there can be some harm to the poorest, but only when $b < c$. In contrast, when producers of pp-goods evolve, all individuals in the population are better off. **c**, In the all-A state, a smaller portion of the population holds at least 50% of the wealth (depicted in red on the graph) for ff- and pf-goods than for pp-goods.

16

Social good	Homogeneous population structures				Heterogeneous population structures			
	Exists?	Can evolve?			Exists?	Can evolve?		
		PC	DB	IM		PC	DB	IM
pp	*w, h*	—	*w*	*w*	*w, h*	*w*	*w*	*w*
ff	*w, h*	—	*w*	*w*	*w, h, wh*	*w, h, wh*	*w, h, wh*	*w, h, wh*
pf	*w, h*	—	*w*	*w*	*w, h, wh*	*w, h, wh*	*w, h, wh*	*w, h, wh*

w = wealth-producing
h = harmful
wh = wealth-producing *and* harmful

Figure 13: Summary of main examples. A good is wealth-producing (*w*) if the total payoff (sum of all benefits minus sum of all costs) is positive when everyone in the population is a producer. It is harmful (*h*) if at least one individual has a negative payoff in the all-producer state. For three kinds of social goods (pp, ff, and pf) and update rules (PC, DB, and IM), this table summarizes when a good can be wealth-producing and/or harmful, as well as when such a good can evolve. Notably, these results are not influenced much by the choice of update rule.
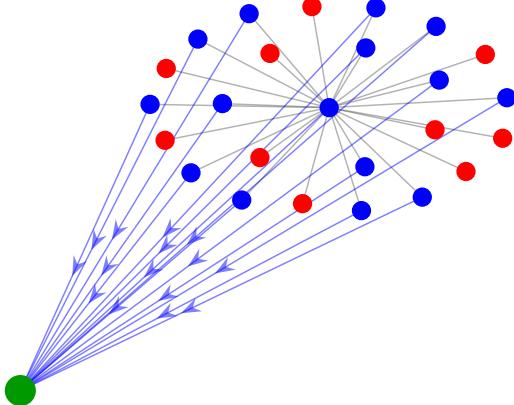
a Producers contribute to a public pool
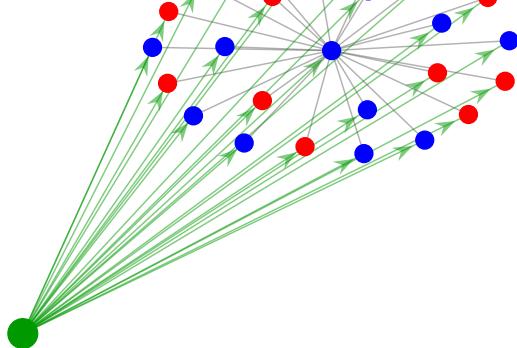
b Pool is divided among all members of the population

c
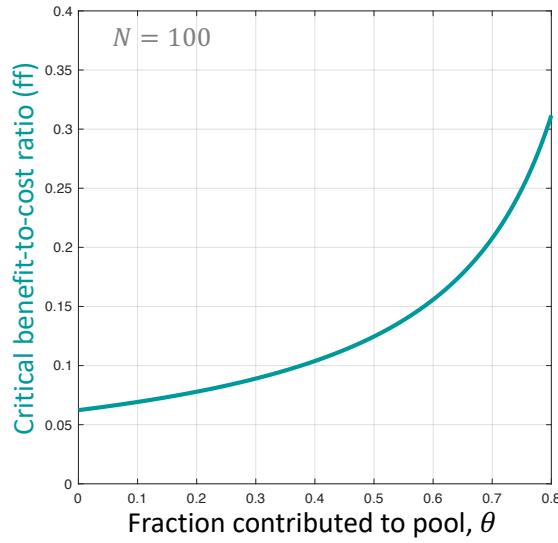
d
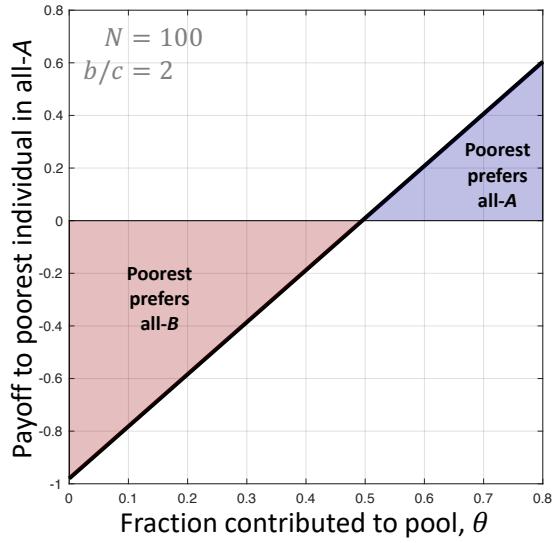
Figure 14: Contributions by producers to a public pool can ameliorate payoff inequality. For ff-goods, suppose that each producer (blue) donates θb to a pool (green) and $(1-\theta) b$ to neighbors, **a**. If the total value of the public pool is divided among all members of the population (green arrows, **b**), then the situation can improve for those who are worst-off in the all-producer state. In particular, such a pool can result in a positive payoff to everyone in the population provided the contribution, quantified by θ, is sufficiently large. The trade-off is that this pool also increases the critical benefit-to-cost ratio required for producers to evolve by a multiplicative factor of $1/(1-\theta)$ (see SI), illustrated in **c** on a star of size $N = 100$ under PC updating. For this population structure, **d** depicts the payoff of the poorest individual ("leaf" player, at the periphery of the star) in the prosocial (all-A) state as a function of the fraction contributed to the pool, θ, when $b = 2$ and $c = 1$. This payoff is negative when $\theta \lesssim 1/2$, which means that 99% of the population is better off in the asocial (all-B) state. However, when $\theta \gtrsim 1/2$, all individuals are better off when producers proliferate.

18

a

b

Fixed benefit, fixed cost (ff)

Concentrated benefit, fixed cost (cf)
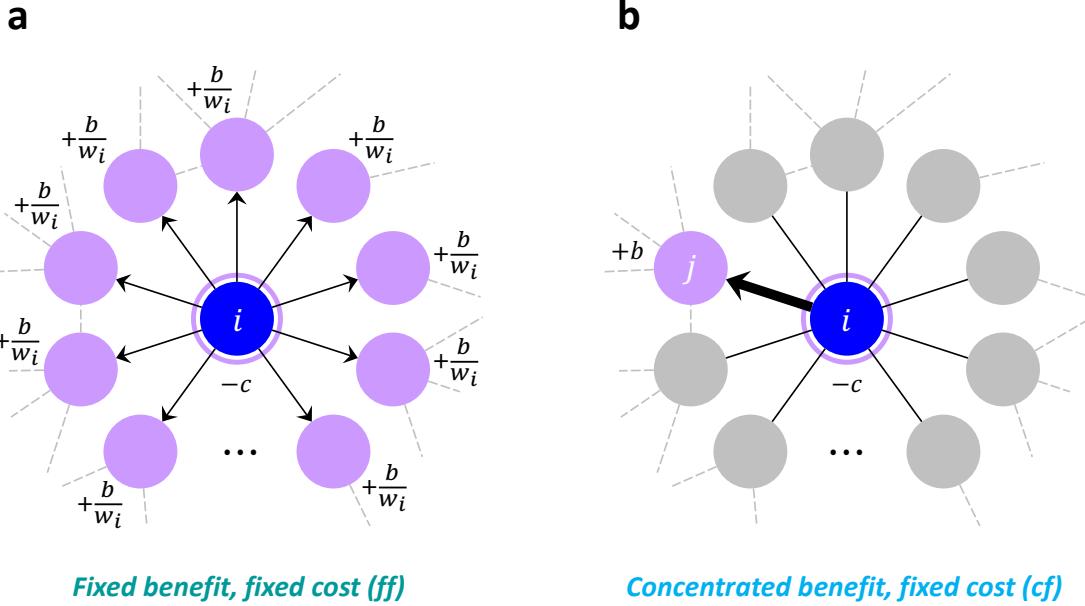
Figure 15: Diffuse versus concentrated benefits. When a producer, i (center), has a rival good of total benefit b and cost c available to donate, two natural ways to distribute this good are **a** to all w_i neighbors, divided evenly, and **b** to a single neighbor, j, chosen at random. The former describes the scheme of ff-goods. The latter represents a stochastic payoff scheme in which one lucky neighbor benefits from the good in its entirety. Remarkably, both methods result in identical critical benefit-to-cost ratios in the limit of weak selection. Therefore, the surprising results reported for ff-goods, such as the existence of critical ratios strictly less than one, also hold for this scheme. More generally, we show in the SI that any stochastic payoff scheme can be replaced by an "equivalent" deterministic scheme.

producer might choose a single random neighbor as the recipient of the good in its entirety (Figure 15). Such a payoff scheme could be driven by indifference, meaning a donor does not care who receives the benefit, or by a mechanism external to the donor. For example, if an individual receives a request to participate in double-blind peer review, then this individual's donation, which is derived from their referee report, is conferred upon recipient(s) who are not directly chosen by the reviewer (the journal chooses). In the SI, we show that, under weak selection, randomly choosing a recipient is equivalent to dividing one's contribution among all possible recipients. This is because, when selection is weak, the conditions for success depend only on expected, rather than actual, payoffs. In particular, from the perspective of evolutionary dynamics, the stochastic donation scheme of Figure 15**b** is equivalent to that of ff-goods, Figure 15**a**, and producers can evolve even when $b < c$.

2.3 Reciprocity

The behavioral types considered so far are quite simple: produce (A) or do not produce (B), unconditionally. When individuals have more than one chance to interact prior to an update to the population, more complex behavioral strategies can emerge. In the iterated prisoner's dilemma, an individual can punish past acts of defection and reward past acts of cooperation, and this mechanism of "direct reciprocity" is well-known for its ability to facilitate the emergence of cooperation (Trivers, 1971, Nowak, 2006, Press and Dyson, 2012, Stewart and Plotkin, 2013, 2016, Hilbe et al., 2018).

There are many different ways to model reciprocity of social good production in heterogeneous populations. As a starting point, consider the strategy "tit-for-tat" (TFT), which cooperates (donates) in the first round and then subsequently copies what the opponent did in the previous round (Axelrod, 1984). Let B_{ij} and C_{ij} be the benefit and cost of i donating to j. In our model, an individual using TFT gives B_{ij} to every j (at a cost of C_{ij}) in the first round. In subsequent rounds, i donates B_{ij} to j (still at cost C_{ij}) if and only if j donated to i in the previous round; otherwise, i gives nothing to j and pays no cost associated to j. Other individuals have no effect on i's choice of whether to donate to j.

For pp-goods, this model gives the classical interpretation of TFT in the prisoner's dilemma, only now the two-player interactions are the pairwise encounters on a graph. For ff-goods, this model may be understood as follows. In the first round, each TFT player produces a good at cost c and divides the benefit, b, among all neighbors. Subsequently, a TFT-player looks around and counts how many neighbors produced a good in the previous round. If a fraction, x, of one's neighbors produced a good, then in the subsequent round this TFT player produces a good of benefit xb and cost xc (i.e. a fraction x of the original good). This benefit is divided among only those neighbors who produced a good in the last round. We avoid pf-goods here because the interpretation of "reciprocity" is much more nuanced for non-excludable goods. The benefits of a clean environment, for instance, normally cannot be denied to an individual.

We consider competition between the strategies ALLD and TFT, where ALLD (which stands for "always defect") is an unconditional non-producer. Since ALLD never produces a good, this iterated game exhibits quite straightforward behavior: In the first round, TFT is a producer and ALLD is a non-producer. In all subsequent rounds, TFT produces for, and donates to, only the other TFT players in their neighborhood. ALLD players get only the benefits they receive in the first round. In order to distinguish between present and future payoff streams, we use a discounting factor, $\lambda \in [0,1]$, which can be interpreted the probability of another encounter before the game ends. When $\lambda = 0$, we recover the original model of producers versus non-producers, with no reciprocity. When $\lambda = 1$, the time horizon of the game is infinite ("undiscounted"). For all values of λ strictly between 0 and 1, the game is finite with $1/(1-\lambda)$ rounds, on average.

For any $\lambda \in [0,1]$, the selection condition for TFT to be favored over ALLD is

$$\sum_{i,j,k,\ell=1}^{N} \pi_i m_k^{ji} \left(-\left(x_{jk} + \lambda x_{j\ell} \right) C_{k\ell} + \left(x_{j\ell} + \lambda x_{jk} \right) B_{\ell k} \right)$$

$$> \sum_{i,j,k,\ell=1}^{N} \pi_i m_k^{ji} \left(-\left(x_{ik} + \lambda x_{i\ell} \right) C_{k\ell} + \left(x_{i\ell} + \lambda x_{ik} \right) B_{\ell k} \right). \tag{2}$$

When $\lambda = 0$, Equation 2 reduces to Equation 1. We derive this result in the SI and provide simplified formulas for PC, DB, and IM updating in Methods. Figure 16 illustrates the effects of increasing the time horizon on the critical ratio for TFT to evolve. In each case, reciprocity lowers the threshold for the evolution of producers. We observe, however, that payoffs in the all-TFT state are the same as those in the all-producer state. Therefore, the potential for wealth-reducing and/or harmful prosociality is not eliminated by reciprocity. On the contrary, reciprocity can enable such outcomes to arise under an expanded range of conditions. Overall, reciprocity typically facilitates the evolution of prosocial behaviors, which may be either helpful or harmful
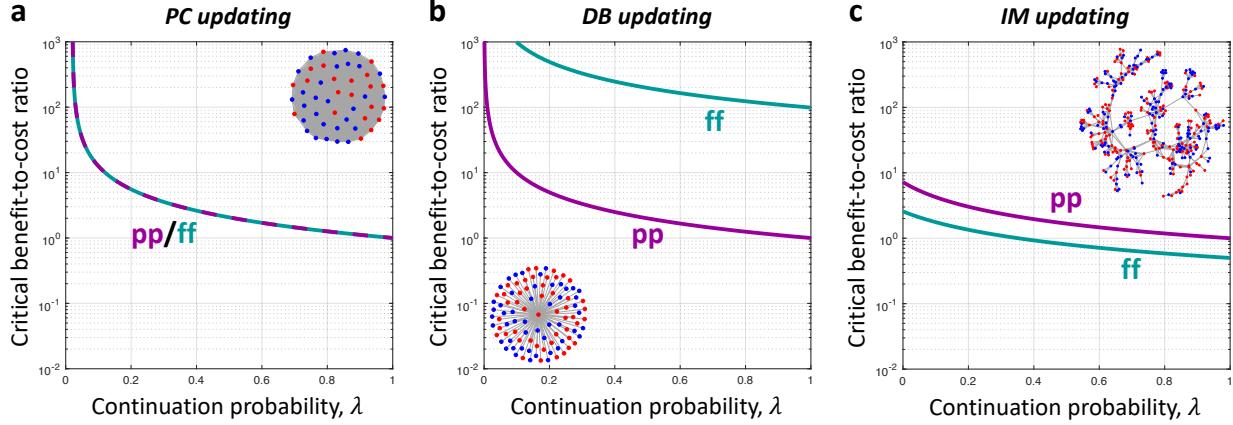
Figure 16: Effects of reciprocity on the selection of prosocial behaviors. Sufficiently large continuation probabilities can facilitate the evolution of TFT by reducing the critical benefit-to-cost ratio. **a,** Whereas producers can never evolve on a homogeneous graph under PC updating when there is just one interaction ($\lambda = 0$), reciprocity can favor producers (TFT) even on well-mixed structures. **b,** The star never supports producers under DB updating, but for non-zero continuation probabilities, λ, TFT can be favored as long as b/c is sufficiently large. **c,** On the empirical coauthorship network considered in Figure 12, increasing λ further promotes the spread of prosocial traits with low benefit-to-cost ratios under IM updating. In particular, since payoffs in the all-TFT state are identical to those of the all-producer state, reciprocity for ff-goods can cause worse outcomes for the poorest in the population.

to the population at large and/or to the least well-off. For example, on the coauthorship network of Figure 12, which corresponds to $\lambda = 0$, ff-goods must be wealth-producing for producers to evolve. But reciprocity, in the form of sufficiently large $\lambda > 0$, can support the evolution of producers even when the underlying social good is wealth-decreasing ($b < c$).

There is one particular condition under which reciprocity cannot be harmful for a population. We say that a prosocial behavior is "pairwise mutually beneficial" (PMB) if $B_{ji} \geqslant C_{ij}$ for all i and j, with strict inequality for at least one pair (i, j). In other words, a behavior is PMB if and only if for each pair of individuals expressing this behavior, both partners receive at least as much as they pay (for that particular interaction). PMB behaviors cannot be wealth-decreasing, nor can they lead to negative payoffs in the all-producer state. The simplest example of a PMB behavior is the production of pp-goods when $b > c$. However, whereas the production of pp-goods when $b > c$ is PMB on any graph, in general whether a behavior is PMB depends on both the good and the population structure. For example, when $b > c$, ff-goods are always PMB on regular graphs, but not necessarily on heterogeneous networks such as the rich club. PMB behaviors are not always favored by selection, as was demonstrated in our baseline model of one-shot interactions. However, we show in the SI that, for any given PMB behavior, TFT is favored for all λ greater than some threshold value $\lambda^* < 1$.

2.4 Stronger selection

The analytical conditions presented so far (e.g. Equation 1) are valid under the assumption that selection is weak. In the case of PC updating, this assumption means that if i has payoff u_i and j, a neighbor of i, has payoff u_j, and if j is chosen as a model individual for comparison to i, then i imitates the behavioral type of j with probability $\left(1 + e^{-\delta(u_j - u_i)}\right)^{-1}$, where δ is positive

but sufficiently small. A natural question arising here is what happens when δ is not necessarily small, which is relevant when individuals are highly inclined to imitate a neighbor with a larger payoff.

Analytical calculations for stronger selection (larger δ) quickly become infeasible for arbitrary heterogeneous graphs Ibsen-Jensen et al. (2015), but remain tractable for structures with a high degree of symmetry, such as the star (Hadjichrysathou et al., 2011). In Figure 17, we consider ff-goods on a star graph of size $N = 500$. We find that stronger selection can promote the evolution of producers significantly, both when the total benefits exceed the total costs and vice versa. We can quantify selection for producers by the fraction of time the population spends in the all-producer state under rare mutation, which is $\rho_A / (\rho_A + \rho_B)$. When $b = 1/10$ and $c = 1$, the star spends approximately 64% of the time in the all-producer state at its peak (Figure 17**c**). When $b = 2$ and $c = 1$, this number is slightly larger at approximately 66% (Figure 17**d**). In both cases, the all-producer state is highly unequal, with a large, positive payoff for the individual at the center of the star and negative payoffs for the $N - 1 = 499$ remaining individuals at the periphery.

3   Discussion

A large and growing body of research (Pinheiro et al., 2014, Peña et al., 2016a, Nowak and May, 1992, Ohtsuki et al., 2006, Taylor et al., 2007, Chen, 2013, Débarre et al., 2014, Santos and Pacheco, 2005, Antal et al., 2006, Gómez-Gardenes et al., 2007, Sood et al., 2008, Cao et al., 2010, Maciejewski et al., 2014, Fan et al., 2017, Allen et al., 2017, Santos et al., 2008, Taylor, 1992, Wilson et al., 1992, Mitteldorf and Wilson, 2000, Irwin and Taylor, 2001, Grafen, 2007, Lion and van Baalen, 2008, Tarnita et al., 2009a, Nowak et al., 2009, Rand et al., 2014, Débarre, 2017, Su et al., 2019a,b, Allen et al., 2019) has shown that spatial structure can promote the evolution of cooperative or prosocial behaviors. Our work, while affirming this principle, reveals it to be more complicated than it may seem. First, we show that the conditions for prosocial behaviors to evolve depend crucially on how the costs and benefits are distributed. This, in turn, depends on whether the goods produced are rival or non-rival, and whether the cost to produce scales with the number of recipients. To link these economic concepts with the evolutionary dynamics literature, we have introduced three natural schemes (ff, pp, and pf) for the production and sharing of goods on networks. Second, and more strikingly, the "prosocial" outcome may not be socially optimal. Selection can favor outcomes in which all individuals contribute to the social good, but some expend more in costs than they receive in benefits, leading to negative payoffs. Additionally, there are structures on which production of ff-goods can evolve when even they are a net detriment to the population ($b < c$).

Qualitatively, these results hold for all update rules considered here; the primary differences among these update rules are in the precise population structures for which the production of goods is favored (see Extended Data Figures and SI). For example, homogeneous graphs can support the evolution of producers under DB and IM updating but not under PC updating. Some heterogeneous structures can favor producers under one update rule but not another. These differences are expected given that different update rules describe different evolutionary dynamics. Our general result (Equation 1) is not restricted to any particular update rule or sharing scheme, and it can be applied to many other populations.
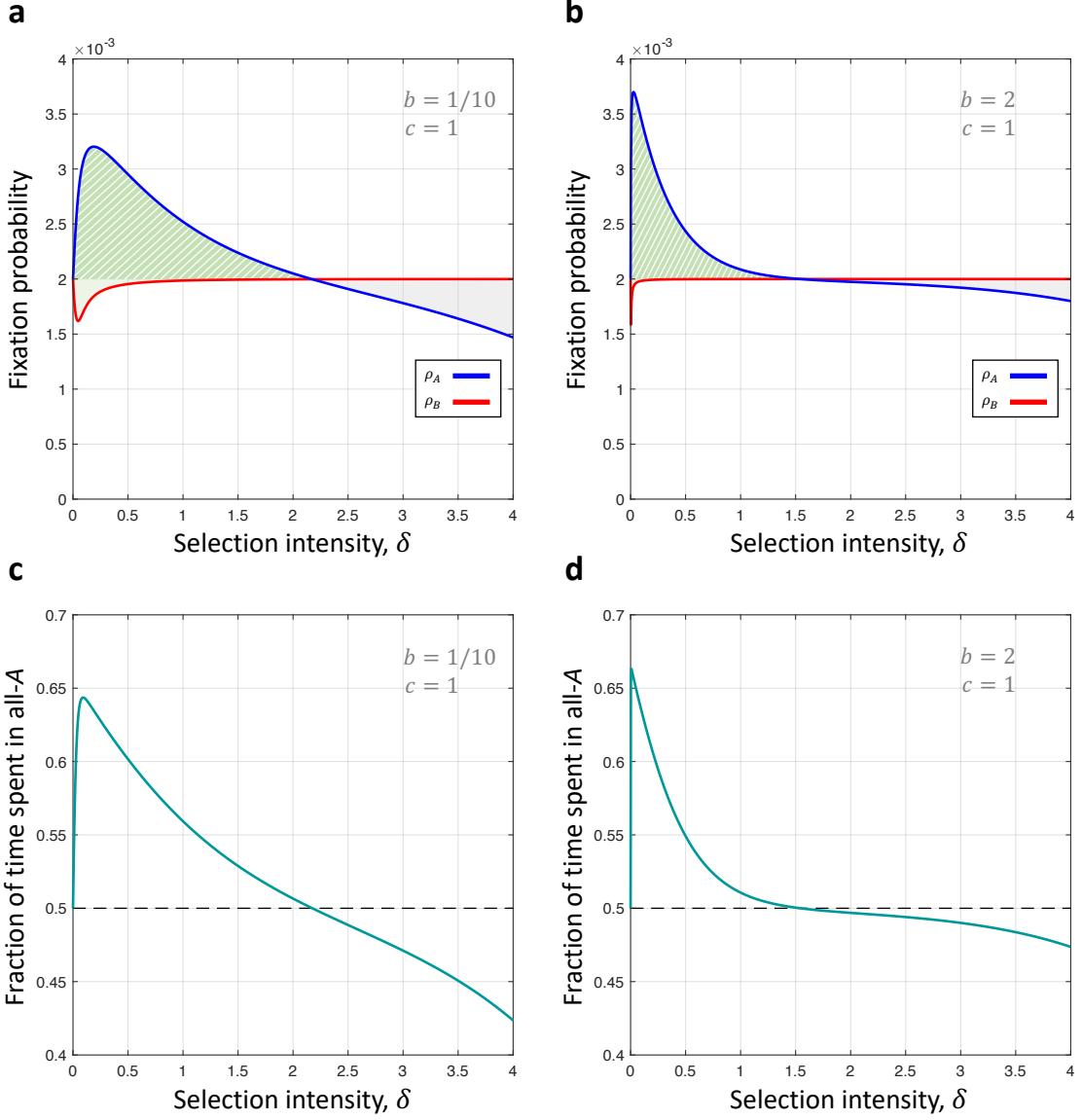
Figure 17: Beyond weak selection on the star graph. Increasing the selection intensity can magnify the wealth inequality due to prosocial behaviors. We illustrate this phenomenon using PC updating on a star graph of size $N = 500$ (i.e. rich club with $m = 1$ and $n = 499$), with ff-goods. Panels **a** and **c** differ from **b** and **d** only in the benefit, b, which is $1/10$ in the former and 2 in the latter. The cost is $c = 1$ in all panels. For both of these cases, our theoretical results imply that increasing the selection strength from $\delta = 0$ favors the evolution of producers. As is illustrated, this effect remains for a significant range of selection intensities, δ. Moreover, at its peak, the amount of time the population spends in the all-producer state is approximately 64% when $b = 1/10$ (**c**) and approximately 66% when $b = 2$ (**d**). When $b = 1/10$, the hub gets a payoff of 48.9, while individuals at the periphery get a payoff of approximately -1. When $b = 2$, the hub gets 997 and individuals at the periphery again get roughly -1. Thus, natural selection of prosocial behaviors can result in extremely bad outcomes for large portions of a population.

One of the main limitations of our study is its restriction to populations of fixed size and structure. Changes to the population size and/or structure as the population evolves could lead to additional interesting behavior with respect to social goods (e.g. population growth rates or the possibility of extinction). Our analysis also focuses on relatively simple social goods (ff, pp, and pf), which represent idealized versions of what might arise in a real population. Congestible goods, for example, are non-rival when there are few consumers but become rival when many stand to benefit from them. A good could also be anti-rival, with the per-capita benefit being an increasing function of the total number of recipients. We see all of these possibilities as promising topics for future studies on the evolutionary dynamics of social goods.

Reciprocity is another mechanism that is known to promote the evolution of cooperation (Trivers, 1971, Nowak, 2006, Press and Dyson, 2012, Stewart and Plotkin, 2013, 2016, Hilbe et al., 2018), which holds in our model as well. In the context of repeated interactions, reciprocal prosocial behaviors can evolve in under a broader range of conditions than their unconditional counterparts. However, this also means that repeated interactions can enable the evolution of harmful and/or wealth-reducing prosociality in conditions where it would not evolve for one-shot interactions. While our model touches upon one aspect of reciprocity in the context of social goods, we also view this topic as an important area for future research.

Our work raises thorny questions about the nature and consequences of cooperative or prosocial behaviors. Under what circumstances should we consider such traits desirable? If they increase a population's total wealth? If they make everyone better off? If they distribute wealth evenly? In general, these conditions are distinct, and each must be considered in the context of the underlying population structure.

In summary, we have shown that heterogeneous population structures act as strong promoters of the evolution of prosocial behaviors. However, the resulting prosocial behaviors can lead to payoff distributions in which a few highly-connected nodes accumulate much of the total wealth (Santos et al., 2008), while poorly-connected nodes end up being harmed. When the population structure is interpreted as describing informal social ties within a group, these examples may be seen as instances of the "tyranny of structurelessness" (Freeman, 2013). In particular, the absence of a formal system of governance can lead to situations in which some (or many) in a group are worse off. While the impact of institutions (Zhang et al., 2013) on evolutionary dynamics is a deep topic, our results provide insight into when an intervention might be necessary. These outcomes call for the design of mechanisms to redistribute wealth in order to maintain a stable society, which engages in and benefits from prosocial behaviors.

4 Methods

4.1 Modeling evolutionary dynamics

We model a general evolutionary process in a population of finite size, N, using the notion of a replacement rule (Allen and Tarnita, 2014, Allen and McAvoy, 2019). If the process is in state $\mathbf{x} \in \{0,1\}^N$ at a given time step, where $x_i = 1$ indicates that i has type A and $x_i = 0$ indicates that i has type B, then a replacement event (R, α) is chosen with probability $p_{(R,\alpha)}(\mathbf{x})$. A replacement event consists of a set of individuals to be replaced, $R \subseteq \{1, \ldots, N\}$, and a parentage map, $\alpha : R \to \{1, \ldots, N\}$, where, for $i \in R$, $\alpha(i) = j$ indicates that i is replaced by

the offspring of j. The distribution $\left\{ p_{(R,\alpha)}(\mathbf{x}) \right\}_{(R,\alpha)}$ defines the replacement rule and specifies the process driving evolution in the population.

For a fixed payoff scheme, the total payoff to i in state $\mathbf{x} \in \{0,1\}^N$ is

$$u_i(\mathbf{x}) = \sum_{j=1}^{N} \left(-x_i C_{ij} + x_j B_{ji} \right), \tag{3}$$

where C_{ij} is the cost i pays to donate to j when i is a producer and B_{ji} is the benefit j provides to i when j is a producer. This payoff is converted to relative fecundity via the formula $F_i(\mathbf{x}) = \exp\{\delta u_i(\mathbf{x})\}$, which is then used to determine the probability of choosing replacement event (R,α) in state $\mathbf{x}$. For example, suppose that $\left(w_{ij} \right)_{i,j=1}^{N}$ is the adjacency matrix for an undirected, unweighted graph. Let $p_{ij} := w_{ij}/w_i$ be the one-step probability of moving from i to j in a random walk on this graph, where $w_i := \sum_{j=1}^{N} w_{ij}$ is the number of individuals neighboring i. Under pairwise-comparison (PC) updating, the probability of choosing (R,α) is

$$p_{(R,\alpha)}(\mathbf{x}) = \begin{cases} \frac{1}{N} p_{i\alpha(i)} \frac{F_{\alpha(i)}(\mathbf{x})}{F_i(\mathbf{x})+F_{\alpha(i)}(\mathbf{x})} & R = \{i\} \text{ for some } i \in \{1,\dots,N\}, \ \alpha(i) \neq i, \\[2mm] \frac{1}{N} \sum_{j=1}^{N} p_{ij} \frac{F_i(\mathbf{x})}{F_i(\mathbf{x})+F_j(\mathbf{x})} & R = \{i\} \text{ for some } i \in \{1,\dots,N\}, \ \alpha(i) = i, \\[2mm] 0 & \text{otherwise.} \end{cases} \tag{4}$$

Similarly, the probability of replacement event (R,α) is

$$p_{(R,\alpha)}(\mathbf{x}) = \begin{cases} \frac{1}{N} \frac{w_{i\alpha(i)} F_{\alpha(i)}(\mathbf{x})}{\sum_{k=1}^{N} w_{ik} F_k(\mathbf{x})} & R = \{i\} \text{ for some } i \in \{1,\dots,N\}, \\[2mm] 0 & \text{otherwise} \end{cases} \tag{5}$$

under DB updating and

$$p_{(R,\alpha)}(\mathbf{x}) = \begin{cases} \frac{1}{N} \frac{w_{i\alpha(i)} F_{\alpha(i)}(\mathbf{x})}{F_i(\mathbf{x})+\sum_{k=1}^{N} w_{ik} F_k(\mathbf{x})} & R = \{i\} \text{ for some } i \in \{1,\dots,N\}, \ \alpha(i) \neq i, \\[2mm] \frac{1}{N} \frac{F_i(\mathbf{x})}{F_i(\mathbf{x})+\sum_{k=1}^{N} w_{ik} F_k(\mathbf{x})} & R = \{i\} \text{ for some } i \in \{1,\dots,N\}, \ \alpha(i) = i, \\[2mm] 0 & \text{otherwise} \end{cases} \tag{6}$$

under IM updating.

4.2 Fixation probabilities, transient dynamics, and the selection condition

We assume that there is at least one individual who can generate a lineage that takes over the population. As a result, the population must eventually end up in one of the two monomorphic states, all-A or all-B. Let ρ_A^i (resp. ρ_B^i) be the probability that a single A (resp. B), placed initially at location i, takes over a population of type B (resp. type A). The mean fixation probabilities for the two types are then $\rho_A = \frac{1}{N} \sum_{i=1}^{N} \rho_A^i$ and $\rho_B = \frac{1}{N} \sum_{i=1}^{N} \rho_B^i$, respectively.

By analyzing the demographic variables (e.g. birth rates and death probabilities) resulting from this process, together with the transient dynamics (i.e. prior to absorption in all-A or all-B), we derive a condition for $\rho_A > \rho_B$ under weak selection ($\delta \ll 1$). Specifically, if *(i)* π_i is the reproductive value of vertex i; *(ii)* x_{ij} is the probability that i and j have the same type in the neutral process (with respect to a distribution described in the SI); and *(iii)* m_k^{ij} is the marginal effect of k's fecundity on i replacing j, then $\rho_A > \rho_B$ under weak selection if and only if Equation 1 holds. Moreover, this condition can be evaluated by solving a linear system of $O(N^2)$ equations. The details of this derivation, including a more formal description of x_{ij}, may be found in the SI.

4.2.1 PC updating

For pairwise-comparison (PC) updating, we have $\rho_A > \rho_B$ under weak selection if and only if

$$\sum_{i=1}^{N} \pi_i \sum_{\ell=1}^{N} \left(-x_{ii}C_{i\ell} + x_{i\ell}B_{\ell i} \right) > \sum_{i,j=1}^{N} \pi_i p_{ij} \sum_{\ell=1}^{N} \left(-x_{ij}C_{j\ell} + x_{i\ell}B_{\ell j} \right). \tag{7}$$

Informally, this condition says that the expected payoff to a producer, i, chosen with probability π_i (which in this case is $w_i / \sum_{j=1}^{N} w_j$), exceeds that of a random neighbor. In the SI, we show that we can evaluate Equation 7 by replacing $1 - x_{ij}$ by τ_{ij}, where $\tau_{ii} = 0$ for every i and

$$\tau_{ij} = 1 + \frac{1}{2} \sum_{k=1}^{N} p_{ik} \tau_{kj} + \frac{1}{2} \sum_{k=1}^{N} p_{jk} \tau_{ik} \tag{8}$$

whenever $i \neq j$. Finally, to see what Equation 7 looks like for a given kind of donation, we just need to give formulas for B_{ij} and C_{ij}. For pp-goods, we have $B_{ij} = w_{ij}b$ and $C_{ij} = w_{ij}c$. For ff-goods, $B_{ij} = w_{ij}b/w_i$ and $C_{ij} = w_{ij}c/w_i$. Finally, for pf-goods, $B_{ij} = w_{ij}b$ and $C_i = w_{ij}c/w_i$. In each case, Equation 7 can be expressed as $\gamma b > \beta c$, where both β and γ are independent of b and c, and $\beta > 0$. With $b, c > 0$, a necessary condition for producers to be favored is thus $\gamma > 0$. Writing $B_{ij} = b\beta_{ij}$ and $C_{ij} = c\gamma_{ij}$, we see that when $\gamma > 0$, the selection condition is

$$\rho_A > \rho_B \iff \frac{b}{c} > \left(\frac{b}{c} \right)^* := \frac{\beta}{\gamma} = \frac{\sum_{i,j=1}^{N} \pi_i p_{ij} \tau_{ij} \sum_{\ell=1}^{N} \gamma_{j\ell}}{\sum_{i,j=1}^{N} \pi_i p_{ij} \sum_{\ell=1}^{N} \left(\tau_{i\ell} - \tau_{j\ell} \right) \beta_{\ell j}}. \tag{9}$$

On a homogeneous graph, one can show that $\gamma < 0$ (see SI), which means that such a structure cannot support the evolution of producers under PC updating. In a prior study, it was shown that on large homogeneous graphs, the evolution of cooperation was possible under PC

updating only in the presence of game transitions (Su et al., 2019c) (in which case the interaction was essentially a coordination game). In contrast, on any homogeneous graph without game transitions, producers (including cooperators) cannot be favored under PC updating. However, as shown in the text, there are many examples of heterogeneous population structures with $\gamma > 0$.

4.2.2 DB updating

For death-birth (DB) updating, we have $\rho_A > \rho_B$ under weak selection if and only if

$$\sum_{i=1}^{N} \pi_i \sum_{\ell=1}^{N} \left(-x_{ii}C_{i\ell} + x_{i\ell}B_{\ell i}\right) > \sum_{i,j=1}^{N} \pi_i p_{ij}^{(2)} \sum_{\ell=1}^{N} \left(-x_{ij}C_{j\ell} + x_{i\ell}B_{\ell j}\right). \tag{10}$$

The interpretation of this condition is analogous to that of the condition for PC updating, except here the comparison is between a producer and a two-step (rather than one-step) neighbor.

Using the same values of π_i and τ_{ij} described above for PC updating, this condition becomes

$$\rho_A > \rho_B \iff \frac{b}{c} > \left(\frac{b}{c}\right)^* := \frac{\beta}{\gamma} = \frac{\sum_{i,j=1}^{N} \pi_i p_{ij}^{(2)} \tau_{ij} \sum_{\ell=1}^{N} \gamma_{j\ell}}{\sum_{i,j=1}^{N} \pi_i p_{ij}^{(2)} \sum_{\ell=1}^{N} \left(\tau_{i\ell} - \tau_{j\ell}\right) \beta_{\ell j}}. \tag{11}$$

4.2.3 IM updating

For both PC and DB updating, the reproductive value, π_i, random walk step-probability, p_{ij}, and recurrence relation for τ_{ij} (which is a proxy for x_{ij}) were the same in both processes. For imitation (IM) updating, these quantities need to be modified slightly. Let $\widetilde{w}$ be matrix obtained from w by adding a self-loop to each vertex, i.e. $\widetilde{w}_{ij} = w_{ij}$ for $i \neq j$ and $\widetilde{w}_{ii} = 1$ for all $i = 1, \ldots, N$. Let $\widetilde{p}_{ij} := \widetilde{w}_{ij}/\widetilde{w}_i$ be the probability of transitioning from i to j in one step of a random walk on this modified graph. The reproductive value of i under IM updating is $\pi_i = \widetilde{w}_i / \sum_{j=1}^{N} \widetilde{w}_j$, where $\widetilde{w}_i = 1 + w_i$. Under weak selection, we have $\rho_A > \rho_B$ if and only if

$$\sum_{i=1}^{N} \pi_i \sum_{\ell=1}^{N} \left(-x_{ii}C_{i\ell} + x_{i\ell}B_{\ell i}\right) > \sum_{i,j=1}^{N} \pi_i \widetilde{p}_{ij}^{(2)} \sum_{\ell=1}^{N} \left(-x_{ij}C_{j\ell} + x_{i\ell}B_{\ell j}\right). \tag{12}$$

The terms x_{ij} are again different here than they were for PC and DB updating. However, interpreting x_{ij} in the same way as before (as the probability that i and j have the same type with respect to a certain distribution in the neutral process), Equation 12 has a natural interpretation as a comparison between the expected payoff to a producer (left-hand side) and the expected payoff to a two-step neighbor of a producer (right-hand side). To evaluate this condition, we can replace $1 - x_{ij}$ by τ_{ij}, where $\tau_{ii} = 0$ for every i and, whenever $i \neq j$, τ_{ij} satisfies the recurrence

$$\tau_{ij} = 1 + \frac{1}{2} \sum_{k=1}^{N} \widetilde{p}_{ik} \tau_{kj} + \frac{1}{2} \sum_{k=1}^{N} \widetilde{p}_{jk} \tau_{ik}. \tag{13}$$

For $B_{ij} = b\beta_{ij}$ and $C_{ij} = c\gamma_{ij}$, the selection condition when $\gamma > 0$ is

$$\rho_A > \rho_B \iff \frac{b}{c} > \left(\frac{b}{c}\right)^* := \frac{\beta}{\gamma} = \frac{\sum_{i,j=1}^{N} \pi_i \widehat{p}_{ij}^{(2)} \tau_{ij} \sum_{\ell=1}^{N} \gamma_{j\ell}}{\sum_{i,j=1}^{N} \pi_i \widehat{p}_{ij}^{(2)} \sum_{\ell=1}^{N} \left(\tau_{i\ell} - \tau_{j\ell}\right) \beta_{\ell j}}. \tag{14}$$

4.3 Asymmetric games

Here, we consider the effects of asymmetric payoffs on rich club graphs with m central nodes and n peripheral nodes. We are especially interested in the case in which m is fixed and n grows.

Suppose that a producer on the periphery produces a good of cost c_p and divides the benefit, b_p, among all neighbors. A producer in the center group pays c_r and distributes the benefit, b_r, evenly. We consider the case in which $b_p = b$ and $c_p = c$, and that $b_r = s(m,n) b_p$ and $c_r = s(m,n) c_p$ for some function $s(m,n)$. If everyone on the rich club is a producer (whose state in $\{0,1\}^N$ is denoted $\mathbf{A}$), then the payoff to individual i is

$$u_i(\mathbf{A}) = \begin{cases} -s(m,n)c + \frac{m-1}{m+n-1}s(m,n)b + \frac{n}{m}b & i \text{ is in the central rich club,} \\ \\ -c + \frac{m}{m+n-1}s(m,n)b & i \text{ is on the periphery.} \end{cases} \tag{15}$$

For a peripheral individual's payoff to remain non-negative as $n \to \infty$, $s(m,n)$ must grow at least linearly in n. At the same time, if $\lim_{n\to\infty} \frac{1}{n} s(m,n) = \infty$, then the payoff to an individual in the central rich club will eventually become negative as $n \to \infty$. Therefore, we consider $s(m,n) = k_1(m) n + k_0(m)$ for some functions k_0 and k_1 of m. Letting $n \to \infty$ gives

$$\lim_{n\to\infty} u_i(\mathbf{A}) = \begin{cases} -\infty & i \text{ is in the central rich club, } \frac{b}{c} < mk_1(m), \\ \\ 0 & i \text{ is in the central rich club, } \frac{b}{c} = mk_1(m), \\ \\ +\infty & i \text{ is in the central rich club, } \frac{b}{c} > mk_1(m), \\ \\ -c + mk_1(m)b & i \text{ is on the periphery.} \end{cases} \tag{16}$$

For the first payoff in Equation 15 to stay non-negative as $n \to \infty$, we require $b/c \geqslant mk_1(m)$. But we also want the second payoff in Equation 15 to be non-negative, i.e. $b/c \geqslant 1/(mk_1(m))$. Thus, we seek k_1 with

$$\frac{c}{b} \leqslant mk_1(m) \leqslant \frac{b}{c}. \tag{17}$$

Such a k_1 exists if and only if $b \geqslant c$, in which case we can set $k_0(m) = 0$ and $k_1(m) = 1/m$. For large n, it follows that each producer in the central clique gives an average of b/m to each neighbor at a cost of c/m. The per-neighbor benefit and cost values are the same (b/m and c/m, respectively) for each peripheral individual as well, which effectively transforms the payoff

28

structure into that of a pp-good with benefit-to-cost ratio $(b/m) / (c/m) = b/c$, and we already know that it is much more difficult for producers of such a good to evolve (if they can at all).

4.4 Reciprocity

For the model of reciprocity defined in the text, we let A and B denote the strategies TFT and ALLD, respectively. Let u_i^t be the payoff to player i in the tth round of the game, i.e.

$$u_i^t(\mathbf{x}) = \begin{cases} \sum_{j=1}^N \left(-x_i C_{ij} + x_j B_{ji}\right) & t = 1, \\ \\ \sum_{j=1}^N x_i x_j \left(-C_{ij} + B_{ji}\right) & t > 1. \end{cases} \tag{18}$$

If the discounting factor (continuation probability) is $\lambda \in [0,1)$, then the overall payoff to i is

$$\begin{aligned} u_i(\mathbf{x}) &= (1-\lambda) \left(\sum_{j=1}^N \left(-x_i C_{ij} + x_j B_{ji}\right) + \sum_{t=1}^\infty \lambda^t \sum_{j=1}^N x_i x_j \left(-C_{ij} + B_{ji}\right) \right) \\ &= (1-\lambda) \sum_{j=1}^N \left(-x_i C_{ij} + x_j B_{ji}\right) + \lambda \sum_{j=1}^N x_i x_j \left(-C_{ij} + B_{ji}\right) \\ &= \sum_{j=1}^N \left(-\left(1 - \lambda + \lambda x_j\right) x_i C_{ij} + \left(1 - \lambda + \lambda x_i\right) x_j B_{ji}\right). \end{aligned} \tag{19}$$

Writing down the payoffs when i uses TFT ($x_i = 1$) and ALLD ($x_i = 0$) separately gives

$$u_i(\mathbf{x}) = \begin{cases} \sum_{j=1}^N \left(-C_{ij} + x_j B_{ji}\right) + \lambda \sum_{j=1}^N \left(1 - x_j\right) C_{ij} & x_i = 1, \\ \\ (1-\lambda) \sum_{j=1}^N x_j B_{ji} & x_i = 0. \end{cases} \tag{20}$$

It follows that, in each state $\mathbf{x} \in \{0,1\}^N$, increasing λ does not decrease the payoff to an A and does not increase the payoff to a B. For any reasonable process favoring individuals with higher payoffs (including PC, DB, and IM updating), it follows that ρ_A is monotonically increasing and ρ_B is monotonically decreasing in λ. Furthermore, we note that when $\lambda = 1$, i gets 0 when using ALLD and $\sum_{j=1}^N x_j \left(-C_{ij} + B_{ji}\right)$ when using TFT. If the underlying behavior is pairwise mutually beneficial (PMB), then in every state each ALLD has a payoff of zero and each TFT has a payoff of at least zero. Therefore, no reasonable process favoring traits with higher payoffs should disfavor TFT relative to ALLD when the interactions have an infinite time horizon.

The selection condition for ALLD versus TFT, Equation 2, is derived in the SI. Here, we explore how this condition can be evaluated for the update rules considered in the text.

4.4.1 PC updating

For PC updating, the condition for selection to favor TFT relative to ALLD is

$$\rho_A > \rho_B \iff \sum_{i=1}^{N} \pi_i \sum_{\ell=1}^{N} \left(-\left(x_{ii} + \lambda x_{i\ell} \right) C_{i\ell} + \left(x_{i\ell} + \lambda x_{ii} \right) B_{\ell i} \right)$$
$$> \sum_{i,j=1}^{N} \pi_i p_{ij} \sum_{\ell=1}^{N} \left(-\left(x_{ij} + \lambda x_{i\ell} \right) C_{j\ell} + \left(x_{i\ell} + \lambda x_{ij} \right) B_{\ell j} \right). \quad (21)$$

For social goods satisfying $B_{ij} = b\beta_{ij}$ and $C_{ij} = c\gamma_{ij}$, if $\gamma > 0$ then this condition is

$$\frac{b}{c} > \left(\frac{b}{c} \right)_{\lambda}^{*} = \frac{\sum_{i,j=1}^{N} \pi_i p_{ij} \sum_{\ell=1}^{N} \left(\tau_{ij} + \lambda \tau_{i\ell} - \lambda \tau_{j\ell} \right) \gamma_{j\ell}}{\sum_{i,j=1}^{N} \pi_i p_{ij} \sum_{\ell=1}^{N} \left(\lambda \tau_{ij} + \tau_{i\ell} - \tau_{j\ell} \right) \beta_{\ell j}}, \quad (22)$$

where π_i and τ_{ij} are the same as they were previously for PC updating with $\lambda = 0$.

4.4.2 DB updating

For DB updating, the condition for selection to favor TFT relative to ALLD is

$$\rho_A > \rho_B \iff \sum_{i=1}^{N} \pi_i \sum_{\ell=1}^{N} \left(-\left(x_{ii} + \lambda x_{i\ell} \right) C_{i\ell} + \left(x_{i\ell} + \lambda x_{ii} \right) B_{\ell i} \right)$$
$$> \sum_{i,j=1}^{N} \pi_i p_{ij}^{(2)} \sum_{\ell=1}^{N} \left(-\left(x_{ij} + \lambda x_{i\ell} \right) C_{j\ell} + \left(x_{i\ell} + \lambda x_{ij} \right) B_{\ell j} \right). \quad (23)$$

For social goods satisfying $B_{ij} = b\beta_{ij}$ and $C_{ij} = c\gamma_{ij}$, if $\gamma > 0$ then this condition is

$$\frac{b}{c} > \left(\frac{b}{c} \right)_{\lambda}^{*} = \frac{\sum_{i,j,\ell=1}^{N} \pi_i p_{ij}^{(2)} \left(\tau_{ij} + \lambda \tau_{i\ell} - \lambda \tau_{j\ell} \right) \gamma_{j\ell}}{\sum_{i,j,\ell=1}^{N} \pi_i p_{ij}^{(2)} \left(\lambda \tau_{ij} + \tau_{i\ell} - \tau_{j\ell} \right) \beta_{\ell j}}, \quad (24)$$

where π_i and τ_{ij} are the same as they were previously for DB updating with $\lambda = 0$.

4.4.3 IM updating

For IM updating, the condition for selection to favor TFT relative to ALLD is

$$\sum_{i,\ell=1}^{N} \pi_i \left(-\left(x_{ii} + \lambda x_{i\ell} \right) C_{i\ell} + \left(x_{i\ell} + \lambda x_{ii} \right) B_{\ell i} \right)$$
$$> \sum_{i,j,\ell=1}^{N} \pi_i \widetilde{p}_{ij}^{(2)} \left(-\left(x_{ij} + \lambda x_{i\ell} \right) C_{j\ell} + \left(x_{i\ell} + \lambda x_{ij} \right) B_{\ell j} \right). \quad (25)$$

For social goods satisfying $B_{ij} = b\beta_{ij}$ and $C_{ij} = c\gamma_{ij}$, if $\gamma > 0$ then this condition is

$$\frac{b}{c} > \left(\frac{b}{c}\right)_{\lambda}^{*} = \frac{\sum_{i,j,\ell=1}^{N} \pi_i \widetilde{p}_{ij}^{(2)} \left(\tau_{ij} + \lambda\tau_{i\ell} - \lambda\tau_{j\ell}\right) \gamma_{j\ell}}{\sum_{i,j,\ell=1}^{N} \pi_i \widetilde{p}_{ij}^{(2)} \left(\lambda\tau_{ij} + \tau_{i\ell} - \tau_{j\ell}\right) \beta_{\ell j}}, \tag{26}$$

where π_i and τ_{ij} are the same as they were previously for IM updating with $\lambda = 0$.

4.5 Numerical examples

Extended Data Figures 8 and 6 depict critical ratios on random graphs. Extended Data Figure 8 was generated using a Barabási-Albert preferential attachment model (Barabási and Albert, 1999); the population begins with $m_0 = 2$ individuals, and each new individual is connected to $m = 1$ existing members of the population according to the standard degree-weighted distribution. Extended Data Figure 6**a** was generated using the $G(N, p)$ Erdös-Rényi model (Bollobás, 2001). Extended Data Figure 6**b** was created using the Watts-Strogatz model (Watts and Strogatz, 1998) starting from a cycle with edge-rewiring probability p.

For the empirical networks considered in the main text and Extended Data Figures (which are available on public databases (Leskovec and Krevl, 2014, Rossi and Ahmed, 2015)), we have used the largest connected component to ensure that fixation is possible in the case of societies with more than one connected component. The sources of all empirical networks used here are provided in the corresponding figure captions.

Acknowledgments

Supporting Information

SI.1 Formally modeling evolutionary dynamics

In this section, we formally establish the conditions for A to be favored over B stated in the main text. The notation and modeling techniques follow the conventions of Allen and McAvoy (2019).

Since we model two types, A and B, in a population of finite size, N, the state of the population is specified by the configuration of A and B. For simplicity, we denote the state of the population by a binary vector, $\mathbf{x} \in \{0,1\}^N$, where $x_i = 1$ means that the individual at location i is has type A, and $x_i = 0$ means that this individual has type B. We denote by $\mathbf{A} := (1,1,\ldots,1)$ and $\mathbf{B} := (0,0,\ldots,0)$ the two monomorphic (or monoallelic) states, respectively.

We describe evolutionary dynamics using the notion of a replacement rule (Allen and Tarnita, 2014, Allen and McAvoy, 2019). A replacement rule is a distribution $\left\{ p_{(R,\alpha)} (\mathbf{x}) \right\}_{(R,\alpha)}$ over pairs (R,α), where R is a subset of $\{1,\ldots,N\}$ and α is a map $R \to \{1,\ldots,N\}$. R is the set of individuals replaced in a given time step, and α is the parentage map, which means that $\alpha(i) = j$ if and only if the offspring of j replaces i. The probability of choosing replacement event (R,α), which is denoted by $p_{(R,\alpha)} (\mathbf{x})$, usually depends on the current population state, $\mathbf{x} \in \{0,1\}^N$, because an individual's type can influence the probability with which they reproduce and/or die.

The notion of a replacement rule is flexible enough to accommodate a wide variety of models of spatially structured populationsAllen and Tarnita (2014), Allen and McAvoy (2019). Although the update rules considered in the main text have at most one individual replaced per time step (in our notation, $p_{(R,\alpha)} (\mathbf{x}) = 0$ whenever $|R| > 1$), the results of this section allow for any fixed or variable number of individuals to be replaced. In Section SI.2 we give specific formulas for $p_{(R,\alpha)} (\mathbf{x})$ for various update rules.

We assume that $\left\{ p_{(R,\alpha)} (\mathbf{x}) \right\}_{(R,\alpha)}$ satisfies the following property (Allen and McAvoy, 2019), which formally states that there is at least one individual that can eventually be an ancestor of the entire population:

Fixation Axiom. There exists $i \in \{1,\ldots,N\}$, an integer $m \geq 1$, and a sequence of replacement events $\{(R_k, \alpha_k)\}_{k=1}^m$ such that

- $p_{(R_k,\alpha_k)} (\mathbf{x}) > 0$ for every $k \in \{1,\ldots,m\}$ and $\mathbf{x} \in \{0,1\}^N$;

- $i \in R_k$ for some $k \in \{1,\ldots,m\}$;

- $\widetilde{\alpha}_1 \circ \widetilde{\alpha}_2 \circ \cdots \circ \widetilde{\alpha}_m (j) = i$ for every $j \in \{1,\ldots,N\}$, where $\widetilde{\alpha} : \{1,\ldots,N\} \to \{1,\ldots,N\}$ denotes the extension of $\alpha : R \to \{1,\ldots,N\}$ to all of $\{1,\ldots,N\}$, i.e.

$$
\widetilde{\alpha}(j) := \begin{cases} \alpha(j) & j \in R, \\ \\ j & j \notin R. \end{cases} \tag{SI.1}
$$

We now describe how a replacement event affects the state of the population. Suppose that the current state is $\mathbf{x} \in \{0,1\}^N$. Following the replacement event (R,α), individual i inherits

32

his or her type from $\widetilde{\alpha}(i)$; thus, the type of i is updated from x_i to $x_{\widetilde{\alpha}(i)}$. We denote by $\mathbf{x}_{\widetilde{\alpha}}$ this resulting state. Given a replacement rule $\left\{ p_{(R,\alpha)}(\mathbf{x}) \right\}_{(R,\alpha)}$, we can then define a Markov chain on $\{0,1\}^N$, with the transition probability between $\mathbf{x}$ and $\mathbf{y}$ in $\{0,1\}^N$ given by

$$P_{\mathbf{x} \to \mathbf{y}} := \sum_{\substack{(R,\alpha) \\ \mathbf{x}_{\widetilde{\alpha}} = \mathbf{y}}} p_{(R,\alpha)}(\mathbf{x}). \tag{SI.2}$$

Since $\mathbf{A}_{\widetilde{\alpha}} = \mathbf{A}$ and $\mathbf{B}_{\widetilde{\alpha}} = \mathbf{B}$ for every replacement event, (R, α), $\mathbf{A}$ and $\mathbf{B}$ are absorbing states for this chain. Moreover, the Fixation Axiom implies that every non-monomorphic state is transient, so the chain will eventually end up in either $\mathbf{A}$ or $\mathbf{B}$ given any initial state. We are particularly interested in rare-mutant states, which for us means that there are $N-1$ of type A and one of type B or $N-1$ of type B and one of type A. Denote by ρ_A^i (resp. ρ_B^i) the probability that $\mathbf{A}$ (resp. $\mathbf{B}$) is reached from the state with just a single A (resp. B) at location i. The mean fixation probabilities of A and B are then $\rho_A := (1/N) \sum_{i=1}^N \rho_A^i$ and $\rho_B := (1/N) \sum_{i=1}^N \rho_B^i$, respectively. This section is dedicated to using a replacement rule to derive a condition for selection to favor type A relative to type B, i.e. $\rho_A > \rho_B$.

SI.1.1 Demographic variables

The marginal probability of transmission from i to j in state $\mathbf{x} \in \{0,1\}^N$ is

$$e_{ij}(\mathbf{x}) := \sum_{\substack{(R,\alpha) \\ \alpha(j)=i}} p_{(R,\alpha)}(\mathbf{x}). \tag{SI.3}$$

From these marginal probabilities come birth rates and death probabilities,

$$b_i(\mathbf{x}) := \sum_{j=1}^N e_{ij}(\mathbf{x}); \tag{SI.4a}$$

$$d_i(\mathbf{x}) := \sum_{j=1}^N e_{ji}(\mathbf{x}). \tag{SI.4b}$$

Since the population size is fixed, the average birth rate, $\overline{b}(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^N b_i(\mathbf{x})$, coincides with the average death probability, $\overline{d}(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^N d_i(\mathbf{x})$, in every state, $\mathbf{x} \in \{0,1\}^N$.

We let $\left\{ p_{(R,\alpha)}(\mathbf{x}) \right\}_{(R,\alpha)}$ depend on a non-negative parameter, $\delta \geqslant 0$, which quantifies the intensity of selection within the population. We assume that $p_{(R,\alpha)}(\mathbf{x})$ is continuously-differentiable in a neighborhood of $\delta = 0$, and we use the "prime" notation (e.g. e_{ij}') to denote the δ-derivative evaluated at $\delta = 0$. We further assume that, for $\delta = 0$, $p_{(R,\alpha)}(\mathbf{x})$ is independent of $\mathbf{x}$ (as are the demographic variables derived from the replacement rule). We refer to the case $\delta = 0$ as neutral drift, and we use the superscript $\circ$ to denote this special case (e.g. e_{ij}°).

33

SI.1.2 Reproductive value

Suppose that $\{\pi_i\}_{i=1}^N$ consists of scalars such that, for every $i = 1, \ldots, N$,

$$\sum_{j=1}^N e_{ij}^\circ \pi_j = \sum_{j=1}^N e_{ji}^\circ \pi_i. \tag{SI.5}$$

The (unique) solution to Equation SI.5 satisfying $\sum_{i=1}^N \pi_i = 1$ is the so-called "reproductive value" (RV) of i (Maciejewski, 2014, Taylor, 1990). π_i can be interpreted as the probability that, under neutral drift, i generates a lineage that takes over the population (Allen and McAvoy, 2019). Let $\widehat{b}_i(\mathbf{x}) := \sum_{j=1}^N e_{ij}(\mathbf{x}) \pi_j$ and $\widehat{d}_i(\mathbf{x}) := \sum_{j=1}^N e_{ji}(\mathbf{x}) \pi_i$ be the RV-weighted birth and death rates, respectively. The mean change in $\sum_{i=1}^N \pi_i x_i$, the RV-weighted abundance of A in state $\mathbf{x}$, is

$$\widehat{\Delta}_{\mathrm{sel}}(\mathbf{x}) := \sum_{i=1}^N x_i \left(\widehat{b}_i(\mathbf{x}) - \widehat{d}_i(\mathbf{x}) \right), \tag{SI.6}$$

which enjoys the convenient property that $\widehat{\Delta}_{\mathrm{sel}}^\circ(\mathbf{x}) = 0$ for every $\mathbf{x}$.

SI.1.3 Mutation and the RMC distribution

For $u \in [0, 1]$, consider the Markov chain on $\{0, 1\}^N$ with transition probabilities

$$\widetilde{P}_{\mathbf{x} \to \mathbf{y}} := \begin{cases} u \frac{1}{N} & \mathbf{x} = \mathbf{A}, \ \overline{y} = \frac{1}{N}, \\[2mm] (1-u) P_{\mathbf{A} \to \mathbf{y}} & \mathbf{x} = \mathbf{A}, \ \overline{y} \neq \frac{1}{N}, \\[2mm] u \frac{1}{N} & \mathbf{x} = \mathbf{B}, \ \overline{y} = 1 - \frac{1}{N}, \\[2mm] (1-u) P_{\mathbf{B} \to \mathbf{y}} & \mathbf{x} = \mathbf{B}, \ \overline{y} \neq 1 - \frac{1}{N}, \\[2mm] P_{\mathbf{x} \to \mathbf{y}} & \mathbf{x} \notin \{\mathbf{A}, \mathbf{B}\}. \end{cases} \tag{SI.7}$$

In other words, transitions from a non-monomorphic state are the same as in the mutation-free process; in a monomorphic state, a single mutant type (B in the all-A state and A in the all-B state) will arise with probability u, and the initial location of this mutant is chosen uniformly-at-random from the vertices (see Supplementary Figure 1). When $u = 0$, we have $\widetilde{P}_{\mathbf{x} \to \mathbf{y}} = P_{\mathbf{x} \to \mathbf{y}}$ for every $\mathbf{x}, \mathbf{y} \in \{0, 1\}^N$. But when $u > 0$, this Markov chain has a single closed communicating class by the Fixation Axiom, and consequently it has a unique stationary distribution, π_{MSS}.
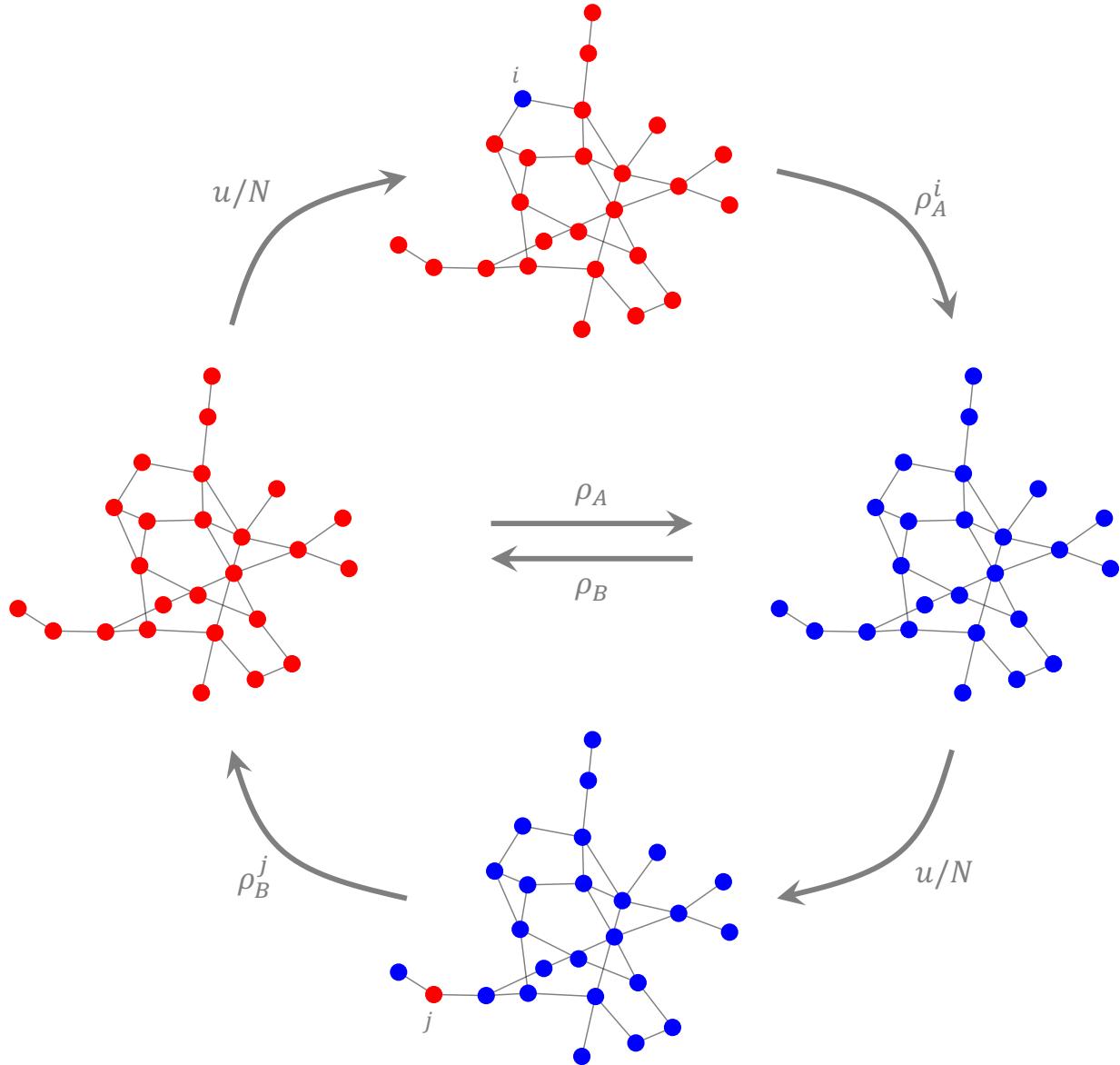
Supplementary Figure 1: Mutation and absorption into one of the two monomorphic states. A single mutant arises with probability u and is placed at location i with probability $1/N$. Mutants do not arise in the non-monomorphic states, and eventually the process returns to a monomorphic state (all-A with probability ρ_A^i and all-B with probability $1 - \rho_A^i$) by the Fixation Axiom. The mean fixation probabilities of A and be are then $\rho_A = (1/N)\sum_{i=1}^N \rho_A^i$ and $\rho_B = (1/N)\sum_{i=1}^N \rho_B^i$, respectively. When $u > 0$, the resulting Markov chain has a unique stationary distribution, π_{MSS}, which can be used to understand when $\rho_A > \rho_B$.

In order to get a better understanding of the transient-state dynamics, we define

$$\pi_{\text{RMC}}(\mathbf{x}) := \lim_{u \to 0} \frac{\pi_{\text{MSS}}(\mathbf{x})}{1 - \pi_{\text{MSS}}(\mathbf{A}) - \pi_{\text{MSS}}(\mathbf{B})}, \tag{SI.8}$$

which we call the rare-mutation conditional (RMC) distribution. One may also characterize the RMC distribution as the stationary distribution for the chain on $\{0,1\}^N - \{\mathbf{A}, \mathbf{B}\}$ defined by

$$P_{\mathbf{x} \to \mathbf{y}}^{\text{RMC}} := \begin{cases} P_{\mathbf{x} \to \mathbf{y}} + \frac{1}{N} P_{\mathbf{x} \to \mathbf{A}} & \bar{y} = \frac{1}{N}, \\[2mm] P_{\mathbf{x} \to \mathbf{y}} + \frac{1}{N} P_{\mathbf{x} \to \mathbf{B}} & \bar{y} = 1 - \frac{1}{N}, \\[2mm] P_{\mathbf{x} \to \mathbf{y}} & \frac{1}{N} < \bar{y} < 1 - \frac{1}{N}. \end{cases} \tag{SI.9}$$

The definitions of both of these chains (Equations. SI.7 and SI.9) are slightly different from their definitions given in Allen and McAvoy (2019), but nonetheless they have the following properties (whose proofs follow from the arguments of Allen and McAvoy (2019)):

Fact 1. $\lim_{u \to 0} \pi_{\text{MSS}}(\mathbf{A}) = \rho_A / (\rho_A + \rho_B)$ and $\lim_{u \to 0} \pi_{\text{MSS}}(\mathbf{B}) = \rho_B / (\rho_A + \rho_B)$;

Fact 2. The quantity

$$K := \lim_{u \to 0} \frac{u}{(1 - \pi_{\text{MSS}}(\mathbf{A}) - \pi_{\text{MSS}}(\mathbf{B}))} \tag{SI.10}$$

is strictly positive and differentiable in δ in a small neighborhood of $\delta = 0$;

Fact 3. $\mathbb{E}_{\text{RMC}}[\varphi] = K \frac{d}{du}\Big|_{u=0} \mathbb{E}_{\text{MSS}}[\varphi]$ for any $\varphi : \{0,1\}^N \to \mathbb{R}$ with $\varphi(\mathbf{A}) = \varphi(\mathbf{B}) = 0$.

Fact 4. $\mathbb{E}_{\text{RMC}}^{\circ}[x_i] = 1/2$ for $i = 1, \dots, N$.

SI.1.4 Type distributions under neutral drift

The probability that i and j have the same type in the neutral RMC distribution is

$$x_{ij} := \mathbb{E}_{\text{RMC}}^{\circ}[x_i x_j + (1 - x_i)(1 - x_j)] = 2\mathbb{E}_{\text{RMC}}^{\circ}[x_i x_j]. \tag{SI.11}$$

Let $K^\circ := \lim_{u \to 0} u / \left(1 - \pi^\circ_{\mathrm{MSS}}(\mathbf{A}) - \pi^\circ_{\mathrm{MSS}}(\mathbf{B})\right)$ be the value of K at $\delta = 0$. If $i \neq j$, then

$$
\begin{aligned}
\mathbb{E}^\circ_{\mathrm{RMC}}\left[x_i x_j\right] &= \sum_{\mathbf{x} \neq \mathbf{A}, \mathbf{B}} \pi^\circ_{\mathrm{RMC}}(\mathbf{x}) \, x_i x_j \\
&= \sum_{\mathbf{x}, \mathbf{y} \neq \mathbf{A}, \mathbf{B}} \pi^\circ_{\mathrm{RMC}}(\mathbf{y}) \, P^{\mathrm{RMC}}_{\mathbf{y} \to \mathbf{x}} x_i x_j \\
&= \sum_{\mathbf{x}, \mathbf{y} \neq \mathbf{A}, \mathbf{B}} \pi^\circ_{\mathrm{RMC}}(\mathbf{y}) \, P_{\mathbf{y} \to \mathbf{x}} x_i x_j + \sum_{\mathbf{y} \neq \mathbf{A}, \mathbf{B}} \pi^\circ_{\mathrm{RMC}}(\mathbf{y}) \sum_{\ell \neq i, j} \frac{1}{N} P_{\mathbf{y} \to \mathbf{B}} \\
&= \sum_{\mathbf{y} \neq \mathbf{A}, \mathbf{B}} \pi^\circ_{\mathrm{RMC}}(\mathbf{y}) \sum_{(R, \alpha)} p^\circ_{(R, \alpha)} y_{\widetilde{\alpha}(i)} y_{\widetilde{\alpha}(j)} - \sum_{\mathbf{y} \neq \mathbf{A}, \mathbf{B}} \pi^\circ_{\mathrm{RMC}}(\mathbf{y}) \, P_{\mathbf{y} \to \mathbf{A}} \\
&\quad + \sum_{\mathbf{y} \neq \mathbf{A}, \mathbf{B}} \pi^\circ_{\mathrm{RMC}}(\mathbf{y}) \sum_{\ell \neq i, j} \frac{1}{N} P_{\mathbf{y} \to \mathbf{B}} \\
&= \sum_{(R, \alpha)} p^\circ_{(R, \alpha)} \mathbb{E}^\circ_{\mathrm{RMC}}\left[x_{\widetilde{\alpha}(i)} x_{\widetilde{\alpha}(j)}\right] - \sum_{\mathbf{y} \neq \mathbf{A}, \mathbf{B}} \pi^\circ_{\mathrm{RMC}}(\mathbf{y}) \, P_{\mathbf{y} \to \mathbf{A}} \\
&\quad + \frac{N-2}{N} \sum_{\mathbf{y} \neq \mathbf{A}, \mathbf{B}} \pi^\circ_{\mathrm{RMC}}(\mathbf{y}) \, P_{\mathbf{y} \to \mathbf{B}} \\
&= \sum_{(R, \alpha)} p^\circ_{(R, \alpha)} \mathbb{E}^\circ_{\mathrm{RMC}}\left[x_{\widetilde{\alpha}(i)} x_{\widetilde{\alpha}(j)}\right] - \frac{K^\circ}{2} + \frac{N-2}{N} \frac{K^\circ}{2} \\
&= \sum_{(R, \alpha)} p^\circ_{(R, \alpha)} \mathbb{E}^\circ_{\mathrm{RMC}}\left[x_{\widetilde{\alpha}(i)} x_{\widetilde{\alpha}(j)}\right] - \frac{K^\circ}{N},
\end{aligned}
\tag{SI.12}
$$

Letting $\tau_{ij} := \left(1 - x_{ij}\right) / 2K^\circ = \mathbb{E}^\circ_{\mathrm{RMC}}\left[x - x_i x_j\right] / K^\circ$ gives, for $i \neq j$, the recurrence relation

$$
\tau_{ij} = \frac{1}{N} + \sum_{(R, \alpha)} p^\circ_{(R, \alpha)} \tau_{\widetilde{\alpha}(i) \widetilde{\alpha}(j)}.
\tag{SI.13}
$$

For $i = j$, we have $\tau_{ii} = 0$ for every $i = 1, \ldots, N$. This system of equations uniquely determines $\left\{\tau_{ij}\right\}_{i,j=1}^N$.

SI.1.5 Payoffs, fecundities, and social goods

In the class of evolutionary processes considered here, interactions result in payoffs, which are then converted into fecundities. These fecundities are subsequently used to update the state of the population. More abstractly, every state $\mathbf{x} \in \{0,1\}^N$ results in a vector of fecundities, $\mathbf{F} \in [0, \infty)^N$, with one entry for every individual. However, this state-to-fecundity mapping need not be deterministic; it could also be stochastic (Figure 15). In this section, we show that stochastic mappings can be reduced to deterministic mappings under weak selection.

In practice, state-based replacement rules (introduced in §SI.1.1) often depend on the state, $\mathbf{x}$, only through its effect on fecundity. In other words, if there is a deterministic state-to-fecundity map, $\mathbf{x} \mapsto \mathbf{F}(\mathbf{x})$, where $\mathbf{F} \in [0, \infty)^N$, then $p_{(R, \alpha)}(\mathbf{x}) = p_{(R, \alpha)}(\mathbf{F}(\mathbf{x}))$. To avoid confusion when

we consider stochastic state-to-fecundity mappings below, we denote by $p_{(R,\alpha)}$ a replacement rule that is a function of state, $\mathbf{x}$, and by $q_{(R,\alpha)}$ a replacement rule that is a function of fecundity, $\mathbf{F}$. For example, DB updating in a graph-structured population, $\left(w_{ij}\right)_{i,j=1}^{N}$, is defined by the rule

$$
q_{(R,\alpha)}\left(\mathbf{F}\right) = \begin{cases} \dfrac{1}{N}\dfrac{F_{\alpha(i)}w_{\alpha(i)i}}{\sum_{j=1}^{N}F_j w_{ji}} & R = \{i\} \text{ for some } i, \\[4mm] 0 & \text{otherwise.} \end{cases} \tag{SI.14}
$$

In most traditional formulations of evolutionary games with DB updating, there exists a payoff function $u : \{0,1\}^N \to \mathbb{R}^N$ that gives a payoff vector for the population as a function of the state, $\mathbf{x}$. The payoff for player i, $u_i\left(\mathbf{x}\right)$, is then converted to fecundity, $F_i\left(\mathbf{x}\right)$, by letting $F_i\left(\mathbf{x}\right) := \exp\{\delta u_i\left(\mathbf{x}\right)\}$ for some selection intensity parameter, $\delta \geqslant 0$. Thus, we can write

$$
p_{(R,\alpha)}\left(\mathbf{x}\right) = \begin{cases} \dfrac{1}{N}\dfrac{F_{\alpha(i)}(\mathbf{x})w_{\alpha(i)i}}{\sum_{j=1}^{N}F_j(\mathbf{x})w_{ji}} & R = \{i\} \text{ for some } i, \\[4mm] 0 & \text{otherwise.} \end{cases} \tag{SI.15}
$$

Suppose now that we have a fixed fecundity-based replacement rule, q, together with a *stochastic* state-to-fecundity mapping. Thus, for every state $\mathbf{x} \in \{0,1\}^N$, there is a distribution over fecundity vectors $\mathbf{F} \in [0,\infty)^N$. If $\mathbb{E}_{\mathbf{x}}$ denotes expectation with respect to this distribution, then one obtains a state-based rule,

$$
p_{(R,\alpha)}\left(\mathbf{x}\right) := \mathbb{E}_{\mathbf{x}}\left[q_{(R,\alpha)}\right]. \tag{SI.16}
$$

We consider fecundity-based replacement rules for which $q_{(R,\alpha)}\left(\mathbf{F}\right) = q_{(R,\alpha)}^{\circ} + \delta q_{(R,\alpha)}'\left(\mathbf{F}\right) + O\left(\delta^2\right)$ for some function $q_{(R,\alpha)}'\left(\mathbf{F}\right)$ whenever $\delta \ll 1$. Moreover, we assume that the distribution on fecundity is determined by randomness in the payoffs. Specifically, every individual, i, receives a payoff, u_i, based on some probability distribution. This payoff is then converted to fecundity via the formula $F_i\left(\delta\right) = \exp\{\delta u_i\}$, where $\delta \geqslant 0$ is the intensity of selection. It follows that

$$
\frac{d}{d\delta}\mathbb{E}_{\mathbf{x}}\left[F_i\left(\delta\right)\right] = \mathbb{E}_{\mathbf{x}}\left[F_i'\left(\delta\right)\right] \tag{SI.17}
$$

for every $i = 1, \ldots, N$.

Consider the "averaged" replacement rule, $\overline{p}_{(R,\alpha)}(\mathbf{x}) := q_{(R,\alpha)}(\mathbb{E}_{\mathbf{x}}[\mathbf{F}])$. By the chain rule,

$$
\begin{aligned}
p_{(R,\alpha)}(\mathbf{x}) &= \mathbb{E}_{\mathbf{x}}\left[q_{(R,\alpha)}\right] \\
&= q_{(R,\alpha)}^{\circ} + \delta \mathbb{E}_{\mathbf{x}}\left[q'_{(R,\alpha)}\right] + O\left(\delta^2\right) \\
&= q_{(R,\alpha)}^{\circ} + \delta \sum_{i=1}^{N} \left.\frac{\partial q_{(R,\alpha)}}{\partial F_i}\right|_{\mathbf{F}=\mathbf{F}(0)} \mathbb{E}_{\mathbf{x}}\left[F_i'(0)\right] + O\left(\delta^2\right) \\
&= q_{(R,\alpha)}^{\circ} + \delta \sum_{i=1}^{N} \left.\frac{\partial q_{(R,\alpha)}}{\partial F_i}\right|_{\mathbf{F}=\mathbf{F}(0)} \left.\frac{d\mathbb{E}_{\mathbf{x}}\left[F_i(\delta)\right]}{d\delta}\right|_{\delta=0} + O\left(\delta^2\right).
\end{aligned}
\tag{SI.18}
$$

It follows that

$$
\left.\frac{d}{d\delta}\right|_{\delta=0} p_{(R,\alpha)}(\mathbf{x}) = \left.\frac{d}{d\delta}\right|_{\delta=0} \overline{p}_{(R,\alpha)}(\mathbf{x}).
\tag{SI.19}
$$

From this equation, together with the fact that fixation probabilities are defined by linear systems whose coefficients are based on the replacement rule, we conclude that the first-order behavior of the fixation probabilities must coincide under realized and expected payoffs. Thus, without a loss of generality, we may assume that the state-to-fecundity mapping is deterministic.

Suppose that type A at location i pays C_{ij} to donate B_{ij} to j. Type B at this same location pays c_{ij} to donate b_{ij} to j. Although this formulation accounts for general additive interactions, we are mostly interested in the case in which A is a producer and B is a non-producer, meaning $b_{ij} = c_{ij} = 0$ for every i and j. In the most general case, however, the cumulative payoff to i is

$$
u_i(\mathbf{x}) = \sum_{j=1}^{N} \left(-x_i C_{ij} - (1-x_i) c_{ij} + x_j B_{ji} + (1-x_j) b_{ji}\right).
\tag{SI.20}
$$

This formulation can account for both accumulated and averaged payoffs due to the dependence of B_{ij}, C_{ij}, b_{ij}, and c_{ij} on both i and j; we give explicit examples along these lines in §SI.3.

SI.1.6 Marginal replacement effects

By the results of §SI.1.5, we may assume (without a loss of generality) that the payoff-to-fecundity map is deterministic. Let $u : \{0,1\}^N \to \mathbb{R}^N$ be a payoff function that assigns a real number, $u_i(\mathbf{x})$, to each individual, i, and state, $\mathbf{x}$. Individual i's payoff is then converted to fecundity by $F_i(\mathbf{x}) := \exp\{\delta u_i(\mathbf{x})\}$. Moreover, we assume that for every $\mathbf{x} \in \{0,1\}^N$, we have $e_{ij}(\mathbf{x}) = e_{ij}(\mathbf{F})$ (meaning e_{ij} depends on $\mathbf{x}$ only through the effects of $\mathbf{x}$ on fecundity).

Letting $m_k^{ij} := \frac{\partial}{\partial F_k}\Big|_{\mathbf{F}=\mathbf{F}(0)} e_{ij}(\mathbf{F})$ be the *marginal effect of k on i replacing j*, we see that

$$e'_{ij}(\mathbf{x}) = \sum_{k=1}^{N} \left(\frac{\partial}{\partial F_k}\Big|_{\mathbf{F}=\mathbf{F}(0)} e_{ij}(\mathbf{F}) \right) \left(\frac{d}{d\delta}\Big|_{\delta=0} F_k(\delta) \right) = \sum_{k=1}^{N} m_k^{ij} u_k(\mathbf{x}). \tag{SI.21}$$

Note, in particular, that m_k^{ij} is independent of the payoffs and structure of the game. It can also be easily calculated for any process from the details of the update rule; we give examples below.

SI.1.7   Condition for evolutionary success of A relative to B

Our condition for the success of A relative to B is based on the following result, which is a modification of Theorem 8 of Allen and McAvoy (2019):

Lemma 1. For the class of processes described herein,

$$\rho'_A > \rho'_B \iff \mathbb{E}^\circ_{\mathrm{RMC}}\left[\widehat{\Delta}'_{\mathrm{sel}}\right] > 0. \tag{SI.22}$$

Sketch of proof. The expected change in RV-weighted abundance of A in state $\mathbf{x}$ is

$$\widehat{\Delta}(\mathbf{x}) = \begin{cases} -u\frac{1}{N} & \mathbf{x} = \mathbf{A}, \\ u\frac{1}{N} & \mathbf{x} = \mathbf{B}, \\ \widehat{\Delta}_{\mathrm{sel}}(\mathbf{x}) & \mathbf{x} \notin \{\mathbf{A}, \mathbf{B}\}. \end{cases} \tag{SI.23}$$

Since this expected change must average out to 0 over the stationary distribution for the chain,

$$0 = \mathbb{E}_{\mathrm{MSS}}\left[\widehat{\Delta}\right] = \mathbb{E}_{\mathrm{MSS}}\left[\widehat{\Delta}_{\mathrm{sel}}\right] - u\frac{1}{N}\pi_{\mathrm{MSS}}(\mathbf{A}) + u\frac{1}{N}\pi_{\mathrm{MSS}}(\mathbf{B}). \tag{SI.24}$$

Differentiating this equation with respect to u at $u = 0$ gives

$$\frac{N}{K}\mathbb{E}_{\mathrm{RMC}}\left[\widehat{\Delta}_{\mathrm{sel}}\right] = \lim_{u\to 0}(\pi_{\mathrm{MSS}}(\mathbf{A}) - \pi_{\mathrm{MSS}}(\mathbf{B}))$$

$$= \frac{\rho_A - \rho_B}{\rho_A + \rho_B}, \tag{SI.25}$$

where, again, $K := \lim_{u\to 0} u/(1 - \pi_{\mathrm{MSS}}(\mathbf{A}) - \pi_{\mathrm{MSS}}(\mathbf{B}))$. Since $\widehat{\Delta}^\circ_{\mathrm{sel}}(\mathbf{x}) = 0$ for every $\mathbf{x} \in \{0,1\}^N$, differentiating Equation SI.25 with respect to δ at $\delta = 0$ gives

$$\frac{N}{K^\circ}\mathbb{E}^\circ_{\mathrm{RMC}}\left[\widehat{\Delta}'_{\mathrm{sel}}\right] = \frac{\rho'_A - \rho'_B}{2\rho^\circ_A}. \tag{SI.26}$$

by Facts 1–3 in §SI.1.3. Since $K^\circ > 0$ and $\rho^\circ_A = (1/N)\sum_{i=1}^{N}\pi_i = 1/N$ (see §SI.1.2), it

follows that $\rho'_A > \rho'_B$ if and only if $\mathbb{E}^{\circ}_{\text{RMC}} \left[\widehat{\Delta}'_{\text{sel}} \right] > 0$, as desired. $\qquad\square$

Theorem 1.

$$\rho'_A > \rho'_B \iff \sum_{i,j,k,\ell=1}^{N} \pi_i m_k^{ji} \left(-\left(x_{jk} - x_{ik} \right) \left(C_{k\ell} - c_{k\ell} \right) + \left(x_{j\ell} - x_{i\ell} \right) \left(B_{\ell k} - b_{\ell k} \right) \right) > 0.$$

(SI.27)

Proof. A straightforward calculation using the definition $m_k^{ij} := \left. \frac{\partial}{\partial F_k} \right|_{\mathbf{F}=\mathbf{F}(0)} e_{ij} \left(\mathbf{F} \right)$ gives

$$
\begin{aligned}
\widehat{\Delta}'_{\text{sel}} \left(\mathbf{x} \right) &= \sum_{i=1}^{N} x_i \left(\widehat{b}'_i \left(\mathbf{x} \right) - \widehat{d}'_i \left(\mathbf{x} \right) \right) \\
&= \sum_{i,j=1}^{N} x_i \left(e'_{ij} \left(\mathbf{x} \right) \pi_j - e'_{ji} \left(\mathbf{x} \right) \pi_i \right) \\
&= \sum_{i,j=1}^{N} \pi_i \left(x_j - x_i \right) e'_{ji} \left(\mathbf{x} \right) \\
&= \sum_{i,j,k=1}^{N} \pi_i m_k^{ji} \left(x_j - x_i \right) u_k \left(\mathbf{x} \right).
\end{aligned}
$$

(SI.28)

Since $x_{ij} := 2\mathbb{E}^{\circ}_{\text{RMC}} \left[x_i x_j \right]$, we see that

$$
\begin{aligned}
\mathbb{E}^{\circ}_{\text{RMC}} \left[\left(x_j - x_i \right) u_k \left(\mathbf{x} \right) \right] &= \mathbb{E}^{\circ}_{\text{RMC}} \left[\left(x_j - x_i \right) \sum_{\ell=1}^{N} \left(\begin{smallmatrix} -x_k C_{k\ell} - (1-x_k) c_{k\ell} \\ +x_\ell B_{\ell k} + (1-x_\ell) b_{\ell k} \end{smallmatrix} \right) \right] \\
&= \frac{1}{2} \sum_{\ell=1}^{N} \left(-x_{jk} C_{k\ell} - \left(1 - x_{jk} \right) c_{k\ell} + x_{j\ell} B_{\ell k} + \left(1 - x_{j\ell} \right) b_{\ell k} \right) \\
&\quad - \frac{1}{2} \sum_{\ell=1}^{N} \left(-x_{ik} C_{k\ell} - \left(1 - x_{ik} \right) c_{k\ell} + x_{i\ell} B_{\ell k} + \left(1 - x_{i\ell} \right) b_{\ell k} \right) \\
&= \frac{1}{2} \sum_{\ell=1}^{N} \left(-x_{jk} \left(C_{k\ell} - c_{k\ell} \right) + x_{j\ell} \left(B_{\ell k} - b_{\ell k} \right) \right) \\
&\quad - \frac{1}{2} \sum_{\ell=1}^{N} \left(-x_{ik} \left(C_{k\ell} - c_{k\ell} \right) + x_{i\ell} \left(B_{\ell k} - b_{\ell k} \right) \right).
\end{aligned}
$$

(SI.29)

The theorem then follows at once from Lemma 1. $\qquad\square$

Corollary 1. When A is a producer and B is a non-producer ($b_{ij} = c_{ij} = 0$ for every i and j),

$$\rho'_A > \rho'_B \iff \sum_{i,j,k,\ell=1}^{N} \pi_i m_k^{ji} \left(-x_{jk} C_{k\ell} + x_{j\ell} B_{\ell k}\right) > \sum_{i,j,k,\ell=1}^{N} \pi_i m_k^{ji} \left(-x_{ik} C_{k\ell} + x_{i\ell} B_{\ell k}\right).$$

$$(SI.30)$$

Remark 1. Although Theorem 1 and Corollary 1 are stated in terms of x_{ij}, they can be evaluated by replacing $1 - x_{ij}$ by τ_{ij} since $\tau_{ij} := \left(1 - x_{ij}\right)/2K^\circ$ for some $K^\circ > 0$. τ can be easily calculated using Equation SI.13 and the fact that $\tau_{ii} = 0$ for $i = 1, \ldots, N$. However, the statements of Theorem 1 and Corollary 1 are somewhat more intuitive using these probabilities, x_{ij}, directly, since then the constituent terms can be interpreted as expected payoffs.

Returning to Equation 1 in the main text, let *(i)* P_{ji} be the probability that a producer in location j replaces a random individual in location i and *(ii)* Q_{ji} be the probability that a random individual in location j replaces a producer in location i. Since $P_{ji}^\circ = Q_{ji}^\circ$ (i.e. when $\delta = 0$) and

$$P'_{ji} = \sum_{k,\ell=1}^{N} m_k^{ji} \left(-x_{jk} C_{k\ell} + x_{j\ell} B_{\ell k}\right);$$

$$(SI.31a)$$

$$Q'_{ji} = \sum_{k,\ell=1}^{N} m_k^{ji} \left(-x_{ik} C_{k\ell} + x_{i\ell} B_{\ell k}\right),$$

$$(SI.31b)$$

we see that the signs of the first-order terms of $\rho_A - \rho_B$ and $\sum_{i,j=1}^{N} \pi_i \left(P_{ji} - Q_{ji}\right)$ agree.

SI.1.8 Relationship to inclusive fitness theory

Inclusive fitness theory (Hamilton, 1964, Lehmann and Rousset, 2014, Birch, 2017) is often used to model the evolution of social behavior. According to this theory, individuals evolve to act as if maximizing a quantity called inclusive fitness, which is a sum of fitness effects caused by an actor, each weighted by relatedness to the recipient. Careful analysis has revealed that to define the inclusive fitness of an individual requires weak selection, additivity of fitness effects, and other assumptionsBirch (2017), Nowak et al. (2010), Allen and Nowak (2016), which hold in the model considered here.

Before identifying inclusive fitness effects, we must define individual fitness. The fitness of a vertex i in a given state $\mathbf{x}$ can be defined asAllen and McAvoy (2019)

$$\omega_i(\mathbf{x}) = \pi_i + \widehat{b}_i(\mathbf{x}) - \widehat{d}_i(\mathbf{x}) = \pi_i + \sum_{j=1}^{N} \left(e_{ij}(\mathbf{x}) \pi_j - e_{ji}(\mathbf{x}) \pi_i\right).$$

$$(SI.32)$$

We observe that for neutral drift, fitness is simply equal to reproductive value, $\omega_i^\circ = \pi_i$. For weak selection, we have

$$\omega'_i(\mathbf{x}) = \sum_{j=1}^{N} \left(e'_{ij}(\mathbf{x}) \pi_j - e'_{ji}(\mathbf{x}) \pi_i\right).$$

$$(SI.33)$$

From Equation SI.21, this can be written as

$$\omega_i'(\mathbf{x}) = \sum_{k=1}^{N} M_k^i u_k(\mathbf{x}), \tag{SI.34}$$

where the quantity

$$M_k^i := \left.\frac{\partial}{\partial F_k}\right|_{\mathbf{F}=\mathbf{F}(0)} \left(e_{ij}(\mathbf{F})\pi_j - e_{ji}(\mathbf{F})\pi_i\right) = \sum_{j=1}^{N}\left(m_k^{ij}\pi_j - m_k^{ji}\pi_i\right) \tag{SI.35}$$

describes how the payoff to vertex k affects the fitness of vertex i. We observe from Equations (SI.28) and (SI.33) that

$$\widehat{\Delta}_{\text{sel}}'(\mathbf{x}) = \sum_{i=1}^{N} x_i \omega_i'(\mathbf{x}). \tag{SI.36}$$

Substituting from Equation SI.20, we obtain

$$\begin{aligned}
\omega_i'(\mathbf{x}) &= \sum_{k=1}^{N}\left[M_k^i \sum_{\ell=1}^{N}\left(-x_k C_{k\ell} - (1-x_k)c_{k\ell} + x_\ell B_{\ell k} + (1-x_\ell)b_{\ell k}\right)\right] \\
&= \sum_{k=1}^{N}\left[M_k^i \sum_{\ell=1}^{N}\left(-x_k C_{k\ell} - (1-x_k)c_{k\ell}\right) + \sum_{\ell=1}^{N} M_\ell^i\left(x_k B_{k\ell} + (1-x_k)b_{k\ell}\right)\right] \\
&= \sum_{k=1}^{N}\left[x_k \sum_{\ell=1}^{N}\left(-C_{k\ell}M_k^i + B_{k\ell}M_\ell^i\right) + (1-x_k)\sum_{\ell=1}^{N}\left(-c_{k\ell}M_k^i + b_{k\ell}M_\ell^i\right)\right].
\end{aligned} \tag{SI.37}$$

The final line of Equation SI.37 expresses the neighbor-modulated fitness of vertex i, in that it identifies the contribution (fitness effect) of each vertex k to the fitness of i. Specifically, this fitness effect is $\sum_{\ell=1}^{N}\left(-C_{k\ell}M_k^i + B_{k\ell}M_\ell^i\right)$ for vertices k of type A, and $\sum_{\ell=1}^{N}\left(-c_{k\ell}M_k^i + b_{k\ell}M_\ell^i\right)$ for vertices k of type B.

To move from neighbor-modulated to inclusive fitness, one must causally attribute every fitness effect on every individual to a particular actor in the population (Birch, 2017, Allen and Nowak, 2016). Even in our simple model, this causal attribution is rather arbitrary and artificial. For example, if individual i gives benefit B_{ij} to individual j, which in turn alters individual k's fitness by an amount $B_{ij}M_j^k$, to whom should this effect on k's fitness be causally attributed? To i, to j, or both? To make progress, we adopt the convention that all fitness effects are attributed to the originator of the social good (e.g., terms of the form $B_{ij}M_j^k$ are attributed to individual i).

To formulate the inclusive fitness of individual i, we now multiply each effect attributable to individual i by the relatedness of i to the recipient. For a given state $\mathbf{x}$, we use a notion of "relatedness in state," such that the relatedness of two individuals is one if they have the same type and zero otherwise; as a formula, the relatedness of i and j is $x_i x_j + (1-x_i)(1-x_j)$. (Later, we will average over states to obtain relatedness coefficients between zero and one.) Multiplying the fitness effects from Eq. (SI.37) by the corresponding relatedness-in-state coefficients, we obtain

the inclusive fitness effect of vertex i in state $\mathbf{x}$ as

$$
\omega_i^{\text{IF}}(\mathbf{x}) =
\begin{cases}
\sum_{k,\ell=1}^{N} \left(-C_{i\ell} M_i^k + B_{i\ell} M_\ell^k \right) x_k & x_i = 1, \\[2ex]
\sum_{k,\ell=1}^{N} \left(-c_{i\ell} M_i^k + b_{i\ell} M_\ell^k \right) (1 - x_k) & x_i = 0.
\end{cases}
\tag{SI.38}
$$

We immediately see that

$$
\begin{aligned}
\sum_{i=1}^{N} x_i \omega_i^{\text{IF}}(\mathbf{x}) &= \sum_{i,k=1}^{N} \left(-C_{i\ell} M_i^k + B_{i\ell} M_\ell^k \right) x_i x_k \\
&= \sum_{i,k=1}^{N} \left(-C_{k\ell} M_k^i + B_{k\ell} M_\ell^i \right) x_i x_k \\
&= \sum_{i=1}^{N} x_i \omega_i'(\mathbf{x}) \\
&= \widehat{\Delta}_{\text{sel}}'(\mathbf{x}).
\end{aligned}
\tag{SI.39}
$$

Inclusive fitness, when it exists, is therefore an alternative accounting method that leads to the same result for the gradient of selection, $\widehat{\Delta}_{\text{sel}}'(\mathbf{x})$.

To obtain fitness quantities that apply to the overall evolutionary process, we average over the neutral RMC distribution, conditioned on the type of vertex i. First, we consider neighbor-modulated fitness. If vertex i has type A then, noting that $\mathbb{E}_{\text{RMC}}^{\circ}\left[x_j \mid x_i = 1 \right] = x_{ij}$, we obtain

$$
\begin{aligned}
\mathbb{E}_{\text{RMC}}^{\circ}\left[\omega_i'(\mathbf{x}) \mid x_i = 1 \right] = \sum_{k=1}^{N} \Bigg[& x_{ik} \sum_{\ell=1}^{N} \left(-C_{k\ell} M_k^i + B_{k\ell} M_\ell^i \right) \\
& + (1 - x_{ik}) \sum_{\ell=1}^{N} \left(-c_{k\ell} M_k^i + b_{k\ell} M_\ell^i \right) \Bigg].
\end{aligned}
\tag{SI.40}
$$

If i has type B, then since $\mathbb{E}_{\text{RMC}}^{\circ}\left[x_j \mid x_i = 0 \right] = 1 - x_{ij}$, we have

$$
\begin{aligned}
\mathbb{E}_{\text{RMC}}^{\circ}\left[\omega_i'(\mathbf{x}) \mid x_i = 0 \right] = \sum_{k=1}^{N} \Bigg[& (1 - x_{ik}) \sum_{\ell=1}^{N} \left(-C_{k\ell} M_k^i + B_{k\ell} M_\ell^i \right) \\
& + x_{ik} \sum_{\ell=1}^{N} \left(-c_{k\ell} M_k^i + b_{k\ell} M_\ell^i \right) \Bigg].
\end{aligned}
\tag{SI.41}
$$

Now, turning to inclusive fitness, we have

$$
\mathbb{E}_{\text{RMC}}^{\circ}\left[\omega_i^{\text{IF}}(\mathbf{x}) \mid x_i = 1 \right] = \sum_{k,\ell=1}^{N} \left(-C_{i\ell} M_i^k + B_{i\ell} M_\ell^k \right) x_{ik},
\tag{SI.42}
$$

44

for type A, and

$$\mathbb{E}^{\circ}_{\text{RMC}}\left[\omega^{\text{IF}}_i(\mathbf{x}) \mid x_i = 0\right] = \sum_{k,\ell=1}^{N} \left(-c_{i\ell}M^k_i + b_{i\ell}M^k_\ell\right) x_{ik}, \tag{SI.43}$$

for type B.

From Equation SI.39 we have $\mathbb{E}^{\circ}_{\text{RMC}}\left[\widehat{\Delta}'_{\text{sel}}\right] = \mathbb{E}^{\circ}_{\text{RMC}}\left[\sum_{i=1}^{N} x_i\omega^{\text{IF}}_i\right] = \mathbb{E}^{\circ}_{\text{RMC}}\left[\sum_{i=1}^{N} x_i\omega'_i\right]$, again demonstrating the equivalence of the neighbor-modulated and inclusive fitness accounting methods, in the case of this model.

SI.2 Specific update rules

We now turn to specific examples of update rules (Figure 3) in graph-structured populations. Let $(w_{ij})^{N}_{i,j=1}$ be an undirected, unweighted, connected graph on N vertices. The degree of vertex i is simply the number of links connected to that vertex, i.e. $w_i := \sum_{j=1}^{N} w_{ij}$. This graph defines the structure of the population, with links indicating neighbor relationships.

SI.2.1 Pairwise-comparison (PC) updating

Under PC updating (see Figure 3), the probability of replacement event (R, α) is

$$p_{(R,\alpha)}(\mathbf{x}) = \begin{cases} \frac{1}{N}p_{i\alpha(i)}\frac{F_{\alpha(i)}(\mathbf{x})}{F_i(\mathbf{x})+F_{\alpha(i)}(\mathbf{x})} & R = \{i\} \text{ for some } i \in \{1,\dots,N\}, \ \alpha(i) \neq i, \\ \\ \frac{1}{N}\sum_{j=1}^{N} p_{ij}\frac{F_i(\mathbf{x})}{F_i(\mathbf{x})+F_j(\mathbf{x})} & R = \{i\} \text{ for some } i \in \{1,\dots,N\}, \ \alpha(i) = i, \\ \\ 0 & \text{otherwise.} \end{cases} \tag{SI.44}$$

For $i, j = 1, \dots, N$, the marginal probability that i transmits its offspring to $j \neq i$ is

$$e_{ij}(\mathbf{x}) = \frac{1}{N}p_{ji}\frac{F_i(\mathbf{x})}{F_i(\mathbf{x}) + F_j(\mathbf{x})}, \tag{SI.45}$$

which gives a marginal effect of k on i replacing j of

$$m^{ij}_k = \begin{cases} \frac{1}{4N}p_{ji} & k = i, \\ \\ -\frac{1}{4N}p_{ji} & k = j, \\ \\ 0 & k \neq i, j. \end{cases} \tag{SI.46}$$

It follows that

$$\sum_{i,j,k,\ell=1}^{N} \pi_i m_k^{ji} \left(-x_{jk}C_{k\ell} + x_{j\ell}B_{\ell k}\right) = \sum_{i,j=1}^{N} \pi_i \left[\frac{1}{4N} p_{ij} \sum_{\ell=1}^{N} \left(-x_{jj}C_{j\ell} + x_{j\ell}B_{\ell j}\right) \right.$$
$$\left. - \frac{1}{4N} p_{ij} \sum_{\ell=1}^{N} \left(-x_{ji}C_{i\ell} + x_{j\ell}B_{\ell i}\right) \right]; \quad \text{(SI.47a)}$$

$$\sum_{i,j,k,\ell=1}^{N} \pi_i m_k^{ji} \left(-x_{ik}C_{k\ell} + x_{i\ell}B_{\ell k}\right) = \sum_{i,j=1}^{N} \pi_i \left[\frac{1}{4N} p_{ij} \sum_{\ell=1}^{N} \left(-x_{ij}C_{j\ell} + x_{i\ell}B_{\ell j}\right) \right.$$
$$\left. - \frac{1}{4N} p_{ij} \sum_{\ell=1}^{N} \left(-x_{ii}C_{i\ell} + x_{i\ell}B_{\ell i}\right) \right]. \quad \text{(SI.47b)}$$

The reproductive value of i under PC updating is $\pi_i = w_i / \sum_{k=1}^{N} w_k$. Since $\pi_i p_{ij} = \pi_j p_{ji}$ for every i and j, it follows that $\rho_A > \rho_B$ for small $\delta > 0$ if and only if

$$\sum_{i=1}^{N} \pi_i \sum_{\ell=1}^{N} \left(-x_{ii}C_{i\ell} + x_{i\ell}B_{\ell i}\right) > \sum_{i,j=1}^{N} \pi_i p_{ij} \sum_{\ell=1}^{N} \left(-x_{ij}C_{j\ell} + x_{i\ell}B_{\ell j}\right), \quad \text{(SI.48)}$$

which gives Equation 7 in the main text. To evaluate this condition, recall that we can replace $1 - x_{ij}$ by τ_{ij}, where, by Equation SI.13,

$$\tau_{ij} = \frac{1}{N} + \frac{1}{2N} \sum_{k=1}^{N} p_{ik}\tau_{kj} + \frac{1}{2N} \sum_{k=1}^{N} p_{jk}\tau_{ik} + \left(1 - \frac{1}{N}\right)\tau_{ij}, \quad \text{(SI.49)}$$

which implies that $\tau_{ij} = 1 + (1/2)\sum_{k=1}^{N} p_{ik}\tau_{kj} + (1/2)\sum_{k=1}^{N} p_{jk}\tau_{ik}$ whenever $i \neq j$. (For $i = j$, we have $\tau_{ii} = 0$ for $i = 1, \dots, N$). Using this substitution, we obtain Equation 9 in the main text.

SI.2.2   Death-birth (DB) updating

Under DB updating (see Figure 3), the probability of replacement event (R, α) is

$$p_{(R,\alpha)}(\mathbf{x}) = \begin{cases} \dfrac{1}{N} \dfrac{w_{i\alpha(i)} F_{\alpha(i)}(\mathbf{x})}{\sum_{k=1}^{N} w_{ik} F_k(\mathbf{x})} & R = \{i\} \text{ for some } i \in \{1, \dots, N\}, \\[2em] 0 & \text{otherwise.} \end{cases} \quad \text{(SI.50)}$$

For $i, j = 1, \dots, N$, the marginal probability that i transmits its offspring to j is

$$e_{ij}(\mathbf{x}) = \frac{1}{N} \frac{w_{ji} F_i(\mathbf{x})}{\sum_{k=1}^{N} w_{jk} F_k(\mathbf{x})}. \quad \text{(SI.51)}$$

Therefore, the marginal effect of k on i replacing j is

$$m_k^{ij} = \begin{cases} \frac{1}{N} p_{ji} \left(1 - p_{ji}\right) & k = i, \\ \\ -\frac{1}{N} p_{ji} p_{jk} & k \neq i. \end{cases} \tag{SI.52}$$

Again, the reproductive value of i is $\pi_i = w_i / \sum_{k=1}^{N} w_k$. Since $\pi_i p_{ij} = \pi_j p_{ji}$ for every i and j,

$$\sum_{i,j,k,\ell=1}^{N} \pi_i m_k^{ji} \left(-x_{jk} C_{k\ell} + x_{j\ell} B_{\ell k}\right) = \sum_{i=1}^{N} \pi_i \frac{1}{N} \sum_{\ell=1}^{N} \left(-x_{ii} C_{i\ell} + x_{i\ell} B_{\ell i}\right)$$

$$- \sum_{i=1}^{N} \pi_i \sum_{k=1}^{N} \frac{1}{N} p_{ik}^{(2)} \sum_{\ell=1}^{N} \left(-x_{ik} C_{k\ell} + x_{i\ell} B_{\ell k}\right); \tag{SI.53a}$$

$$\sum_{i,j,k,\ell=1}^{N} \pi_i m_k^{ji} \left(-x_{ik} C_{k\ell} + x_{i\ell} B_{\ell k}\right) = 0. \tag{SI.53b}$$

Consequently, we see that $\rho_A > \rho_B$ for small $\delta > 0$ if and only if

$$\sum_{i=1}^{N} \pi_i \sum_{\ell=1}^{N} \left(-x_{ii} C_{i\ell} + x_{i\ell} B_{\ell i}\right) > \sum_{i,j=1}^{N} \pi_i p_{ij}^{(2)} \sum_{\ell=1}^{N} \left(-x_{ij} C_{j\ell} + x_{i\ell} B_{\ell j}\right), \tag{SI.54}$$

which gives Equation 10 in the main text. To evaluate this condition, for $i \neq j$ we can replace $1 - x_{ij}$ by τ_{ij}, where, by Equation SI.13,

$$\tau_{ij} = \frac{1}{N} + \frac{1}{N} \sum_{k=1}^{N} p_{ik} \tau_{kj} + \frac{1}{N} \sum_{k=1}^{N} p_{jk} \tau_{ik} + \left(1 - \frac{2}{N}\right) \tau_{ij}, \tag{SI.55}$$

which implies that $\tau_{ij} = 1/2 + (1/2) \sum_{k=1}^{N} p_{ik} \tau_{kj} + (1/2) \sum_{k=1}^{N} p_{jk} \tau_{ik}$ whenever $i \neq j$. (Again, we have $\tau_{ii} = 0$ for $i = 1, \ldots, N$.) Since we can evaluate Equation 11 using any non-zero multiple of τ, it suffices to replace τ by 2τ and use the recurrence $\tau_{ij} = 1 + (1/2) \sum_{k=1}^{N} p_{ik} \tau_{kj} + (1/2) \sum_{k=1}^{N} p_{jk} \tau_{ik}$ when $i \neq j$, the same as that of PC updating (and also of Allen et al., 2017, Fotouhi et al., 2018, Allen and McAvoy, 2019). This substitution leads to Equation 11 in the main text.

SI.2.3 Imitation (IM) updating

Under IM updating (see Figure 3), the probability of replacement event (R, α) is

$$
p_{(R,\alpha)}(\mathbf{x}) = \begin{cases} \dfrac{1}{N} \dfrac{w_{i\alpha(i)} F_{\alpha(i)}(\mathbf{x})}{F_i(\mathbf{x}) + \sum_{k=1}^N w_{ik} F_k(\mathbf{x})} & R = \{i\} \text{ for some } i \in \{1, \dots, N\}, \ \alpha(i) \neq i, \\[3mm] \dfrac{1}{N} \dfrac{F_i(\mathbf{x})}{F_i(\mathbf{x}) + \sum_{k=1}^N w_{ik} F_k(\mathbf{x})} & R = \{i\} \text{ for some } i \in \{1, \dots, N\}, \ \alpha(i) = i, \\[3mm] 0 & \text{otherwise.} \end{cases} \quad \text{(SI.56)}
$$

For $i, j = 1, \dots, N$, the probability that i transmits its offspring to $j \neq i$ is

$$
e_{ij}(\mathbf{x}) = \frac{1}{N} \frac{w_{ji} F_i(\mathbf{x})}{F_j(\mathbf{x}) + \sum_{k=1}^N w_{jk} F_k(\mathbf{x})}, \quad \text{(SI.57)}
$$

and the marginal effect of k on i replacing $j \neq i$ is

$$
m_k^{ij} = \begin{cases} \dfrac{1}{N} \dfrac{w_{ji}}{w_j+1} \left(1 - \dfrac{w_{ji}}{w_j+1}\right) & k = i, \\[3mm] -\dfrac{1}{N} \left(\dfrac{w_{ji}}{w_j+1}\right)^2 & k = j, \\[3mm] -\dfrac{1}{N} \dfrac{w_{ji}}{w_j+1} \dfrac{w_{jk}}{w_j+1} & k \neq i, j. \end{cases} \quad \text{(SI.58)}
$$

Let $(\widetilde{w}_{ij})_{i,j=1}^N$ be the matrix defined by $\widetilde{w}_{ii} = 1$ and $\widetilde{w}_{ij} = w_{ij}$ for $i \neq j$. We also define an analogue of $p_{ij} = w_{ij}/w_i$, namely $\widetilde{p}_{ij} := \widetilde{w}_{ij}/\widetilde{w}_i$. Under IM updating, reproductive value is now $\pi_i = \widetilde{w}_i / \sum_{k=1}^N \widetilde{w}_k$. Since $\pi_i \widetilde{p}_{ij} = \pi_j \widetilde{p}_{ji}$ for every i and j, a straightforward calculation gives

$$
\sum_{i,j,k,\ell=1}^N \pi_i m_k^{ji} \left(-x_{jk} C_{k\ell} + x_{j\ell} B_{\ell k}\right) - \sum_{i,j,k,\ell=1}^N \pi_i m_k^{ji} \left(-x_{ik} C_{k\ell} + x_{i\ell} B_{\ell k}\right)
$$
$$
= \frac{1}{N} \sum_{i,\ell=1}^N \pi_i \left(-x_{ii} C_{i\ell} + x_{i\ell} B_{\ell i}\right) - \frac{1}{N} \sum_{i,j,\ell=1}^N \pi_i \widetilde{p}_{ij}^{(2)} \left(-x_{ij} C_{j\ell} + x_{i\ell} B_{\ell j}\right) \quad \text{(SI.59)}
$$

Therefore, $\rho_A > \rho_B$ for small $\delta > 0$ if and only if

$$
\sum_{i=1}^N \pi_i \sum_{\ell=1}^N \left(-x_{ii} C_{i\ell} + x_{i\ell} B_{\ell i}\right) > \sum_{i,j=1}^N \pi_i \widetilde{p}_{ij}^{(2)} \sum_{\ell=1}^N \left(-x_{ij} C_{j\ell} + x_{i\ell} B_{\ell j}\right), \quad \text{(SI.60)}
$$

48

which gives Equation 12 in the main text. To evaluate this condition, we replace $1 - x_{ij}$ by τ_{ij}, where $\tau_{ii} = 0$ for $i = 1, \ldots, N$ and, for $i \neq j$,

$$\tau_{ij} = \frac{1}{N} + \frac{1}{N} \sum_{k=1}^{N} \widetilde{p}_{ik} \tau_{kj} + \frac{1}{N} \sum_{k=1}^{N} \widetilde{p}_{jk} \tau_{ik} + \left(1 - \frac{2}{N}\right) \tau_{ij}, \qquad \text{(SI.61)}$$

which gives the recurrence $\tau_{ij} = 1/2 + (1/2) \sum_{k=1}^{N} \widetilde{p}_{ik} \tau_{kj} + (1/2) \sum_{k=1}^{N} \widetilde{p}_{jk} \tau_{ik}$. Again, we can replace τ by 2τ (since Equation 14 can be evaluated using any non-zero multiple of τ) and use $\tau_{ij} = 1 + (1/2) \sum_{k=1}^{N} \widetilde{p}_{ik} \tau_{kj} + (1/2) \sum_{k=1}^{N} \widetilde{p}_{jk} \tau_{ik}$ when $i \neq j$. We therefore obtain Equation 14 in the main text. We note that Equation 14, as well as τ, can be obtained from the corresponding results for DB updating by replacing w with $\widetilde{w}$ and p with $\widetilde{p}$.

SI.3 Specific social goods

For the three main kinds of social goods we consider, we have $B_{ij} = b\beta_{ij}$ and $C_{ij} = c\gamma_{ij}$, where

$$\left(\beta_{ij}, \gamma_{ij}\right) = \begin{cases} \left(w_{ij}, w_{ij}\right) & \text{pp,} \\ \left(w_{ij}/w_i, w_{ij}/w_i\right) & \text{ff,} \\ \left(w_{ij}, w_{ij}/w_i\right) & \text{pf.} \end{cases} \qquad \text{(SI.62)}$$

When considering the differences between fixed and proportional benefits and costs, there is one more case: fp (fixed benefits and proportional costs), i.e. $\left(\beta_{ij}, \gamma_{ij}\right) = \left(w_{ij}/w_i, w_{ij}\right)$. However, this kind of good is somewhat less natural than the others because it involves the production of a good with fixed total benefit whose cost rises with the number of recipients. A good of this form could model a situation in which the cost is tied to the transmission of the good rather than its production. For example, consider a rival, divisible good that is either costless to produce (or else is readily available to a donor, who does not need to "produce" it). Then, if there are k neighbors, each one gets b/k, where b is the benefit of the good. But if the process of transmitting or delivering the good costs c per recipient, then the total benefit is b and the total cost is kc. While our general theory can account for this kind of good, we focus primarily on the other three.

SI.3.1 PC updating

On regular graphs of degree d, all individuals have exactly d neighbors, and there is a simple correspondence between different kinds of donations. For pp-goods, a producer pays dc in total in order to donate b to each and every neighbor (for a total of db). Thus, ff-goods can be obtained from pp-goods by scaling both b and c by $1/d$. pf-goods can be obtained from pp-goods by scaling c (but not b) by $1/d$. But we can say more: from the recurrence

$\tau_{ij} = 1 + (1/2) \sum_{k=1}^{N} p_{ik}\tau_{kj} + (1/2) \sum_{k=1}^{N} p_{jk}\tau_{ik}$, we see that

$$\sum_{i,j,\ell=1}^{N} \pi_i p_{ij} p_{\ell j} \tau_{i\ell} = \sum_{i,j=1}^{N} \pi_i p_{ij} \tau_{ij} - 1. \tag{SI.63}$$

Therefore,

$$\sum_{i,j=1}^{N} \pi_i p_{ij} \sum_{\ell=1}^{N} \left(\tau_{i\ell} - \tau_{j\ell} \right) p_{\ell j} = -1. \tag{SI.64}$$

It follows that the denominator of $(b/c)^*$ in Equation 9 is negative for pp-, ff-, pf-, and fp-goods on regular graphs. Thus, regular graphs can never support producers over non-producers under PC updating.

We now turn to a couple of examples of PC updating on heterogeneous graphs:

Example 1 (Cluster of stars). On heterogeneous graphs, the results are much more interesting. Consider, for example, a cluster of stars conjoined by a complete graph at their hubs (Extended Data Figure 9). Let $i \sim j$ indicate that i and j reside on the same star, and let H and L be the set of hubs and leaves of the structure, respectively. If there are m stars in total, each of size n, then

$$\tau_{ij} = \begin{cases} \frac{m^3 + 4m^2n - 3m^2 + 4mn^2 - 6mn - 2m + 2n^2 - 9n + 8}{2m^2 + 4mn - 6m + 2n^2 - 5n + 4} & i \in H,\ j \in L,\ i \sim j, \\[2ex] \frac{2m^3 + 9m^2n - 10m^2 + 12mn^2 - 28mn + 16m + 4n^3 - 14n^2 + 16n - 8}{2m^2 + 4mn - 6m + 2n^2 - 5n + 4} & i \in H,\ j \in L,\ i \not\sim j, \\[2ex] \frac{2m^3 + 9m^2n - 13m^2 + 12mn^2 - 34mn + 25m + 4n^3 - 16n^2 + 21n - 10}{2m^2 + 4mn - 6m + 2n^2 - 5n + 4} & i \in H,\ j \in H,\ i \neq j, \\[2ex] \frac{m^3 + 4m^2n - m^2 + 4mn^2 - 2mn - 8m + 4n^2 - 14n + 12}{2m^2 + 4mn - 6m + 2n^2 - 5n + 4} & i \in L,\ j \in L,\ i \sim j,\ i \neq j, \\[2ex] \frac{2m^3 + 9m^2n - 8m^2 + 12mn^2 - 24mn + 10m + 4n^3 - 12n^2 + 11n - 4}{2m^2 + 4mn - 6m + 2n^2 - 5n + 4} & i \in L,\ j \in L,\ i \not\sim j, \\[2ex] 0 & i = j \end{cases} \tag{SI.65}$$

by Equation SI.13. These quantities give all of the times τ by symmetry.

Using Equation 9, we find that

$$
\left(\frac{b}{c}\right)^* = \begin{cases}
\dfrac{\binom{2m^2+4mn-6m}{+2n^2-5n+4}\binom{m^3+4m^2n-7m^2+4mn^2}{-14mn+13m-n^2+3n-3}}{\binom{m^4n-3m^4+5m^3n^2-21m^3n+22m^3+8m^2n^3-47m^2n^2+85m^2n}{-53m^2+4mn^4-34mn^3+88mn^2-96mn+42m-2n^4+7n^3-5n^2-n}} & \text{pp,} \\[3em]
\dfrac{\begin{pmatrix} 2m^5+13m^4n-21m^4+31m^3n^2-96m^3n+78m^3 \\ +32m^2n^3-145m^2n^2+221m^2n-119m^2+12mn^4 \\ -72mn^3+153mn^2-144mn+56m-6n^3+25n^2-32n+12 \end{pmatrix}}{\begin{pmatrix} m^5n-m^5+6m^4n^2-17m^4n+9m^4+13m^3n^3-61m^3n^2+83m^3n-29m^3 \\ +12m^2n^4-81m^2n^3+180m^2n^2-157m^2n+39m^2+4mn^5-38mn^4 \\ +120mn^3-167mn^2+103mn-18m-2n^5+11n^4-21n^3+18n^2-7n \end{pmatrix}} & \text{ff,} \\[3em]
\dfrac{\binom{2m^4+11m^3n-17m^3+20m^2n^2-57m^2n+44m^2}{+12mn^3-48mn^2+63mn-31m-6n^2+13n-6}}{\binom{m^4n-3m^4+5m^3n^2-21m^3n+22m^3+8m^2n^3-47m^2n^2+85m^2n}{-53m^2+4mn^4-34mn^3+88mn^2-96mn+42m-2n^4+7n^3-5n^2-n}} & \text{pf.}
\end{cases} \tag{SI.66}
$$

As n grows, this critical ratio approaches $(4m-1)/(2m-1)$ for pp-goods. For pf-goods and ff-goods, this ratio is asymptotic to $(6m/(2m-1))/n$ as $n \to \infty$; in particular, it approaches 0. Therefore, for any $b, c > 0$, there exists n such that producers are favored over non-producers.

Example 2 (Rich club). The "rich club" (Fotouhi et al., 2018) is a structure consisting of a well-connected group of m individuals, together with n individuals at the periphery (see Figure 4). Each peripheral individual is connected to all m members of the central rich club and nobody else. Every member of the rich club is connected to all other individuals in the population. An example of such a structure is shown in Figure 4. A straightforward calculation shows that the unique solution to Equation SI.13 with $\tau_{ii} = 0$ for $i = 1, \ldots, N$ is

$$
\tau_{ij} = \begin{cases}
\dfrac{m^3+4m^2n-m^2+4mn^2-4mn-n^2+2n-1}{m^2+3mn+n^2-n} & i \in R,\ j \in P, \\[2em]
\dfrac{m^3+4m^2n+4mn^2-mn+n-1}{m^2+3mn+n^2-n} & i \in P,\ j \in P,\ i \neq j, \\[2em]
\dfrac{m\left(m^2+4mn-m+4n^2-4n\right)}{m^2+3mn+n^2-n} & i \in R,\ j \in R,\ i \neq j, \\[2em]
0 & i = j.
\end{cases} \tag{SI.67}
$$

Again, these quantities give all of the times τ by symmetry.

By Equation 9, it follows that

$$
\left(\frac{b}{c}\right)^* = \begin{cases}
\dfrac{\left(m^2+3mn+n^2-n\right)\left(m^3+4m^2n-3m^2+4mn^2-7mn+3m-n^2+2n-1\right)}{-m^4-6m^3n+2m^3-11m^2n^2+9m^2n-m^2-5mn^3+9mn^2-4mn+n^4-3n^2+2n} & \text{pp,} \\[2em]
\dfrac{m(m+n-1)\left(\begin{smallmatrix}m^4+6m^3n-2m^3+12m^2n^2-10m^2n+m^2\\+8mn^3-12mn^2+4mn-2n^3+4n^2-2n\end{smallmatrix}\right)}{\left(\begin{smallmatrix}-m^5-6m^4n+2m^4-12m^3n^2+10m^3n-m^3-8m^2n^3\\+12m^2n^2-4m^2n+2mn^3-4mn^2+2mn+n^5-2n^4+2n^2-n\end{smallmatrix}\right)} & \text{ff,} \\[2em]
\dfrac{m^4+6m^3n-2m^3+12m^2n^2-10m^2n+m^2+8mn^3-12mn^2+4mn-2n^3+4n^2-2n}{-m^4-6m^3n+2m^3-11m^2n^2+9m^2n-m^2-5mn^3+9mn^2-4mn+n^4-3n^2+2n} & \text{pf.}
\end{cases} \tag{SI.68}
$$

As n grows, this ratio approaches $4m-1$ for pp-goods. This ratio is asymptotic to $\left(2m\left(4m-1\right)\right)/n$ for ff-goods and $\left(2\left(4m-1\right)\right)/n$ for pf-goods.

SI.3.2 DB updating

For pp-goods on a d-regular graph, $(b/c)^* = d\left(N-2\right)/\left(N-2d\right)$ (Chen, 2013, Chen et al., 2016, McAvoy and Allen, 2019), which therefore gives $(b/c)^*$ for ff-goods as well. For pf-goods, $(b/c)^* = \left(N-2\right)/\left(N-2d\right)$. For all three kinds of donation, it is clear that selection can favor producers over non-producers on a d-regular graph only if $b > c$. However, we can find heterogeneous graphs on which producers are favored over non-producers even when $b < c$.

Example 3 (Cluster of stars)**.** By Equations SI.65 and 11, under DB updating we have

$$
\left(\frac{b}{c}\right)^* = \begin{cases}
\dfrac{\left(\begin{smallmatrix}2m^5+11m^4n-21m^4+21m^3n^2-80m^3n+78m^3\\+16m^2n^3-93m^2n^2+186m^2n-127m^2+4mn^4-34mn^3\\+114mn^2-171mn+96m-4n^4+26n^3-70n^2+82n-36\end{smallmatrix}\right)}{(m-1)\left(\begin{smallmatrix}m^3n-3m^3+5m^2n^2-17m^2n+18m^2+8mn^3\\-34mn^2+58mn-37m+4n^4-22n^3+52n^2-62n+30\end{smallmatrix}\right)} & \text{pp,} \\[3em]
\dfrac{\left(\begin{smallmatrix}2m^3+6m^2n-10m^2+6mn^2\\-19mn+16m+2n^3-9n^2+14n-8\end{smallmatrix}\right)\left(\begin{smallmatrix}2m^4+11m^3n-19m^3+20m^2n^2\\-65m^2n+56m^2+12mn^3-58mn^2\\+92mn-53m-4n^3+10n^2-10n+6\end{smallmatrix}\right)}{(m-1)\left(\begin{smallmatrix}2m^2+4mn-6m\\+2n^2-5n+4\end{smallmatrix}\right)\left(\begin{smallmatrix}m^4n-m^4+6m^3n^2-16m^3n+8m^3+13m^2n^3\\-55m^2n^2+75m^2n-27m^2+12mn^4-70mn^3+155mn^2\\-150mn+48m+4n^5-30n^4+92n^3-144n^2+116n-36\end{smallmatrix}\right)} & \text{ff,} \\[3em]
\dfrac{\left(\begin{smallmatrix}2m^4+11m^3n-19m^3+20m^2n^2-65m^2n+56m^2\\+12mn^3-58mn^2+92mn-53m-4n^3+10n^2-10n+6\end{smallmatrix}\right)}{(m-1)\left(\begin{smallmatrix}m^3n-3m^3+5m^2n^2-17m^2n+18m^2+8mn^3\\-34mn^2+58mn-37m+4n^4-22n^3+52n^2-62n+30\end{smallmatrix}\right)} & \text{pf.}
\end{cases} \tag{SI.69}
$$

As $n \to \infty$, this ratio approaches 1 under pp. For both pf-goods and ff-goods, this ratio is asymptotic to $\left(\left(3m-1\right)/\left(m-1\right)\right)/n$ as $n \to \infty$. Thus, for any $m > 1$, there exists n for which this population structure can support producers over non-producers even when $0 < b < c$.

Example 4 (Rich club). For the rich club evolving according to DB updating,

$$
\left(\frac{b}{c}\right)^* = \begin{cases}
\dfrac{\left(\begin{array}{c} -m^5-7m^4n+4m^4-17m^3n^2+23m^3n-5m^3-16m^2n^3+40m^2n^2 \\ -23m^2n+2m^2-4mn^4+21mn^3-29mn^2+12mn+4n^4-9n^3+6n^2-n \end{array}\right)}{(m-1)(m+3n-2)(m^2+3mn+n^2-n)} & \text{pp,} \\[20pt]
\dfrac{m(m+n-1)\left(\begin{array}{c} -m^4-6m^3n+3m^3-12m^2n^2+15m^2n-2m^2 \\ -8mn^3+19mn^2-8mn+4n^3-7n^2+3n \end{array}\right)}{(m-1)\left(\begin{array}{c} m^4+6m^3n-2m^3+11m^2n^2-8m^2n \\ +5mn^3-6mn^2+mn-n^4+n^3+n^2-n \end{array}\right)} & \text{ff,} \\[20pt]
\dfrac{\left(\begin{array}{c} -m^4-6m^3n+3m^3-12m^2n^2+15m^2n-2m^2 \\ -8mn^3+19mn^2-8mn+4n^3-7n^2+3n \end{array}\right)}{(m-1)(m+3n-2)(m^2+3mn+n^2-n)} & \text{pf.}
\end{cases}
\tag{SI.70}
$$

For fixed $m > 1$, we see that

$$
\lim_{n\to\infty} \left(\frac{b}{c}\right)^* = \begin{cases}
-\infty & \text{pp,} \\[8pt]
\dfrac{4m(2m-1)}{m-1} & \text{ff,} \\[8pt]
-\dfrac{4(2m-1)}{3(m-1)} & \text{pf.}
\end{cases}
\tag{SI.71}
$$

SI.3.3 On the necessity of $b > c$ for pp-goods

When $0 < b \leqslant c$, pp-goods have the property that no producer can have a payoff of more than 0. Since every non-producer has a payoff of at least 0, it follows that every non-producer has at least the payoff of the best-performing producer in the population. Informally speaking, when higher payoffs result in more reproductive success, it should follow that producers of pp-goods cannot be favored over non-producers when $0 < b \leqslant c$. Here, we make this claim more formally.

Recall that the expected change RV-weighted abundance of A due to selection is

$$
\begin{aligned}
\widehat{\Delta}_{\text{sel}}(\mathbf{x}) &= \sum_{i,j=1}^{N} \pi_i \left(x_j - x_i\right) e_{ji}(\mathbf{x}) \\
&= \sum_{i,j=1}^{N} \pi_i \left((1-x_i)x_j - x_i(1-x_j)\right) e_{ji}(\mathbf{x}) \\
&= \sum_{i=1}^{N} \pi_i \left[(1-x_i)\sum_{j=1}^{N} x_j e_{ji}(\mathbf{x}) - x_i \sum_{j=1}^{N}(1-x_j)e_{ji}(\mathbf{x})\right].
\end{aligned}
\tag{SI.72}
$$

The term $\sum_{j=1}^{N} x_j e_{ji}(\mathbf{x})$ is the probability that a producer replaces i, while $\sum_{j=1}^{N}(1-x_j)e_{ji}(\mathbf{x})$ is the probability that a non-producer replaces i. When i is a non-producer the former should be reduced by selection; when i is a producer, the latter should be increased by selection. This property, of course, is not guaranteed to hold for *any* replacement rule, but it is reasonable to expect it to hold in models for which higher payoffs result in greater competitive abilities.

For example, this property is easily seen to hold for PC, DB, and IM updating for pp-goods when $b \leqslant c$. We do not include all of the details here, but it is straightforward to verify this claim using Equations SI.45, SI.51, and SI.57. Since then $\widehat{\Delta}'_{\text{sel}}(\mathbf{x}) \leqslant 0$ for every $\mathbf{x} \in \{0, 1\}^N$, producers cannot be favored over non-producers; such behavior would require $b > c$.

The same argument works for fp-goods as well. However, we have already seen numerous examples of when selection favors producers of ff- and pf-goods even when $0 < b \leqslant c$. The reason the argument presented here does not apply to ff- and pf-goods is that there can still be states in which some producers have higher payoffs than non-producers when $0 < b \leqslant c$.

SI.3.4 Monomorphic states and prosocial inequality

Let $\mathbf{A}$ and $\mathbf{B}$ denote the all-A and all-B states, respectively. As a simple example of what the monomorphic states look like in a heterogeneous population, consider the star graph (Figure 4 or Extended Data Figure 9 with $m = 1$). For all three kinds of donation, $u_{\text{hub}}(\mathbf{B}) = u_{\text{leaf}}(\mathbf{B}) = 0$. We look at the other monomorphic state, $\mathbf{A}$, separately:

For pp-goods, we have $u_{\text{hub}}(\mathbf{A}) = (N - 1)(b - c)$ and $u_{\text{leaf}}(\mathbf{A}) = b - c$, which gives

$$\overline{u}(\mathbf{A}) = 2\left(1 - \frac{1}{N}\right)(b - c); \tag{SI.73a}$$

$$u_{\text{hub}}(\mathbf{A}) - u_{\text{leaf}}(\mathbf{A}) = (N - 2)(b - c); \tag{SI.73b}$$

$$\max\{u_{\text{hub}}(\mathbf{A}), u_{\text{leaf}}(\mathbf{A})\} = \begin{cases} (N - 1)(b - c) & b \geqslant c, \\ b - c & b < c; \end{cases} \tag{SI.73c}$$

$$\min\{u_{\text{hub}}(\mathbf{A}), u_{\text{leaf}}(\mathbf{A})\} = \begin{cases} b - c & b \geqslant c, \\ (N - 1)(b - c) & b < c. \end{cases} \tag{SI.73d}$$

For ff-goods, $u_{\text{hub}}(\mathbf{A}) = (N - 1)b - c$ and $u_{\text{leaf}}(\mathbf{A}) = b/(N - 1) - c$, giving

$$\overline{u}(\mathbf{A}) = b - c; \tag{SI.74a}$$

$$u_{\text{hub}}(\mathbf{A}) - u_{\text{leaf}}(\mathbf{A}) = \frac{N(N - 2)}{N - 1}b; \tag{SI.74b}$$

$$\max\{u_{\text{hub}}(\mathbf{A}), u_{\text{leaf}}(\mathbf{A})\} = (N - 1)b - c; \tag{SI.74c}$$

$$\min\{u_{\text{hub}}(\mathbf{A}), u_{\text{leaf}}(\mathbf{A})\} = \frac{1}{N - 1}b - c. \tag{SI.74d}$$

Finally, for pf-goods, $u_{\text{hub}}(\mathbf{A}) = (N-1)b - c$ and $u_{\text{leaf}}(\mathbf{A}) = b - c$, so we have

$$\overline{u}(\mathbf{A}) = 2\left(1 - \frac{1}{N}\right)b - c; \tag{SI.75a}$$

$$u_{\text{hub}}(\mathbf{A}) - u_{\text{leaf}}(\mathbf{A}) = (N-2)b; \tag{SI.75b}$$

$$\max\{u_{\text{hub}}(\mathbf{A}), u_{\text{leaf}}(\mathbf{A})\} = (N-1)b - c; \tag{SI.75c}$$

$$\min\{u_{\text{hub}}(\mathbf{A}), u_{\text{leaf}}(\mathbf{A})\} = b - c. \tag{SI.75d}$$

SI.3.5 Accumulated versus averaged payoffs

For each kind of social good considered thus far, we have used accumulation to determine overall payoff in a population. An alternative method, which is popular for pp-goods (Maciejewski et al., 2014, Allen et al., 2017, Szabó and Fáth, 2007, Tomassini et al., 2007, Antonioni and Tomassini, 2012, Allen et al., 2013a, Wu et al., 2013b), involves averaging these payoffs instead of adding them. In other words, if, under accumulated payoffs, B_{ij} is the benefit to j due to the donation of i and C_{ij} is the corresponding cost (to i), then, under averaged payoffs, these terms undergo the transformation

$$\overline{B}_{ij} = \frac{B_{ij}}{w_j}; \tag{SI.76a}$$

$$\overline{C}_{ij} = \frac{C_{ij}}{w_i}. \tag{SI.76b}$$

Using averaged payoffs, we obtain $\rho'_A > \rho'_B$ under PC updating if and only if $\gamma b > \beta c$, where

$$\beta \; ; \; \gamma = \begin{cases} \sum_{i,j=1}^{N} \pi_i p_{ij} \tau_{ij} \; ; \; -1 & \text{pp,} \\[2ex] \sum_{i,j=1}^{N} \pi_i p_{ij} \tau_{ij} \frac{1}{w_j} \; ; \; \sum_{i,j,\ell=1}^{N} \pi_i p_{ij} \left(\tau_{i\ell} - \tau_{j\ell}\right) \frac{p_{j\ell}}{w_\ell} & \text{ff,} \\[2ex] \sum_{i,j=1}^{N} \pi_i p_{ij} \tau_{ij} \frac{1}{w_j} \; ; \; -1 & \text{pf.} \end{cases} \tag{SI.77}$$

For both pp-goods and pf-goods with payoff averaging, we have $\gamma < 0$, which means that producers can never be favored over non-producers in the limit of weak selection. In contrast, for ff-goods with averaging, there do exist structures on which $\gamma > 0$. For example, on the star of size N, we have $(b/c)^* = \left(3N^2 - 4N\right)/\left(N^2 - 8N + 8\right)$, which converges to 3 as $N \to \infty$. Coincidentally, this ratio is the same as that of pp-goods with accumulated payoffs on the star (c.f. Equation SI.66). In general, however, the conditions for $\rho_A > \rho_B$ are distinct in these two cases (pp-goods with accumulation and ff-goods with averaging).

SI.3.6 Additional contribution to a public pool

Consider ff-goods on a star. Suppose that for some $\theta \in [0, 1]$, each producer contributes $(1-\theta)b$ to their neighborhood and θb to a common pool. The overall cost is c, and the common

pool gets divided among all members of the population. This scenario is depicted in Extended Data Figure 14. If θ is the contribution rate to the common pool, then, under PC updating, selection favors producers over non-producers on a star of size N when

$$\frac{b}{c} > \left(\frac{b}{c}\right)^*_\theta := \frac{1}{1-\theta}\frac{6N^2 - 14N + 8}{N^3 - 6N^2 + 6N}. \tag{SI.78}$$

In the all-non-producer state, everyone has a payoff of 0. In the all-producer state, the payoff to the hub is $(N-1)(1-\theta)b - c + \theta b$ and the payoff each leaf is $(1-\theta)b/(N-1) - c + \theta b$. The payoff to each leaf individual is at least zero if and only if $b > c$ and

$$\theta > \theta^* = \frac{(N-1)c - b}{b(N-2)}. \tag{SI.79}$$

Note that $0 < \theta^* < 1$ if and only if $c < b < (N-1)c$. At the contribution level $\theta = \theta^*$, the hub gets $N(b-c)$ and the leaves all get 0 in the all-producer state.

For example, suppose that $b = 2$ and $c = 1$. Then,

$$\theta^* = \frac{N-3}{2(N-2)}, \tag{SI.80}$$

which approaches $c/b = 1/2$ as $N \to \infty$. Therefore, everyone is at least as well off as in the all-non-producer state, and moreover producers can evolve whenever N is not too small, since

$$\frac{b}{c} = 2 > \left(\frac{b}{c}\right)^* = \frac{2(N-2)}{N-1}\frac{6N^2 - 14N + 8}{N^3 - 6N^2 + 6N} \asymp \frac{12}{N}. \tag{SI.81}$$

Note that for pp-goods without contribution to a common pool, we require $b/c > 3$ on the star.

SI.4   Reciprocity

We now consider more complex behavioral types as outlined in the model of reciprocity described in the main text. Let A denote the strategy "tit-for-tat" (TFT) and let B denote the strategy "always defect" (ALLD). The payoff to player i in the tth round of the game is

$$u_i^t(\mathbf{x}) = \begin{cases} \sum_{j=1}^N \left(-x_i C_{ij} + x_j B_{ji}\right) & t = 1, \\ \\ \sum_{j=1}^N x_i x_j \left(-C_{ij} + B_{ji}\right) & t > 1. \end{cases} \tag{SI.82}$$

If the discounting factor (continuation probability) is $\lambda \in [0, 1)$, then the overall payoff is

$$u_i(\mathbf{x}) = (1 - \lambda) \left(\sum_{j=1}^{N} \left(-x_i C_{ij} + x_j B_{ji} \right) + \sum_{t=1}^{\infty} \lambda^t \sum_{j=1}^{N} x_i x_j \left(-C_{ij} + B_{ji} \right) \right)$$

$$= (1 - \lambda) \sum_{j=1}^{N} \left(-x_i C_{ij} + x_j B_{ji} \right) + \lambda \sum_{j=1}^{N} x_i x_j \left(-C_{ij} + B_{ji} \right). \tag{SI.83}$$

The case $\lambda = 0$ recovers the standard setup of social goods, where A is a producer and B is a non-producer. Since intermediate λ can be captured by a linear combination of $\lambda = 0$ and $\lambda = 1$, we now focus on the condition for $\lambda = 1$. When $\lambda = 1$, the δ-derivative of the change in the RV-weighted abundance of A due to selection at $\delta = 0$ is

$$\widehat{\Delta}'_{\text{sel}}(\mathbf{x}) = \sum_{i,j,k=1}^{N} \pi_i m_k^{ji} \left(x_j - x_i \right) u_k(\mathbf{x})$$

$$= \sum_{i,j,k=1}^{N} \pi_i m_k^{ji} \left(x_j - x_i \right) \sum_{\ell=1}^{N} x_k x_\ell \left(-C_{k\ell} + B_{\ell k} \right)$$

$$= \sum_{i,j,k,\ell=1}^{N} \pi_i m_k^{ji} \left(-C_{k\ell} + B_{\ell k} \right) \left(x_j x_k x_\ell - x_i x_k x_\ell \right). \tag{SI.84}$$

Using the identity (Allen and McAvoy, 2019)

$$\mathbb{E}_{\text{RMC}}^{\circ} \left[x_i x_j x_k \right] = \frac{1}{2} \left(\mathbb{E}_{\text{RMC}}^{\circ} \left[x_i x_j \right] + \mathbb{E}_{\text{RMC}}^{\circ} \left[x_i x_k \right] + \mathbb{E}_{\text{RMC}}^{\circ} \left[x_j x_k \right] \right) - \frac{1}{4}, \tag{SI.85}$$

we see that

$$\mathbb{E}_{\text{RMC}}^{\circ} \left[x_j x_k x_\ell - x_i x_k x_\ell \right] = \frac{1}{2} \left(\mathbb{E}_{\text{RMC}}^{\circ} \left[x_j x_k \right] + \mathbb{E}_{\text{RMC}}^{\circ} \left[x_j x_\ell \right] \right.$$

$$\left. - \mathbb{E}_{\text{RMC}}^{\circ} \left[x_i x_k \right] - \mathbb{E}_{\text{RMC}}^{\circ} \left[x_i x_\ell \right] \right). \tag{SI.86}$$

With $x_{ij} = 2 \mathbb{E}_{\text{RMC}}^{\circ} \left[x_i x_j \right]$, it follows that

$$\mathbb{E}_{\text{RMC}}^{\circ} \left[\widehat{\Delta}'_{\text{sel}} \right] = \frac{1}{4} \sum_{i,j,k,\ell=1}^{N} \pi_i m_k^{ji} \left(-C_{k\ell} + B_{\ell k} \right) \left(x_{jk} + x_{j\ell} - x_{ik} - x_{i\ell} \right). \tag{SI.87}$$

Therefore, for $\lambda = 1$, we see that $\mathbb{E}_{\text{RMC}}^{\circ} \left[\widehat{\Delta}'_{\text{sel}} \right] > 0$ if and only if

$$\sum_{i,j,k,\ell=1}^{N} \pi_i m_k^{ji} \left(-C_{k\ell} + B_{\ell k} \right) \left(x_{jk} + x_{j\ell} \right) > \sum_{i,j,k,\ell=1}^{N} \pi_i m_k^{ji} \left(-C_{k\ell} + B_{\ell k} \right) \left(x_{ik} + x_{i\ell} \right). \tag{SI.88}$$

Consequently, for any $\lambda \in [0,1]$, we have $\mathbb{E}^{\circ}_{\mathrm{RMC}}\left[\widehat{\Delta}'_{\mathrm{sel}}\right] > 0$ (and thus $\rho_A > \rho_B$) if and only if

$$\sum_{i,j,k,\ell=1}^{N} \pi_i m_k^{ji} \left(- \left(x_{jk} + \lambda x_{j\ell} \right) C_{k\ell} + \left(x_{j\ell} + \lambda x_{jk} \right) B_{\ell k} \right)$$

$$> \sum_{i,j,k,\ell=1}^{N} \pi_i m_k^{ji} \left(- \left(x_{ik} + \lambda x_{i\ell} \right) C_{k\ell} + \left(x_{i\ell} + \lambda x_{ik} \right) B_{\ell k} \right), \tag{SI.89}$$

which is Equation 2 in the main text.

We now turn to specific update rules. In each case, the terms π_i, m_k^{ji}, x_{ij}, and τ_{ij} are exactly the same as they were when there was no reciprocity involved (i.e. $\lambda = 0$). Therefore, in each case we take these quantities to be the same as they were previously and instead focus on the selection condition.

SI.4.1   PC updating

Using the approach in §SI.2.1, for PC updating we find that $\rho'_A > \rho'_B$ if and only if

$$\sum_{i=1}^{N} \pi_i \sum_{\ell=1}^{N} \left(- \left(x_{ii} + \lambda x_{i\ell} \right) C_{i\ell} + \left(x_{i\ell} + \lambda x_{ii} \right) B_{\ell i} \right)$$

$$> \sum_{i,j=1}^{N} \pi_i p_{ij} \sum_{\ell=1}^{N} \left(- \left(x_{ij} + \lambda x_{i\ell} \right) C_{j\ell} + \left(x_{i\ell} + \lambda x_{ij} \right) B_{\ell j} \right). \tag{SI.90}$$

SI.4.2   DB updating

Using the approach in §SI.2.2, for DB updating we find that $\rho'_A > \rho'_B$ if and only if

$$\sum_{i,\ell=1}^{N} \pi_i \left(- \left(x_{ii} + \lambda x_{i\ell} \right) C_{i\ell} + \left(x_{i\ell} + \lambda x_{ii} \right) B_{\ell i} \right)$$

$$> \sum_{i,j,\ell=1}^{N} \pi_i p_{ij}^{(2)} \left(- \left(x_{ij} + \lambda x_{i\ell} \right) C_{j\ell} + \left(x_{i\ell} + \lambda x_{ij} \right) B_{\ell j} \right). \tag{SI.91}$$

SI.4.3   IM updating

Using the approach in §SI.2.3, for IM updating we find that $\rho'_A > \rho'_B$ if and only if

$$\sum_{i,\ell=1}^{N} \pi_i \left(- \left(x_{ii} + \lambda x_{i\ell} \right) C_{i\ell} + \left(x_{i\ell} + \lambda x_{ii} \right) B_{\ell i} \right)$$

$$> \sum_{i,j,\ell=1}^{N} \pi_i \widetilde{p}_{ij}^{(2)} \left(- \left(x_{ij} + \lambda x_{i\ell} \right) C_{j\ell} + \left(x_{i\ell} + \lambda x_{ij} \right) B_{\ell j} \right). \tag{SI.92}$$

SI.4.4 Undiscounted games ($\lambda = 1$)

When $\lambda = 1$, the conditions stated above for PC, DB, and IM updating each have the form

$$\sum_{i,\ell=1}^{N} \pi_i \left(-\left(x_{ii} + x_{i\ell} \right) C_{i\ell} + \left(x_{i\ell} + x_{ii} \right) B_{\ell i} \right)$$
$$> \sum_{i,j,\ell=1}^{N} \pi_i P_{ij} \left(-\left(x_{ij} + x_{i\ell} \right) C_{j\ell} + \left(x_{i\ell} + x_{ij} \right) B_{\ell j} \right) \tag{SI.93}$$

for some $N \times N$ stochastic matrix $\left(P_{ij} \right)_{i,j=1}^{N}$ with $\pi_i P_{ij} = \pi_j P_{ji}$ for every i and j. Replacing $1 - x_{ij}$ by τ_{ij} and rearranging gives an equivalent condition for $\rho_A > \rho_B$, namely

$$\sum_{i,j,\ell=1}^{N} \pi_i P_{ij} \left(\tau_{ij} + \tau_{j\ell} - \tau_{i\ell} \right) \left(B_{\ell i} - C_{i\ell} \right) > 0. \tag{SI.94}$$

Lemma 2. If $N \geqslant 3$, then $\sum_{j=1}^{N} \pi_i P_{ij} \left(\tau_{ij} + \tau_{j\ell} - \tau_{i\ell} \right) > 0$ for every i and ℓ.

Proof. We first show that $1 - x_{ij} - x_{j\ell} + x_{i\ell} \geqslant 0$. Note that

$$1 - x_{ij} - x_{j\ell} + x_{i\ell} = \mathbb{E}_{\text{RMC}}^{\circ} \begin{bmatrix} x_i\left(1-x_j\right) + \left(1-x_i\right)x_j \\ -x_j x_\ell - \left(1-x_j\right)\left(1-x_\ell\right) \\ +x_i x_\ell + \left(1-x_i\right)\left(1-x_\ell\right) \end{bmatrix}$$
$$= 2\mathbb{E}_{\text{RMC}}^{\circ} \left[x_j - x_i x_j + x_i x_\ell - x_j x_\ell \right]. \tag{SI.95}$$

The term $x_j - x_i x_j + x_i x_\ell - x_j x_\ell$ is 1 if and only if $x_i = x_\ell$ and $x_j \neq x_i, x_\ell$; otherwise, it is 0. Therefore, its expectation (with respect to any distribution) is non-negative. Furthermore, for any j, the state $\mathbf{e}_j$ with $\left(\mathbf{e}_j \right)_j = 1$ and $\left(\mathbf{e}_j \right)_k = 0$ for $k \neq j$ satisfies $\pi_{\text{RMC}}^{\circ} \left(\mathbf{e}_j \right) > 0$ since it arises from state $\mathbf{B}$ after a mutation to A in location j. Thus, $\tau_{ij} + \tau_{j\ell} > \tau_{i\ell}$ whenever $j \neq i, \ell$.

Now, fix i and ℓ, and let $j \neq i, \ell$ be such that $P_{ij} > 0$. Since P_{ij} is the probability of transitioning from i to j in either a one- or two-step random walk, such a choice of j is possible as long as $N \geqslant 3$. We also have $\tau_{ij} + \tau_{j\ell} > \tau_{i\ell}$, hence $\pi_i P_{ij} \left(\tau_{ij} + \tau_{j\ell} - \tau_{i\ell} \right) > 0$, as desired. $\square$

As a consequence of Lemma 2, we see that PMB goods (those with $B_{ji} \geqslant C_{ij}$ for every i and j, with strict inequality for at least one pair (i, j)) are always favored by selection when $\lambda = 1$. Furthermore, for $B_{ij} = b\beta_{ij}$ and $C_{ij} = c\gamma_{ij}$ (not necessarily PMB), the denominator of

$$\left(\frac{b}{c} \right)_{\lambda=1}^{*} = \frac{\sum_{i,j,\ell=1}^{N} \pi_i P_{ij} \left(\tau_{ij} + \tau_{j\ell} - \tau_{i\ell} \right) \gamma_{i\ell}}{\sum_{i,j,\ell=1}^{N} \pi_i P_{ij} \left(\tau_{ij} + \tau_{j\ell} - \tau_{i\ell} \right) \beta_{\ell i}} \tag{SI.96}$$

is positive (provided β is not identically zero). Thus, the critical ratio gives a lower bound on b/c for the evolution of TFT relative to ALLD (as opposed to an upper bound, which occurs when the denominator of the critical ratio is negative). In particular, for any social good with $B_{ij} = b\beta_{ij}$ and $C_{ij} = c\gamma_{ij}$ for every i and j, where β and γ are independent of b and c, producers of this social good can evolve in the undiscounted game whenever b/c is sufficiently large.

SI.5   Relationship to prior literature

Broadly speaking, our contribution is twofold: *(i)* a general theory of the evolutionary success of producers and *(ii)* applications of this theory that uncover surprising results about producers of certain social goods. Our framework and results are general in two distinct ways. First, each type (A and B) can be producers of arbitrary social goods. If i has type A, then he or she provides B_{ij} to j at a cost of C_{ij}. If i has type B, then he or she gives b_{ij} to j at a cost of c_{ij}. Second, we do not assume that evolution proceeds according to a specific update rule (e.g. PC, DB, or IM). We treat specific update rules in the examples, but our conditions for evolutionary success hold in a much more general context, namely for distributions over replacement events, $\left\{ p_{(R,\alpha)}(\mathbf{x}) \right\}_{(R,\alpha)}$.

Here, we review prior related literature for evolutionary games in structured populations and highlight how our work relates to these previous results.

SI.5.1   Pairwise games in structured populations

There is an extensive literature on 2×2 matrix games in graph-structured populations (Hauert and Doebeli, 2004, Nowak and May, 1992, Ohtsuki et al., 2006, Débarre et al., 2014, Santos and Pacheco, 2005, Gómez-Gardenes et al., 2007, Allen et al., 2017, Nowak et al., 2009, Szabó and Fáth, 2007, Blume, 1993, Tarnita et al., 2009b). Such a matrix game can be written in general form as

$$
\begin{array}{c}
\begin{array}{cc} A & B \end{array} \\
\begin{array}{c} A \\ B \end{array}
\begin{pmatrix} R & S \\ T & P \end{pmatrix},
\end{array}
\tag{SI.97}
$$

For $T > R > P > S$, this game is a prisoner's dilemma, with strategy A representing co-operation. When a player has many neighbors, as is the case in graph-structured populations, a total payoff is obtained by either accumulating or averaging the payoffs from many pairwise interactions. This kind of model corresponds to our notion of pp-goods (see also §SI.3.5).

A specific instance of the prisoner's dilemma known as the *donation game* is often used since it provides a simple description of an altruistic act: a cooperator (A) pays a cost of c to donate b to the opponent; defectors (B) pay no costs and provide no donations. In terms of Equation SI.97, the donation game satisfies $R = b - c$, $S = -c$, $T = b$, and $P = 0$. This donation game is typically easier to analyze than the general matrix game Equation SI.97. However, once conditions for $\rho_A > \rho_B$ have been obtained for the donation game, these conditions can be generalized to an arbitrary game of the form Equation SI.97 by means of the Structure Coefficient Theorem (Nowak et al., 2009, Tarnita et al., 2009b).

Whereas these studies focus on pairwise interactions described by a fixed matrix game, we consider dilemmas of social goods, in which the goods produced by an individual are distributed to his or her neighbors in a manner that depends on the population structure. If i and j are

neighbors, then, instead of their interaction being described by a symmetric payoff matrix

$$
\begin{array}{cc}
& \begin{array}{cc} A & \qquad\quad B \end{array} \\
\begin{array}{c} A \\ B \end{array} & \left(\begin{array}{cc} b - c,\ b - c & -c,\ b \\ b,\ -c & 0,\ 0 \end{array} \right),
\end{array}
\tag{SI.98}
$$

in a social goods dilemma it is described by an asymmetric payoff matrix,

$$
M^{ij} =
\begin{array}{cc}
& \begin{array}{cc} A & \qquad\qquad B \end{array} \\
\begin{array}{c} A \\ B \end{array} & \left(\begin{array}{cc} B_{ji} - C_{ij},\ B_{ij} - C_{ji} & -C_{ij},\ B_{ij} \\ B_{ji},\ -C_{ji} & 0,\ 0 \end{array} \right).
\end{array}
\tag{SI.99}
$$

In other words, type A (producer) at location i pays C_{ij} to donate B_{ij} to j. Type B (non-producer) at this same location pays nothing and gives nothing. Our theory applies to more general bimatrix games (Ohtsuki, 2010, Sekiguchi and Ohtsuki, 2015, McAvoy and Hauert, 2015a,b), although we are particularly focused on asymmetric games arising from the production of social goods.

Below, we give a brief summary of work on symmetric games in structured populations.

SI.5.1.1 Regular and homogeneous graphs

A regular graph is one in which all individuals share the same number of neighbors, i.e. $w_i = d$ for every $i = 1, \ldots, N$ and some fixed d. Using the pair-approximation method on regular graphs, Ohtsuki et al. (2006) showed that weak selection favors the evolution of cooperation if $b/c > d$. Moreover, the difference between accumulated and averaged payoffs on a regular graph amounts to a simple rescaling of both b and c, so this critical threshold, namely $(b/c)^* = d$, is the same for both methods of obtaining a net payoff from many interactions. This result has also been refined to capture finite-population effects on weighted vertex-transitive graphs (Taylor et al., 2007, Débarre et al., 2014, Allen and Nowak, 2014) and unweighted regular graphs (Chen, 2013).

SI.5.1.2 Heterogeneous graphs

Although there are many simulation-based investigations of pairwise games on heterogeneous graphs Abramson and Kuperman, 2001, Santos and Pacheco, 2005, Gómez-Gardenes et al., 2007, Maciejewski et al., 2014, Szabó and Fáth, 2007, an analytic solution for weak selection was not found until the work of Allen et al. (2017). Conditions for a strategy to be favored under weak selection were found for DB and BD updating, using either averaged or accumulated payoffs, in terms of coalescence times. The properties of this solution were further explored in follow-up works by Fotouhi et al. (2018, 2019).

The results of Allen et al. (2017), in the case of the donation game with averaged payoffs, can be recovered from our results by setting $B_{ij} = bw_{ij}/w_j$ and $C_{ij} = cw_{ij}/w_i$ for each i and j. Equation SI.13 is equivalent, under DB or BD updating, to the systems of equations for coalescence times used by Allen et al. (2017).

SI.5.2 Public goods games in structured populations

Numerous works (Peña et al., 2016a, Santos et al., 2008, Rong and Wu, 2009, Szolnoki and Perc, 2011, Perc et al., 2013, Peña et al., 2016b) have considered public goods games on graphs. In additional to classical public goods, one may also consider public goods that travel through the network by diffusion. This framework can be used to model the production of diffusible chemical goods in microbes (Le Gac and Doebeli, 2010, Julou et al., 2013, Kümmerli et al., 2014). Conditions for the production of such goods on weighted vertex-transitive graphs were obtained by Allen et al. (2013b).

Since our focus is on social goods donated to immediate neighbors with no self-donation, we do not give a complete analysis of public goods games. However, there is one study of public goods games, carried out by Santos et al. (2008), that is particularly relevant to our model since it involves a numerical comparison of fixed- and proportional-cost goods in multiplayer interactions.

Santos et al. (2008) considered a model of public goods games in heterogeneous populations in which every individual is involved in multiple games. Each individual initiates an interaction with neighbors, and everyone involved benefits from the public good. An individual with k neighbors therefore participates in $k+1$ interactions, one initiated by themselves and k initiated by neighbors. In their model, a cooperator contributes either c per game or c in total (i.e. $c/(k+1)$ per game); non-producers do nothing. The total contribution is then multiplied by an enhancement factor, r, and divided among all individuals involved in the game.

If γ_i denotes the per-game contribution when i cooperates, this model satisfies

$$B_{ij} = \sum_{k=1}^{N} \widetilde{w}_{ik}\widetilde{w}_{kj}\left(r\frac{\gamma_i}{\widetilde{w}_k}\right); \tag{SI.100a}$$

$$C_{ij} = \sum_{k=1}^{N} \widetilde{w}_{ik}\widetilde{w}_{kj}\left(\frac{\gamma_i}{\widetilde{w}_k}\right), \tag{SI.100b}$$

where, again, $\widetilde{w}_{ii} = 1$ for $i = 1, \ldots, N$ and $\widetilde{w}_{ij} = w_{ij}$ if $i \neq j$. Therefore, this model can be seen as a special case of ours.

For both PC and DB updating, let $\tau_{ii} = 0$ for $i = 1, \ldots, N$ and $\tau_{ij} = 1 + (1/2)\sum_{k=1}^{N} p_{ik}\tau_{kj} + (1/2)\sum_{k=1}^{N} p_{jk}\tau_{ik}$ for $i \neq j$. Under PC updating, for example, we have

$$\rho'_A > \rho'_B \iff r\sum_{i,j=1}^{N}\pi_i p_{ij}\sum_{\ell=1}^{N} x_{i\ell}\left(C_{\ell i} - C_{\ell j}\right) > \sum_{i,j=1}^{N}\pi_i p_{ij}\sum_{\ell=1}^{N}\left(x_{ii}C_{i\ell} - x_{ij}C_{j\ell}\right)$$

$$\iff r\sum_{i,j=1}^{N}\pi_i p_{ij}\sum_{\ell=1}^{N}\tau_{i\ell}\left(C_{\ell j} - C_{\ell i}\right) > \sum_{i,j=1}^{N}\pi_i p_{ij}\sum_{\ell=1}^{N}\tau_{ij}C_{j\ell}. \tag{SI.101}$$

Under DB updating,

$$\rho'_A > \rho'_B \iff r \sum_{i,j=1}^{N} \pi_i p_{ij}^{(2)} \sum_{\ell=1}^{N} x_{i\ell} \left(C_{\ell i} - C_{\ell j}\right) > \sum_{i,j=1}^{N} \pi_i p_{ij}^{(2)} \sum_{\ell=1}^{N} \left(x_{ii} C_{i\ell} - x_{ij} C_{j\ell}\right)$$

$$\iff r \sum_{i,j=1}^{N} \pi_i p_{ij}^{(2)} \sum_{\ell=1}^{N} \tau_{i\ell} \left(C_{\ell j} - C_{\ell i}\right) > \sum_{i,j=1}^{N} \pi_i p_{ij}^{(2)} \sum_{\ell=1}^{N} \tau_{ij} C_{j\ell}. \qquad \text{(SI.102)}$$

Finally, we note that there is nearly always wealth inequality in heterogeneous populations, even for pp-goods with accumulated payoffs. Santos et al. (2008) observed that the nature of how players participate in public goods games in heterogeneous populations can amplify inequality. Our notion of "prosocial inequality" is somewhat different, however. Whereas inequality can be present even when selection improves the payoffs of all players in the population, we show that selection can actually decrease the payoff of some while improving the payoff of others. Thus, prosocial inequality is not a measure of relative inequality within the all-A state; it is a measure of inequality in the all-A state as compared to the all-B state.

SI.5.3   Deterministic versus stochastic payoffs

One final way in which our model and results differ from those of prior studies is that we allow for stochastic payoffs in addition to deterministic payoffs. In fact, we show in §SI.1.5 that a model with stochastic payoffs can be replaced by one with deterministic payoffs under weak selection. This result is particularly relevant for goods with concentrated, stochastic benefits (Figure 15), in which case the critical threshold for producers to evolve is identical to that of ff-goods.

References

K. Sigmund. *The calculus of selfishness*. Princeton University Press, 2010.

A. L. Radzvilavicius, A. J. Stewart, and J. B. Plotkin. Evolution of empathetic moral evaluation. *eLife*, 8, Apr 2019. doi: 10.7554/elife.44269.

R. Axelrod. *The Evolution of Cooperation*. Basic Books, 1984.

G. Szabó and C. Tőke. Evolutionary prisoner's dilemma game on a square lattice. *Physical Review E*, 58(1):69–73, Jul 1998. doi: 10.1103/physreve.58.69.

G. Abramson and M. Kuperman. Social games in a social network. *Physical Review E*, 63(3): 030901, 2001.

M. Broom and J. Rychtář. *Game-Theoretical Models in Biology*. Taylor & Francis, 2013.

J. Maynard Smith. *Evolution and the Theory of Games*. Cambridge University Press, 1982. doi: 10.1017/cbo9780511806292.

C. Hauert and M. Doebeli. Spatial structure often inhibits the evolution of cooperation in the snowdrift game. *Nature*, 428(6983):643–646, Apr 2004. doi: 10.1038/nature02360.

M. Doebeli, C. Hauert, and T. Killingback. The Evolutionary Origin of Cooperators and Defectors. *Science*, 306(5697):859–862, Oct 2004. doi: 10.1126/science.1101456.

W. F. Lloyd. *Two Lectures on the Checks to Population*. Oxford University Press, 1833.

G. Hardin. The Tragedy of the Commons. *Science*, 162(3859):1243–1248, Dec 1968. doi: 10.1126/science.162.3859.1243.

G. Szabó and C. Hauert. Phase transitions and volunteering in spatial public goods games. *Physical Review Letters*, 89(11), Aug 2002. doi: 10.1103/physrevlett.89.118101.

F. L. Pinheiro, V. V. Vasconcelos, F. C. Santos, and J. M. Pacheco. Evolution of All-or-None Strategies in Repeated Public Goods Dilemmas. *PLoS Computational Biology*, 10(11): e1003945, Nov 2014. doi: 10.1371/journal.pcbi.1003945.

J Peña, B. Wu, J. Arranz, and A. Traulsen. Evolutionary Games of Multiplayer Cooperation on Graphs. *PLOS Computational Biology*, 12(8):e1005059, Aug 2016a. doi: 10.1371/journal.pcbi.1005059.

L.-X. Zhong, W.-J. Xu, Y.-X. He, C.-Y. Zhong, R.-D. Chen, T. Qiu, Y.-D. Shi, and F. Ren. A generalized public goods game with coupling of individual ability and project benefit. *Chaos, Solitons & Fractals*, 101:73–80, Aug 2017. doi: 10.1016/j.chaos.2017.05.025.

M. A. Nowak and R. M. May. Evolutionary games and spatial chaos. *Nature*, 359(6398):826–829, Oct 1992. doi: 10.1038/359826a0.

M. Nakamaru, H. Matsuda, and Y. Iwasa. The Evolution of Cooperation in a Lattice-Structured Population. *Journal of Theoretical Biology*, 184(1):65–81, Jan 1997. doi: 10.1006/jtbi.1996.0243.

E. Lieberman, C. Hauert, and M. A. Nowak. Evolutionary dynamics on graphs. *Nature*, 433 (7023):312–316, Jan 2005. doi: 10.1038/nature03204.

H. Ohtsuki, C. Hauert, E. Lieberman, and M. A. Nowak. A simple rule for the evolution of cooperation on graphs and social networks. *Nature*, 441(7092):502–505, May 2006. doi: 10.1038/nature04605.

P. D. Taylor, T. Day, and G. Wild. Evolution of cooperation in a finite homogeneous graph. *Nature*, 447(7143):469–472, May 2007. doi: 10.1038/nature05784.

Y.-T. Chen. Sharp benefit-to-cost rules for the evolution of cooperation on regular graphs. *The Annals of Applied Probability*, 23(2):637–664, Apr 2013. doi: 10.1214/12-aap849.

F. Débarre, C. Hauert, and M. Doebeli. Social evolution in structured populations. *Nature Communications*, 5, Mar 2014. doi: 10.1038/ncomms4409.

F. C. Santos and J. M. Pacheco. Scale-free networks provide a unifying framework for the emergence of cooperation. *Physical Review Letters*, 95(9):098104, 2005.

T. Antal, S. Redner, and V. Sood. Evolutionary Dynamics on Degree-Heterogeneous Graphs. *Physical Review Letters*, 96(18), May 2006. doi: 10.1103/physrevlett.96.188104.

J. Gómez-Gardenes, M. Campillo, L. M. Floría, and Y. Moreno. Dynamical organization of cooperation in complex topologies. *Physical Review Letters*, 98(10):108103, 2007.

V. Sood, T. Antal, and S. Redner. Voter models on heterogeneous networks. *Physical Review E*, 77(4), Apr 2008. doi: 10.1103/physreve.77.041121.

X.-B. Cao, W.-B. Du, and Z.-H. Rong. The evolutionary public goods game on scale-free networks with heterogeneous investment. *Physica A: Statistical Mechanics and its Applications*, 389(6):1273–1280, Mar 2010. doi: 10.1016/j.physa.2009.11.044.

W. Maciejewski, F. Fu, and C. Hauert. Evolutionary game dynamics in populations with heterogenous structures. *PLoS Computational Biology*, 10(4):e1003567, Apr 2014. doi: 10.1371/journal.pcbi.1003567.

R. Fan, Y. Zhang, M. Luo, and H. Zhang. Promotion of cooperation induced by heterogeneity of both investment and payoff allocation in spatial public goods game. *Physica A: Statistical Mechanics and its Applications*, 465:454–463, Jan 2017. doi: 10.1016/j.physa.2016.08.039.

B. Allen, G. Lippner, Y.-T. Chen, B. Fotouhi, N. Momeni, S.-T. Yau, and M. A. Nowak. Evolutionary dynamics on any population structure. *Nature*, 544(7649):227–230, Mar 2017. doi: 10.1038/nature21723.

A. Goldfarb and C. Tucker. Digital Economics. *Journal of Economic Literature*, 57(1):3–43, Mar 2019. doi: 10.1257/jel.20171452.

F. C. Santos, M. D. Santos, and J. M. Pacheco. Social diversity promotes the emergence of cooperation in public goods games. *Nature*, 454(7201):213–216, Jul 2008. doi: 10.1038/nature06940.

C. Li, B. Zhang, R. Cressman, and Y. Tao. Evolution of Cooperation in a Heterogeneous Graph: Fixation Probabilities under Weak Selection. *PLoS ONE*, 8(6):e66560, Jun 2013. doi: 10.1371/journal.pone.0066560.

A. Stutzer, L. Goette, and M. Zehnder. Active Decisions and Prosocial Behaviour: A Field Experiment on Blood Donation. *The Economic Journal*, 121(556):F476–F493, Oct 2011. doi: 10.1111/j.1468-0297.2011.02477.x.

R. I. M. Dunbar. Functional Significance of Social Grooming in Primates. *Folia Primatologica*, 57(3):121–131, 1991. doi: 10.1159/000156574.

L. Horn, C. Scheer, T. Bugnyar, and J. J. M. Massen. Proactive prosociality in a cooperatively breeding corvid, the azure-winged magpie (Cyanopica cyana). *Biology Letters*, 12(10): 20160649, Oct 2016. doi: 10.1098/rsbl.2016.0649.

G. S. Wilkinson. Reciprocal food sharing in the vampire bat. *Nature*, 308(5955):181–184, Mar 1984. doi: 10.1038/308181a0.

D. Fudenberg and L. A. Imhof. Imitation processes with small mutations. *Journal of Economic Theory*, 131(1):251–262, Nov 2006. doi: 10.1016/j.jet.2005.04.006.

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

W. Maciejewski. Reproductive value in graph-structured populations. *Journal of Theoretical Biology*, 340:285–293, Jan 2014. doi: 10.1016/j.jtbi.2013.09.032.

G. Wild and A. Traulsen. The different limits of weak selection and the evolutionary dynamics of finite populations. *Journal of Theoretical Biology*, 247(2):382–390, Jul 2007. doi: 10.1016/j.jtbi.2007.03.015.

F. Fu, L. Wang, M. A. Nowak, and C. Hauert. Evolutionary dynamics on graphs: Efficient method for weak selection. *Physical Review E*, 79(4), Apr 2009. doi: 10.1103/physreve.79.046707.

B. Wu, P. M. Altrock, L. Wang, and A. Traulsen. Universality of weak selection. *Physical Review E*, 82(4), Oct 2010. doi: 10.1103/physreve.82.046106.

B. Wu, J. García, C. Hauert, and A. Traulsen. Extrapolating weak selection in evolutionary games. *PLoS Computational Biology*, 9(12):e1003381, Dec 2013a. doi: 10.1371/journal.pcbi.1003381.

C. Mullon and L. Lehmann. The robustness of the weak selection approximation for the evolution of altruism against strong selection. *Journal of Evolutionary Biology*, 27(10):2272–2282, Aug 2014. doi: 10.1111/jeb.12462.

Y. Iwasa, M. Nakamaru, and S. A. Levin. Allelopathy of bacteria in a lattice population: Competition between colicin-sensitive and colicin-producing strains. *Evolutionary Ecology*, 12(7):785–802, Oct 1998. doi: 10.1023/a:1006590431483.

P. Forber and R. Smead. The evolution of fairness through spite. *Proceedings of the Royal Society B: Biological Sciences*, 281(1780):20132439, Apr 2014. doi: 10.1098/rspb.2013.2439.

D. Fudenberg and L. A. Imhof. Monotone imitation dynamics in large populations. *Journal of Economic Theory*, 140(1):229–245, May 2008. doi: 10.1016/j.jet.2007.08.002.

C. P. Roca, J. A. Cuesta, and A. Sánchez. Evolutionary game theory: Temporal and spatial effects beyond replicator dynamics. *Physics of Life Reviews*, 6(4):208–249, Dec 2009. doi: 10.1016/j.plrev.2009.08.001.

T. Traulsen, D. Semmann, R. D. Sommerfeld, H.-J. Krambeck, and M. Milinski. Human strategy updating in evolutionary games. *Proceedings of the National Academy of Sciences*, 107(7):2962–2966, Feb 2010. doi: 10.1073/pnas.0912515107.

S. Zhou and R. J. Mondragon. The Rich-Club Phenomenon in the Internet Topology. *IEEE Communications Letters*, 8(3):180–182, Mar 2004. doi: 10.1109/lcomm.2004.823426.

V. Colizza, A. Flammini, M. A. Serrano, and A. Vespignani. Detecting rich-club ordering in complex networks. *Nature Physics*, 2(2):110–115, Jan 2006. doi: 10.1038/nphys209.

J. J. McAuley, L. da Fontoura Costa, and T. S. Caetano. Rich-club phenomenon across complex network hierarchies. *Applied Physics Letters*, 91(8):084103, Aug 2007. doi: 10.1063/1.2773951.

B. Fotouhi, N. Momeni, B. Allen, and M. A. Nowak. Conjoining uncooperative societies facilitates evolution of cooperation. *Nature Human Behaviour*, 2(7):492–499, Jul 2018. doi: 10.1038/s41562-018-0368-6.

C. Ansell, R. Bichir, and S. Zhou. Who Says Networks, Says Oligarchy? Oligarchies as "Rich Club" Networks. *Connections*, 35(2):20–32, 2016. doi: 10.17266/35.2.2.

Y. Dong, J. Tang, N. V. Chawla, T. Lou, Y. Yang, and B. Wang. Inferring Social Status and Rich Club Effects in Enterprise Communication Networks. *PLOS ONE*, 10(3):e0119446, Mar 2015. doi: 10.1371/journal.pone.0119446.

L. M. Vaquero and M. Cebrian. The rich club phenomenon in the classroom. *Scientific Reports*, 3(1), jan 2013. doi: 10.1038/srep01174.

A. Ma, R. J. Mondragón, and V. Latora. Anatomy of funded research in science. *Proceedings of the National Academy of Sciences*, 112(48):14760–14765, Oct 2015. doi: 10.1073/pnas.1513651112.

M. Szell and R. Sinatra. Research funding goes to rich clubs. *Proceedings of the National Academy of Sciences*, 112(48):14749–14750, Nov 2015. doi: 10.1073/pnas.1520118112.

Z.-Q. Jiang and W.-X. Zhou. Statistical significance of the rich-club phenomenon in complex networks. *New Journal of Physics*, 10(4):043002, Apr 2008. doi: 10.1088/1367-2630/10/4/043002.

M. E. J. Newman. Finding community structure in networks using the eigenvectors of matrices. *Physical Review E*, 74(3):036104, 2006. doi: 10.1103/PhysRevE.74.036104.

W. F. Laurance. Second thoughts on who goes where in author lists. *Nature*, 442(7098):26–26, Jul 2006. doi: 10.1038/442026d.

V. Venkatraman. Conventions of Scientific Authorship. *Science*, Apr 2010. doi: 10.1126/science.caredit.a1000039.

L. Bošnjak and A. Marušić. Prescribed practices of authorship: review of codes of ethics from professional bodies and journal guidelines across disciplines. *Scientometrics*, 93(3):751–763, May 2012. doi: 10.1007/s11192-012-0773-y.

M. K. McNutt, M. Bradford, J. M. Drazen, B. Hanson, B. Howard, K. H. Jamieson, V. Kiermer, E. Marcus, B. K. Pope, R. Schekman, S. Swaminathan, P. J. Stang, and I. M. Verma. Transparency in authors' contributions and responsibilities to promote integrity in scientific publication. *Proceedings of the National Academy of Sciences*, 115(11):2557–2560, Feb 2018. doi: 10.1073/pnas.1715374115.

R. L. Trivers. The Evolution of Reciprocal Altruism. *The Quarterly Review of Biology*, 46(1): 35–57, Mar 1971. doi: 10.1086/406755.

M. A. Nowak. Five rules for the evolution of cooperation. *Science*, 314(5805):1560–1563, Dec 2006. doi: 10.1126/science.1133755.

W. H. Press and F. J. Dyson. Iterated prisoner's dilemma contains strategies that dominate any evolutionary opponent. *Proceedings of the National Academy of Sciences*, 109(26):10409–10413, May 2012. doi: 10.1073/pnas.1206569109.

A. J. Stewart and J. B. Plotkin. From extortion to generosity, evolution in the Iterated Prisoner's Dilemma. *Proceedings of the National Academy of Sciences*, 110(38):15348–15353, Sep 2013. doi: 10.1073/pnas.1306246110.

A. J. Stewart and J. B. Plotkin. Small groups and long memories promote cooperation. *Scientific Reports*, 6:26889, Jun 2016. doi: 10.1038/srep26889.

C. Hilbe, K. Chatterjee, and M. A. Nowak. Partners and rivals in direct reciprocity. *Nature Human Behaviour*, 2018. doi: 10.1038/s41562-018-0320-9.

R. Ibsen-Jensen, K. Chatterjee, and M. A. Nowak. Computational complexity of ecological and evolutionary spatial dynamics. *Proceedings of the National Academy of Sciences*, page 201511366, Dec 2015. doi: 10.1073/pnas.1511366112.

C. Hadjichrysathou, M. Broom, and J. Rychtář. Evolutionary games on star graphs under various updating rules. *Dynamic Games and Applications*, 1(3):386–407, Jun 2011. doi: 10.1007/s13235-011-0022-7.

P. D. Taylor. Altruism in viscous populations — an inclusive fitness model. *Evolutionary Ecology*, 6(4):352–356, Jul 1992. doi: 10.1007/bf02270971.

D. S. Wilson, G. B. Pollock, and L. A. Dugatkin. Can altruism evolve in purely viscous populations? *Evolutionary Ecology*, 6(4):331–341, Jul 1992. doi: 10.1007/bf02270969.

J. Mitteldorf and D. S. Wilson. Population viscosity and the evolution of altruism. *Journal of Theoretical Biology*, 204(4):481–496, Jun 2000. doi: 10.1006/jtbi.2000.2007.

A. J. Irwin and P. D. Taylor. Evolution of Altruism in Stepping-Stone Populations with Overlapping Generations. *Theoretical Population Biology*, 60(4):315–325, Dec 2001. doi: 10.1006/tpbi.2001.1533.

A. Grafen. An inclusive fitness analysis of altruism on a cyclical network. *Journal of Evolutionary Biology*, 20(6):2278–2283, Nov 2007. doi: 10.1111/j.1420-9101.2007.01413.x.

S. Lion and M. van Baalen. Self-structuring in spatial evolutionary ecology. *Ecology Letters*, 11 (3):277–295, Mar 2008. doi: 10.1111/j.1461-0248.2007.01132.x.

C. E. Tarnita, T. Antal, H. Ohtsuki, and M. A. Nowak. Evolutionary dynamics in set structured populations. *Proceedings of the National Academy of Sciences*, 106(21):8601–8604, May 2009a. doi: 10.1073/pnas.0903019106.

M. A. Nowak, C. E. Tarnita, and T. Antal. Evolutionary dynamics in structured populations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1537):19–30, Nov 2009. doi: 10.1098/rstb.2009.0215.

D. G. Rand, M. A. Nowak, J. H. Fowler, and N. A. Christakis. Static network structure can stabilize human cooperation. *Proceedings of the National Academy of Sciences*, 111(48): 17093–17098, Nov 2014. doi: 10.1073/pnas.1400406111.

F. Débarre. Fidelity of parent-offspring transmission and the evolution of social behavior in structured populations. *Journal of Theoretical Biology*, 420:26–35, May 2017. doi: 10.1016/j.jtbi.2017.02.027.

Q. Su, L. Zhou, and L. Wang. Evolutionary multiplayer games on graphs with edge diversity. *PLOS Computational Biology*, 15(4):e1006947, Apr 2019a. doi: 10.1371/journal.pcbi.1006947.

Q. Su, A. Li, L. Wang, and H. E. Stanley. Spatial reciprocity in the evolution of cooperation. *Proceedings of the Royal Society B: Biological Sciences*, 286(1900):20190041, Apr 2019b. doi: 10.1098/rspb.2019.0041.

B. Allen, G. Lippner, and M. A. Nowak. Evolutionary games on isothermal graphs. *Nature Communications*, 10(1), Nov 2019. doi: 10.1038/s41467-019-13006-7.

Jo Freeman. The Tyranny of Structurelessness. *WSQ: Women's Studies Quarterly*, 41(3-4): 231–246, 2013. doi: 10.1353/wsq.2013.0072.

B. Zhang, C. Li, H. De Silva, P. Bednarik, and K. Sigmund. The evolution of sanctioning institutions: an experimental approach to the social contract. *Experimental Economics*, Jun 2013. doi: 10.1007/s10683-013-9367-7.

B. Allen and C. E. Tarnita. Measures of success in a class of evolutionary models with fixed population size and structure. *Journal of Mathematical Biology*, 68(1-2):109–143, Jan 2014. doi: 10.1007/s00285-012-0622-x.

B. Allen and A. McAvoy. A mathematical formalism for natural selection with arbitrary spatial and genetic structure. *Journal of Mathematical Biology*, 78(4):1147–1210, Mar 2019. doi: 10.1007/s00285-018-1305-z.

Q. Su, A. McAvoy, L. Wang, and M. A. Nowak. Evolutionary dynamics with game transitions. *Proceedings of the National Academy of Sciences*, 116(51):25398–25404, Nov 2019c. doi: 10.1073/pnas.1908936116.

A.-L. Barabási and R. Albert. Emergence of Scaling in Random Networks. *Science*, 286(5439): 509–512, Oct 1999. doi: 10.1126/science.286.5439.509.

B. Bollobás. *Random Graphs*. Cambridge University Press, 2001. doi: 10.1017/cbo9780511814068.

D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393 (6684):440–442, Jun 1998. doi: 10.1038/30918.

J. Leskovec and A. Krevl. SNAP Datasets: Stanford Large Network Dataset Collection. `http://snap.stanford.edu/data`, June 2014.

R. A. Rossi and N. K. Ahmed. The Network Data Repository with Interactive Graph Analytics and Visualization. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015. URL `http://networkrepository.com`.

W. W. Zachary. An Information Flow Model for Conflict and Fission in Small Groups. *Journal of Anthropological Research*, 33(4):452–473, Dec 1977. doi: 10.1086/jar.33.4.3629752.

M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002. doi: 10.1073/pnas.122653799.

P. D. Taylor. Allele-Frequency Change in a Class-Structured Population. *The American Naturalist*, 135(1):95–106, Jan 1990. doi: 10.1086/285034.

W. D. Hamilton. The genetical evolution of social behaviour, I. *Journal of Theoretical Biology*, 7(1):1–16, Jul 1964. doi: 10.1016/0022-5193(64)90038-4.

L. Lehmann and F. Rousset. The genetical theory of social behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1642):20130357, 2014.

J. Birch. *The Philosophy of Social Evolution*. Oxford University Press, 2017.

M. A. Nowak, C. E. Tarnita, and E. O. Wilson. The evolution of eusociality. *Nature*, 466(7310): 1057–1062, Aug 2010. doi: 10.1038/nature09205.

B. Allen and M. A. Nowak. There is no inclusive fitness at the level of the individual. *Current Opinion in Behavioral Sciences*, 12:122–128, 2016.

Y.-T. Chen, A. McAvoy, and M. A. Nowak. Fixation Probabilities for Any Configuration of Two Strategies on Regular Graphs. *Scientific Reports*, 6(1), Dec 2016. doi: 10.1038/srep39181.

A. McAvoy and B. Allen. Fixation probabilities in evolutionary dynamics under weak selection. arXiv preprint arXiv:1908.03827, 2019.

G. Szabó and G. Fáth. Evolutionary games on graphs. *Physics Reports*, 446(4-6):97–216, Jul 2007. doi: 10.1016/j.physrep.2007.04.004.

M. Tomassini, E. Pestelacci, and L. Luthi. Social dilemmas and cooperation in complex networks. *International Journal of Modern Physics C*, 18(07):1173–1185, Jul 2007. doi: 10.1142/s0129183107011212.

A. Antonioni and M. Tomassini. Cooperation on social networks and its robustnesss. *Advances in Complex Systems*, 15(supp01):1250046, Jun 2012. doi: 10.1142/s0219525912500464.

B. Allen, M. A. Nowak, and U. Dieckmann. Adaptive Dynamics with Interaction Structure. *The American Naturalist*, 181(6):E139–E163, Jun 2013a. doi: 10.1086/670192.

B. Wu, A. Traulsen, and C. S. Gokhale. Dynamic properties of evolutionary multi-player games in finite populations. *Games*, 4(2):182–199, May 2013b. doi: 10.3390/g4020182.

L. E. Blume. The Statistical Mechanics of Strategic Interaction. *Games and Economic Behavior*, 5(3):387–424, Jul 1993. doi: 10.1006/game.1993.1023.

C. E. Tarnita, H. Ohtsuki, T. Antal, F. Fu, and M. A. Nowak. Strategy selection in structured populations. *Journal of Theoretical Biology*, 259(3):570–581, Aug 2009b. doi: 10.1016/j.jtbi. 2009.03.035.

H. Ohtsuki. Stochastic evolutionary dynamics of bimatrix games. *Journal of Theoretical Biology*, 264(1):136–142, May 2010. doi: 10.1016/j.jtbi.2010.01.016.

T. Sekiguchi and H. Ohtsuki. Fixation Probabilities of Strategies for Bimatrix Games in Finite Populations. *Dynamic Games and Applications*, 7(1):93–111, Sep 2015. doi: 10.1007/ s13235-015-0170-2.

A. McAvoy and C. Hauert. Asymmetric evolutionary games. *PLoS Computational Biology*, 11 (8):e1004349, Aug 2015a. doi: 10.1371/journal.pcbi.1004349.

A. McAvoy and C. Hauert. Structural symmetry in evolutionary games. *Journal of The Royal Society Interface*, 12(111):20150420, Sep 2015b. doi: 10.1098/rsif.2015.0420.

B. Allen and M. A. Nowak. Games on graphs. *EMS Surveys in Mathematical Sciences*, 1(1): 113–151, 2014. doi: 10.4171/emss/3.

B. Fotouhi, N. Momeni, B. Allen, and M. A. Nowak. Evolution of cooperation on large networks with community structure. *Journal of the Royal Society Interface*, 16(152):20180677, 2019.

Z. Rong and Z.-X. Wu. Effect of the degree correlation in public goods game on scale-free networks. *EPL (Europhysics Letters)*, 87(3):30001, 2009.

A. Szolnoki and M. Perc. Group-size effects on the evolution of cooperation in the spatial public goods game. *Physical Review E*, 84(4):047102, 2011.

M. Perc, J. Gomez-Gardenes, A. Szolnoki, L. M. Floría, and Y. Moreno. Evolutionary dynamics of group interactions on structured populations: a review. *Journal of the Royal Society Interface*, 10(80):20120997, 2013.

J. Peña, B. Wu, and A. Traulsen. Ordering structured populations in multiplayer cooperation games. *Journal of The Royal Society Interface*, 13(114):20150881, Jan 2016b. doi: 10.1098/rsif.2015.0881.

M. Le Gac and M. Doebeli. Environmental viscosity does not affect the evolution of cooperation during experimental evolution of colicigenic bacteria. *Evolution*, 64(2):522–533, 2010.

T. Julou, T. Mora, L. Guillon, V. Croquette, I. J. Schalk, D. Bensimon, and N. Desprat. Cell–cell contacts confine public goods diffusion inside pseudomonas aeruginosa clonal microcolonies. *Proceedings of the National Academy of Sciences*, 110(31):12577–12582, 2013.

R. Kümmerli, K. T. Schiessl, T. Waldvogel, K. McNeill, and M. Ackermann. Habitat structure and the evolution of diffusible siderophores in bacteria. *Ecology Letters*, 17(12):1536–1544, 2014.

B. Allen, J. Gore, and M. A. Nowak. Spatial dilemmas of diffusible public goods. *eLife*, 2, Dec 2013b. doi: 10.7554/elife.01169.