Social goods dilemmas in heterogeneous societies

Alex McAvoy, Benjamin Allen and Martin A. Nowak

Prosocial behaviours 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 that 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, in which some individuals have many more interaction partners than others, can enhance the evolution of prosocial behaviours. 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 favour 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 behaviours, but they also create the possibility of generating large inequality.

Prosocial behaviours are often studied using two-player or many-player games. In the first case, we encounter the donation game, prisoner’s dilemma or relaxed social dilemma. In the second case, we are typically in the world of public goods games. 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 neighbours. However, even within spatially heterogeneous populations, it is often assumed that every group (or pair) 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 neighbours. 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: (1) is the benefit of receiving the social good from a specific donor independent of the number of recipients (non-rival), or does one neighbour’s access to the good decrease that of another (rival)? and (2) 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, 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. In fact, distinguishing among various kinds of social goods is a common practice in economics, one that has not fully permeated evolutionary game theory. 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 Fig. 1.

As an example, consider the prosocial act of donating blood. 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 behaviour 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 modelled better as a pp good, whereas the latter could be viewed as an ff good. Similar arguments can be made for other kinds of social behaviours 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, food delivery among magpies and blood donation (as food) among vampire bats.

All neighbours 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 (for example, 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. This is also the case for information transmission within a social network; the initial acquisition of the information could entail a cost but its full benefit can be enjoyed by more than one individual. Entertainment, such as podcasts, radio programmes 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 also ordinarily represent pf goods.

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 with its transmission to a recipient. We focus our examples primarily on pp, ff and pf goods, since all of the interesting behaviour we observe can

1Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA. 2Department of Mathematics, Emmanuel College, Boston, MA, USA. 3Department of Mathematics, Harvard University, Cambridge, MA, USA. E-mail: alexmcavoy@g.harvard.edu; martin_nowak@harvard.edu
Fig. 1 | Social goods and prosocial behaviours. a. For pp goods, a producer pays cost $c$ for each neighbour to receive benefit $b$. For ff goods, a producer pays a fixed cost $c$ irrespective of the number of neighbours $k$; each neighbour receives benefit $b/k$. For pf goods, the cost $c$ is again independent of the number of neighbours, but now each neighbour 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 neighbours; 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 $b$ (or, respectively, $c$) gives an equivalence between pf goods and ff goods (or, respectively, 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.

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.

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, denoted $(w_{ij})$, satisfies $w_{ij} = 1$ if $i$ and $j$ are neighbours and $w_{ij} = 0$ if $i$ and $j$ are not neighbours. 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 neighbours 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 neighbours 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 $w_i$ neighbours; thus, each neighbour receives $b$. For an ff good, the producer pays cost $c$, and the total benefit $b$ is split among the $w_i$ neighbours; thus, each neighbour receives $b/w_i$. A pf good is a hybrid of these two goods; the total cost is $c$ and each neighbour gets $b$. The behaviour 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 we address is: when is a prosocial behaviour 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 $b > c$. By contrast, pf goods can be wealth producing even when $b < c$ since the total benefit, $b \sum_{i=1}^{N} w_{ij}$, is based on the number of edges, whereas 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 $b/c > N/\sum_{i=1}^{N} w_{ij}$.

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 (in which everyone uses strategy $B$) than in the all-$A$ state (in which everyone uses strategy $A$). 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 $\rho_A$ 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 $\rho_i$ the fixation probability of type B, defined in the same way as $\rho_0$, but with A and B swapped. We say that selection favours 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.

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 $B_i$ to j at a cost of $C_{ij}$ (Fig. 2). Let $\pi_{ij}$ be the fixation probability of a neutral trait starting in location i (also known the ‘reproductive value’ of i)\textsuperscript{39,40}. In the Supplementary Information, 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 $m^B_{ij}$ be the marginal effect of the fecundity of the fc on the probability that i replaces j. Using these quantities, which are described in detail in the Supplementary Information, we show that producers (A) are favoured over non-producers (B) under weak selection\textsuperscript{11-15} whenever

$$
\sum_{i,j,k,l=1}^{N} \pi_{ij}(\sum_{k,l=1}^{N} \pi_{kl} m^B_{ik} - x_{ik} C_{ki} + x_{ik} B_{ik}) > \sum_{i,j,k,l=1}^{N} \pi_{ij} m^B_{ij} - x_{ij} C_{ij} + x_{ij} B_{ij}
$$

(1)

This condition can be evaluated by solving a linear system of $O(N^2)$ equations, giving an overall complexity of $O(N^3)$ (since solving a linear system of n equations requires $O(n^3)$ 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 $B_i = b \beta_i$ and $C_{ij} = c \gamma_{ij}$ for some $b, c > 0$, where $\beta_i$ and $\gamma_{ij}$ are independent of $b$ and $c$. As a consequence, equation (1) can be written as $yb > bc$, where $\beta$ and $\gamma$ are independent of $b$ and $c$. When $\beta > 0$, 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 behaviours—that is, those with $b > 0$ and $c > 0$—cannot evolve; instead, selection can favour spiteful behaviours with $b < 0$ and $c > 0$ (‘costly harm’).\textsuperscript{16,47}. This property is one nuance of critical ratios, namely that they are lower bounds on $b/c$ when $\gamma > 0$ and upper bounds when $\gamma < 0$.

**Evolutionary outcomes on heterogeneous structures.** We consider several natural update rules that drive evolution through imitation. Under pairwise-comparison (PC) updating,\textsuperscript{4} a random individual is chosen to update their strategy. The individual compares its own payoff with that of a single, randomly chosen neighbour. If the neighbour has a higher payoff, then the focal individual adopts the neighbour’s strategy. If the neighbour has a lower payoff, then the focal individual retains its current strategy. The payoff comparison is subject to noise. Death-birth (DB) and imitation (IM) updating are similar, but they differ in the number of neighbours chosen for comparison and/or whether imitating some neighbour is compulsory (see Fig. 3).

These update rules are highly idealized, but they capture important qualitative features of behaviour imitation.\textsuperscript{48-50} For one thing, a learner is more likely to imitate a model individual’s behaviour as the model’s payoff increases. If a learner cannot compare his or her payoff to all neighbours at once (for instance, if he or she encounters neighbours only occasionally), then PC updating is relevant. When information is more readily available—for example, within scientific collaboration networks—IM updating could serve as a better model. In both cases, a learner is not compelled to imitate a behaviour. DB updating, which requires imitating some neighbour, 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 behaviour 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 show that producers of pp, ff and pf goods are never favoured 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.
Articles

NATURE HUMAN BEHAVIOUR

Fig. 3 | Four update rules driving evolutionary dynamics through imitation. When considering how to imitate a neighbour’s action (A or B) on the basis of payoff, four natural update rules arise, which have each been considered extensively in the literature. PC updating involves an individual choosing a random neighbour (yellow) with whom he or she compares payoffs. There is an option to imitate the neighbour, but the focal individual (green) may also choose to retain their existing action. IM updating is similar to this rule, except that the payoff comparison involves a focal individual and all neighbours. Again, this individual can imitate a neighbour but does not have to do so. If one insists that one of these neighbours must be imitated, then we have DB updating. In this case, the green individual is effectively chosen for death because retaining its current behaviour is not an option. The final logical case is when, similar to PC updating, only a single neighbour is chosen for comparison. This time, however, the focal individual must imitate this neighbour. This model turns out to be equivalent to DB updating when there is no game (that is, 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.

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 Supplementary Information, we also establish this result for pp goods on heterogeneous graphs under 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\(^\text{53,54}\).

Consider, for example, a ‘rich-club’ network\(^\text{41-43}\), 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 only to 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’\(^\text{44}\). Structures like the rich club arise within corporate hierarchies\(^\text{45}\), among students in a classroom (on the basis of academic performance)\(^\text{46}\), and among academic institutions (on the basis of funding)\(^\text{47,50}\). Surprisingly, \((bc)^*\) can fall between zero and one for both ff and pf goods in such populations (see Fig. 4 and Extended Data Fig. 1). 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 Fig. 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 \(bc\) 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 \(b > 0\) and \(c > 0\) for which selection favours producers depends on only the benefits of the social good; the costs influence the magnitude of \((bc)^*\) but not its sign (see Methods and Supplementary Information 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 Figs. 2 and 3). 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 Fig. 4) and other rich-club structures\(^\text{46}\), such as dense clusters of stars (Extended Data Fig. 5). Further comparisons of critical ratios for different kinds of social goods are shown in Extended Data Fig. 6 (small graphs) and Extended Data Fig. 7 (division of a group into factions).

Even when producers improve the total wealth of a population, their evolution can leave the poorest individuals worse off. In Fig. 5, we illustrate this phenomenon using IM updating on an empirical coauthorship network of 379 scientists\(^\text{57}\). Producers of ff goods can evolve only when \(b/c > 2.5703\), in which case the total wealth of the population increases. However, 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 the all-A state. Therefore, wealth-producing goods are not necessarily optimal for everyone in the population (see Extended Data Fig. 8 for a summary of possible evolutionary outcomes).

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, \(B_i\) and \(C_i\) 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; for example, 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\(^\text{58-60}\).

In addition to benefits and costs, the nature of the social goods themselves might vary from location to location. Returning to the rich club (Fig. 4), the central clique might represent a network of content producers (for example, radio content), while those at the periphery are consumers. If listeners donate money, then this scenario could be reasonably modelled using pf goods in the centre 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 Fig. 4, relative to when everyone produces a pp good (respectively, ff good), it is generally easier (respectively, harder) for producers to evolve when those in the central 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 more difficult for producers to evolve at all. Such a population leads to a trade-off between the following two scenarios: (1) 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. (2) Well-connected producers ensure that each neighbour 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
Supplementary Information, we also discuss how an institution can mitigate the harmful effects of certain prosocial behaviours (such as the production of ff goods), for example, through a ‘tax’ (Extended Data Fig. 9).

Another form of asymmetry arises from stochasticity in the recipient of a donation. Instead of either producing a social good for each neighbour or dividing it up among the neighbourhood, a producer might choose a single random neighbour as the recipient of the good in its entirety (Fig. 6). Such a payoff scheme could be driven by indifference, meaning that 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 a 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 Supplementary Information, 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 Fig. 6b is equivalent to that of ff goods (Fig. 6a) and producers can evolve even when $b < c$.

Reciprocity. The behavioural 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 behavioural 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 cooperation9–17.

There are many different ways to model reciprocity of social good production in heterogeneous populations. As a starting point, consider the ‘tit-for-tat’ (TFT) strategy, which cooperates (donates)
in the first round and then subsequently copies what the opponent did in the previous round. Let $B_i$ and $C_i$ be the benefit and cost of $i$ donating to $j$. In our model, an individual using TFT gives $B_i$ to every $j$ (at a cost of $C_i$) in the first round. In subsequent rounds, $i$ donates $B_i$ to $j$ (still at cost $C_i$) if and only if $j$ donated to $i$ in the previous round; otherwise, $i$ gives nothing to $j$ and pays no cost associated with $j$. 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 neighbours. Subsequently, a TFT player looks around and counts how many neighbours produced a good in the previous round. If a fraction $x$ of one's neighbours produced a good, then in the subsequent round this TFT player produces a good of benefit $xb$ and cost $xc$ (that is, a fraction $x$ of the original good). This benefit is divided among only those neighbours 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 that competition between the always defect (ALLD) and TFT strategies is an unconditional non-producer. Since ALLD never produces a good, this iterated game exhibits quite straightforward behaviour: 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 neighbourhood. ALLD players get only the benefits they receive in the first round.

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 $\lambda$ 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 favoured over ALLD is

\[
\sum_{i,j,k=1}^{N} \pi_i m_{jk}^B \left( -\left( x_{jk} + \lambda x_{jk} \right) C_{jk} + \left( x_{jk} + \lambda x_{jk} \right) B_{jk} \right)
\]

When \( \lambda = 0 \), equation (2) reduces to equation (1). We derive this result in the Supplementary Information and provide simplified formulas for PC, DB, and IM updating in Methods. Figure 7 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 behaviours, which may be either helpful or harmful to the population at large and/or to the least well-off. For example, on the coauthorship network of Fig. 5, 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 behaviour is pairwise mutually beneficial (PMB) if and only if for each pair of individuals expressing this behaviour, both partners receive at least as much as they pay (for that particular interaction). PMB behaviours cannot be wealth decreasing, nor can they lead to negative payoffs in the all-producer state. The simplest example of a PMB behaviour 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 behaviour is PMB depends on both the good and the population structure. For example, when \( b > c \), if goods are always PMB on regular graphs, but not necessarily on heterogeneous networks such as the rich club. PMB behaviours are not always favoured by selection, as was demonstrated in our baseline model of one-shot interactions. However, we show in the Supplementary Information that for any given PMB behaviour, TFT is favoured for all \( \lambda \) greater than some threshold value \( \lambda' < 1 \).

**Stronger selection.** The analytical conditions presented so far (for example, 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 neighbour of \( i \) has payoff \( u_j \), and if \( j \) is chosen as a model individual for comparison with \( i \), then \( i \) imitates the behavioural type of \( j \) with probability \( \frac{1}{1 + e^{-\delta(u_i - u_j)}} \), where \( \delta \) is positive but sufficiently small. A natural question arising here is what happens when \( \delta \) is not necessarily small, which is relevant when individuals are highly inclined to imitate a neighbour with a larger payoff.

Analytical calculations for stronger selection (larger \( \delta \)) quickly become infeasible for arbitrary heterogeneous graphs\(^2\), but remain
tractable for structures with a high degree of symmetry, such as the star\(^7\). In Fig. 8, we consider if 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 when total costs exceed total benefits. We can quantify selection for producers by the fraction of time the population spends in the all-producer state under rare mutation, which is \(p_j/(p_j + p_0)\). When \(b=0.1\) and \(c=1\), the star spends approximately 64\% of the time in the all-producer state at its peak (Fig. 8c). When \(b=2\) and \(c=1\), this number is slightly larger at approximately 66\% (Fig. 8d). In both cases, the all-producer state is highly unequal, with a large, positive payoff for the individual at the centre of the star and negative payoffs for the \(N-1=499\) remaining individuals at the periphery.

**Discussion**

A large and growing body of research\(^1\) has shown that spatial structure can promote the evolution of cooperative or prosocial behaviours. Our work, while affirming this principle, reveals it to be more complicated than it may seem. First, we show that the conditions for prosocial behaviours 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 benefit is an increasing function of the total 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 favour 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 if goods can evolve when even they are a net detriment to the population (\(b<0\)).

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 favoured (see Extended Data and Supplementary Information). For example, homogeneous graphs can support the evolution of producers under DB and IM updating but not under PC updating. Some heterogeneous structures can favour 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 can be applied to many other populations.

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 behaviour with respect to social goods (for example, 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\(^6\)–\(^7\), which holds in our model as well. In the context of repeated interactions, reciprocal prosocial behaviours 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 prosociability 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 difficult questions about the nature and consequences of cooperative or prosocial behaviours. 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, heterogeneous population structures act as strong promoters of the evolution of prosocial behaviours. However, the resulting prosocial behaviours can lead to payoffs distributions in which a few highly connected nodes accumulate much of the total wealth\(^6\), 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\(^6\). 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\(^8\) 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 behaviours.

**Methods**

**Modelling evolutionary dynamics.** We model a general evolutionary process in a population of finite size \(N\) using the notion of a replacement rule\(^9\). If the process is in state \(x\in\{0,1\}^N\) at a given time step, where \(x_i=1\) indicates that the population is in state \(x\) and \(x_i=0\) indicates that \(i\)'s type \(A\) and \(B\) is the benefit \(j\) provides to \(i\) when \(j\) is a producer. This payoff is converted to relative fecundity via the formula \(F(x)=\exp(\text{dir}(x))\), which is then used to determine the probability of choosing replacement event \((R,\alpha)\) in state \(x\). For example, suppose that \(w_{\alpha,i}^{(\text{ff})}\) is the adjacency matrix for an undirected, unweighted graph. Let \(p_{R,\alpha}(\text{fp})\) be the one-step probability of moving from \(i\) to \(j\) in a random walk on this graph, where \(w_{\alpha,i}^{(\text{fp})}:=\sum_{j=1}^{N-1}w_{\alpha,j}\) is the number of individuals neighbouring \(i\). Under PC updating, the probability of choosing \((R,\alpha)\) is

\[
\begin{align*}
\left[1 - \frac{1}{N} \sum_{j=1}^{N-1} p_{R,\alpha}(\text{fp}) \right] & R = \{i\} \text{ for some } i \in \{1, \ldots, N\}, \alpha(i) \neq i \\
\frac{1}{N} & R = \{i\} \text{ for some } i \in \{1, \ldots, N\}, \alpha(i) = i \\
0 & \text{otherwise}
\end{align*}
\]

Similarly, the probability of replacement event \((R,\alpha)\) under DB updating is

\[
\begin{align*}
\frac{1}{N} & R = \{i\} \text{ for some } i \in \{1, \ldots, N\}, \alpha(i) \neq i \\
0 & \text{otherwise}
\end{align*}
\]

under IM updating and

\[
\begin{align*}
\frac{1}{N} & R = \{i\} \text{ for some } i \in \{1, \ldots, N\}, \alpha(i) \neq i \\
0 & \text{otherwise}
\end{align*}
\]
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 $p_A^{(1)}$ (respectively, $p_B^{(1)}$) be the probability that a single A (respectively, B), placed initially at location $i$, takes over a population of type B (respectively, type A). The mean fixation probabilities for the two types are then $p_A = \frac{1}{N} \sum_{i=1}^{N} p_A^{(i)}$ and $p_B = \frac{1}{N} \sum_{i=1}^{N} p_B^{(i)}$, respectively.

By analysing the demographic variables (for example, birth rates and death probabilities) resulting from this process, together with the transient dynamics (that is, prior to absorption in all-A or all-B), we derive a condition for $p_A > p_B$ under weak selection ($\delta \ll 1$). Specifically, if (1) $\pi$ is the reproductive value of vertex $i$; (2) $x_i$ is the probability that $i$ and $j$ have the same type in the neutral process (with respect to a distribution described in the Supplementary Information); and (3) $m_i$ is the marginal effect of $i$’s fecundity on $i$ replacing $j$, then $p_A > p_B$ 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 $x_i$, may be found in the Supplementary Information.

PC updating. For PC updating, we have $p_A > p_B$ under weak selection if and only if

$$\sum_{i=1}^{N} x_i \sum_{j=1}^{N} \left( -x_j C_{ij} + x_i B_{ij} \right) > \sum_{i=1}^{N} \sum_{j=1}^{N} \pi_j \pi_i \sum_{j=1}^{N} \left( -x_j C_{ij} + x_i B_{ij} \right)$$  

(7)

Informally, this condition says that the expected payoff to a producer, $i$, chosen with probability $\pi_i$ (which in this case is $w_i / \sum_{i=1}^{N} w_i$), exceeds that of a random neighbour. In the Supplementary Information, we show that we can evaluate equation (7) by replacing $x_i$ by $\tau_i$, where $\tau_i = 0$ for every $i$ and

$$\tau_i = 1 + \frac{1}{2} \sum_{j=1}^{N} \frac{1}{p_A} \tau_{ij} + \frac{1}{2} \sum_{j=1}^{N} \frac{1}{p_B} \tau_{ji}$$

(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 $B_i$ and $C_i$. For pp goods, we have $B_i = w_i b$ and $C_i = w_i c_i$. For fi goods, $B_i = w_i b_i$ and $C_i = w_i c_i$. In each case, equation (7) can be expressed as $\pi_i b_i c_i$, where both $\pi_i$ and $\tau_i$ are independent of $b_i$ and $c_i$. With $b > 0$ and $c > 0$, a necessary condition for producers to be favoured is therefore $b > c$. Writing $B_i = b_j$ and $C_i = c_j$, we see that when $b > c$, the selection condition is

$$p_A > p_B \iff \frac{b}{c} > \left( \pi_i \right) c_i$$

(9)

On a homogeneous graph, one can show that $b > c$, so that the same condition can be used to show 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 (in which case the interaction was essentially a coordination game). In contrast, on any homogeneous graph without game transitions, producers (including cooperators) cannot be favoured under PC updating. However, as shown in the text, there are many examples of heterogeneous population structures with $b > c$.

DB updating. For DB updating, we have $p_A > p_B$ under weak selection if and only if

$$\sum_{i=1}^{N} x_i \sum_{j=1}^{N} \left( -x_j C_{ij} + x_i B_{ij} \right) > \sum_{i=1}^{N} \sum_{j=1}^{N} \pi_j \pi_i \sum_{j=1}^{N} \left( -x_j C_{ij} + x_i B_{ij} \right)$$

(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) neighbour.

Using the same values of $\pi_i$ and $\tau_i$ described above for PC updating, this condition becomes

$$p_A > p_B \iff \frac{b}{c} > \left( \pi_i \right) c_i$$

(11)

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

$$\tau_i = 1 + \frac{1}{3} \sum_{j=1}^{N} p_A \tau_{ij} + \frac{1}{3} \sum_{j=1}^{N} p_B \tau_{ji}$$

(13)

For $B_i = b j_i$ and $C_i = c j_i$, the selection condition when $\gamma > 0$ is

$$p_A > p_B \iff \frac{b}{c} > \left( \pi_i \right) c_i$$

(14)

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 $c_j$ and divides the benefit $b_j$ among all neighbours. A producer in the centre group pays $c_i$ and distributes the benefit $b_i$ evenly. We consider the case in which $b_j > b_i$ and $c_j > c_i$, and that $b_i = c(m,n)b$ and $c_j = c(m,n)c$, for some function $c(m,n)$. If everyone on the rich club is a producer (whose state in [0,1] is denoted $A$), then the payoff to individual $i$ is

$$u_i(A) = \left\{ \begin{array}{ll} -c(m,n) + \frac{1}{m} \sum_{j=1}^{m} (m,n) b_j + \frac{b_j}{m} & , i \text{ is in the central rich club} \\
-c(m,n) c_j & , i \text{ is on the periphery} \end{array} \right.$$

(15)

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

$$\lim_{n \to \infty} u_i(A) = \left\{ \begin{array}{ll} -c(m,n) c_k & , i \text{ is in the central rich club} \\
-c(m,n) k(m) & , i \text{ is in the central club} \\
-c(m,n) k(m) + \frac{b_j}{m} & , i \text{ is on the periphery} \end{array} \right.$$

(16)

For the first payoff in equation (15) to stay non-negative as $n \to \infty$, we require $b/c \geq mk(m)$. But we also want the second payoff in equation (15) to be non-negative, that is, $b/c \geq 1/(mk(m))$. Thus, we seek $k_i$ with

$$\frac{c}{b} \leq k(m) \leq \frac{b}{c}$$

(17)

Such $k_i$ exists if and only if $b/c \geq 2$, in which case we can set $k_i(m) = 0$ and $k_i(m) = 1/m$. For large $n$, it follows that each producer in the central clique gives an average of $b/m$ to each neighbour at a cost of $c/m$. The peer-neighbour benefit and cost values are the same ($b/m$ and $c/m$), respectively for each peripheral individual as well, which effectively transforms the payoff 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).

Reciprocity. For the model of reciprocity defined in the text, we let $A$ and $B$ denote the strategies TFT and ALLD, respectively. Let $b$ be the payoff to player $i$ in the rth round of the game, that is

$$u_i'(x) = \left\{ \begin{array}{ll} \sum_{j=1}^{N} x_j (-x_i C_{ij} + x_j B_{ij}) & , i = 1 \\
\sum_{j=1}^{N} x_j (-x_i C_{ij} + x_j B_{ij}) & , i > 1 \end{array} \right.$$

(18)
If the discounting factor (continuation probability) is \( \lambda \in [0,1) \), then the overall payoff to \( i \) is

\[
u_i(\{x\}) = (1 - \lambda) \left( \sum_{j=1}^{N} (C_{ij} + x_{ij}B_{ij}) + \lambda \sum_{j=1}^{N} x_{ij}(C_{ij} + B_{ij}) \right)
\]

\[
= (1 - \lambda) \sum_{j=1}^{N} (C_{ij} + x_{ij}B_{ij}) + \lambda \sum_{j=1}^{N} x_{ij}(C_{ij} + B_{ij}) + \lambda \sum_{j=1}^{N} x_{ij}C_{ij} + (1 - \lambda + \lambda x_i)B_{ij}
\]

\[
= \sum_{j=1}^{N} (1 - \lambda + \lambda x_i) x_{ij}C_{ij} + (1 - \lambda + \lambda x_i)B_{ij}
\]

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

\[
u_i(\{x\}) = \begin{cases} 
(1 - \lambda) \sum_{j=1}^{N} x_{ij}C_{ij} + (1 - \lambda + \lambda x_i)B_{ij} & \text{if } x_i = 1 \\
(1 - \lambda) \sum_{j=1}^{N} x_{ij}B_{ij} & \text{if } x_i = 0 
\end{cases}
\]

It follows that, in each state \( x \in \{0,1\}^N \), increasing \( \lambda \) does not decrease the payoff to an A and does not increase the payoff to a B. For any reasonable process favouring individuals with higher payoffs (including PC, DB and IM updating), it follows that \( \rho_A \) is monotonically increasing and \( \rho_B \) is monotonically decreasing in \( \lambda \). Furthermore, we note that when \( \lambda = 1 \), \( \rho_A \) gets 0 when using ALLD and \( \sum_{j=1}^{N} x_{ij}(C_{ij} + B_{ij}) \) when using TFT. If the underlying behaviour is PMR, 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 favouring traits with higher payoffs should disfavour TFT relative to ALLD when the interactions have an infinite time horizon.

The selection condition for ALLD versus TFT, equation (Z), is derived in the Supplementary Information. Here, we explore how this condition can be evaluated for the update rules considered in the text.

**PC updating.** For PC updating, the condition for selection to favour TFT relative to ALLD is

\[
\rho_A > \rho_B \iff \sum_{j=1}^{N} \pi_j \sum_{i=1}^{N} (x_{ij} + \lambda x_i)C_{ji} + (x_{ij} + \lambda x_i)B_{ji},
\]

\[
> \sum_{j=1}^{N} \pi_j \sum_{i=1}^{N} (x_{ij} + \lambda x_i)C_{ji} + (x_{ij} + \lambda x_i)B_{ji}
\]

\[
B > \sum_{j=1}^{N} \pi_j \sum_{i=1}^{N} (x_{ij} + \lambda x_i)C_{ji} + (x_{ij} + \lambda x_i)B_{ji}
\]

For social goods satisfying \( B_i = b_{ji} \) and \( C_i = c_{pi} \), if \( r > 0 \) then this condition is

\[
\frac{b}{c} > \frac{\sum_{j=1}^{N} \pi_j b_{ji}^r (\tau_j + \lambda \tau_i - \lambda \tau_j) \gamma_j}{\sum_{j=1}^{N} \pi_j b_{ji}^r (\lambda \tau_i + \tau_j - \lambda \tau_j) \beta_j}
\]

where \( \tau_i \) and \( \tau_j \) are the same as they were previously for PC updating with \( \lambda = 0 \).

**DB updating.** For DB updating, the condition for selection to favour TFT relative to ALLD is

\[
\rho_A > \rho_B \iff \sum_{j=1}^{N} \pi_j \sum_{i=1}^{N} (x_{ij} + \lambda x_i)C_{ji} + (x_{ij} + \lambda x_i)B_{ji},
\]

\[
> \sum_{j=1}^{N} \pi_j \sum_{i=1}^{N} (x_{ij} + \lambda x_i)C_{ji} + (x_{ij} + \lambda x_i)B_{ji}
\]

\[
B > \sum_{j=1}^{N} \pi_j \sum_{i=1}^{N} (x_{ij} + \lambda x_i)C_{ji} + (x_{ij} + \lambda x_i)B_{ji}
\]

For social goods satisfying \( B_i = b_{ji} \) and \( C_i = c_{pi} \), if \( r > 0 \) then this condition is

\[
\frac{b}{c} > \frac{\sum_{j=1}^{N} \pi_j b_{ji}^r (\tau_j + \lambda \tau_i - \lambda \tau_j) \gamma_j}{\sum_{j=1}^{N} \pi_j b_{ji}^r (\lambda \tau_i + \tau_j - \lambda \tau_j) \beta_j}
\]

where \( \tau_i \) and \( \tau_j \) are the same as they were previously for DB updating with \( \lambda = 0 \).

**IM updating.** For IM updating, the condition for selection to favour TFT relative to ALLD is

\[
\sum_{j=1}^{N} \pi_j (x_{ij} + \lambda x_i)C_{ji} + (x_{ij} + \lambda x_i)B_{ji},
\]

\[
> \sum_{j=1}^{N} \pi_j b_{ji}^r (x_{ij} + \lambda x_i)C_{ji} + (x_{ij} + \lambda x_i)B_{ji}
\]

For social goods satisfying \( B_i = b_{ji} \) and \( C_i = c_{pi} \), if \( r > 0 \) then this condition is

\[
\frac{b}{c} > \frac{\sum_{j=1}^{N} \pi_j b_{ji}^r (\tau_j + \lambda \tau_i - \lambda \tau_j) \gamma_j}{\sum_{j=1}^{N} \pi_j b_{ji}^r (\lambda \tau_i + \tau_j - \lambda \tau_j) \beta_j}
\]
Articles

56. Dong, Y. et al. Inferring social status and rich club effects in enterprise networks. Phys. Rev. Lett. 95, 098104 (2005).
57. Vaquero, L. M. & Cebrian, M. The rich club phenomenon in the classroom. Sci. Rep. 3, 1174 (2013).
58. Ma, A., Mondragon, R. J. & Latora, V. Anatomy of funded research in science. Proc. Natl Acad. Sci. USA 112, 14760–14765 (2015).
59. Szell, M. & Sinatra, R. Research funding goes to rich clubs. Proc. Natl Acad. Sci. USA 112, 14749–14750 (2015).
60. Jiang, Z.-Q. & Zhou, W.-X. Statistical significance of the rich-club phenomenon in complex networks. New J. Phys. 10, 043002 (2008).
61. Newman, M. E. J. Finding community structure in networks using the eigenvectors of matrices. Phys. Rev. E 74, 036104 (2006).
62. Laurance, W. F. Second thoughts on who goes where in author lists. Nature 442, 26 (2006).
63. Venkatraman, V. Conventions of scientific authorship. Science https://doi.org/10.1126/science.aat10039 (16 April 2010).
64. Bošnjak, L. & Marušić, A. Prescribed practices of authorship: review of codes from professional bodies and journal guidelines across disciplines. Scientometrics 93, 751–763 (2012).
65. McNutt, M. K. et al. Transparency in authors’ contributions and responsibilities to promote integrity in scientific publication. Proc. Natl Acad. Sci. USA 115, 2557–2560 (2018).
66. Trivers, R. L. The evolution of reciprocal altruism. Q. Rev. Biol. 45, 35–57 (1971).
67. Nowak, M. A. Five rules for the evolution of cooperation. Science 314, 1560–1563 (2006).
68. Press, W. H. & Dyson, F. J. Iterated prisoner’s dilemma contains strategies that dominate any evolutionary opponent. Proc. Natl Acad. Sci. USA 109, 10409–10413 (2012).
69. Stewart, A. J. & Plotkin, J. B. From extortion to generosity, evolution in the Iterated Prisoner’s Dilemma. Proc. Natl Acad. Sci. USA 110, 15348–15353 (2013).
70. Stewart, A. J. & Plotkin, J. B. Small groups and long memories evolve. Proc. Nat. Acad. Sci. Rep. 6, 26889 (2016).
71. Hilbe, C., Chatterjee, K. & Nowak, M. A. Partners and rivals in direct reciprocity. Nat. Hum. Behav. 2, 469–477 (2018).
72. Ibsen-Jensen, R., Chatterjee, K. & Nowak, M. A. Computational complexity of ecological and evolutionary spatial dynamics. Proc. Natl Acad. Sci. USA 112, 201511366 (2015).
73. Hadjichrysanthou, C., Broom, M. & Rychtář, J. Evolutionary games on star graphs under various updating rules. Dyn. Games Appl. 1, 386–407 (2011).
74. Taylor, P. D. Altruism in viscous populations—an inclusive fitness model. Evol. Ecol. 6, 352–356 (1992).
75. Wilson, D. S., Pollock, G. B. & Dugatkin, L. A. Can altruism evolve in purely viscous populations? Evol. Ecol. 6, 331–341 (1992).
76. Mitteldorf, J. & Wilson, D. S. Population viscosity and the evolution of altruism. J. Theor. Biol. 204, 481–496 (2000).
77. Irwin, A. J. & Taylor, P. D. Evolution of altruism in stepping-stone populations with overlapping generations. Theor. Popul. Biol. 60, 315–325 (2001).
78. Grafen, A. An inclusive fitness analysis of altruism on a cyclical network. J. Evol. Biol. 20, 2278–2283 (2007).
79. Lion, S. & van Baalen, M. Self-structuring in spatial evolutionary ecology. Evol. Lett. 11, 277–295 (2017).
80. Tarnita, C. E., Antal, T., Ohtsuki, H. & Nowak, M. A. Evolutionary dynamics in set structured populations. Proc. Natl Acad. Sci. USA 106, 8601–8604 (2009).
81. Nowak, M. A., Tarnita, C. E. & Antal, T. Evolutionary dynamics in structured populations. Phil. Trans. R. Soc. B 365, 19–30 (2009).
82. Rand, D. G., Nowak, M. A., Fowler, J. H. & Christakis, N. A. Static network structure can stabilize human cooperation. Proc. Natl Acad. Sci. USA 111, 17093–17098 (2014).
83. Désbarres, M. & Guillaume, J.-L. Fidelity of parent-offspring transmission and the evolution of social behavior in structured populations. J. Theor. Biol. 420, 26–35 (2017).
84. Su, Q., Zhou, L. & Wang, L. Evolutionary multiplayer games on graphs with edge diversity. PLoS Comput. Biol. 15, e1006947 (2019).
85. Su, Q., Li, A., Wang, L. & Stanley, H. E. Spatial reciprocity in the evolution of cooperation. Proc. R. Soc. B 286, 20190041 (2019).
86. Allen, B., Lippner, G. & Nowak, M. A. Evolutionary games on isothermal graphs. Nat. Commun. 10, 5107 (2019).
87. Freeman, J. The tyranny of structurelessness. Women’s Stud. Q. 41, 231–246 (2013).
88. Zhang, B., Li, C., De Silva, H., Rednapp, P. & Sigmund, K. The evolution of sanctioning institutions: an experimental approach to the social contract. Exp. Econ. 17, 285–303 (2014).
89. Allen, B. & Tarnita, C. E. Measures of success in a class of evolutionary models with fixed population size and structure. J. Math. Biol. 68, 109–143 (2014).
90. Allen, B. & McAvoy, A. A mathematical formalism for natural selection with arbitrary spatial and genetic structure. J. Math. Biol. 78, 1147–1210 (2019).
91. Su, Q., McAvoy, A., Wang, L. & Nowak, M. A. Evolutionary dynamics with game transitions. Proc. Natl Acad. Sci. USA 116, 25398–25404 (2019).
Acknowledgements
We thank J. Plotkin for constructive feedback and B. Fotouhi and C. Hilbe for helpful conversations. This work was supported by the Army Research Laboratory (grant W911NF-18-2-0265), the Bill & Melinda Gates Foundation (grant OPP1148627), the John Templeton Foundation (grant 61443), the National Science Foundation (grant DMS-1715315), and the Office of Naval Research (grant N00014-16-1-2914). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions
A.M. designed the study and derived the initial results; A.M., B.A. and M.A.N. analysed the model; A.M., B.A. and M.A.N. wrote the main text; and A.M. and B.A. wrote the supplementary information.

Competing interests
The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41562-020-0881-2.
Supplementary information is available for this paper at https://doi.org/10.1038/s41562-020-0881-2.
Correspondence and requests for materials should be addressed to A.M. or M.A.N.
Reprints and permissions information is available at www.nature.com/reprints.
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
Editor recognition statement Primary Handling Editor: Charlotte Payne.
© The Author(s), under exclusive licence to Springer Nature Limited 2020
Extended Data Fig. 1 | Evolution of producers of ff-goods with 0 < b ≤ 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/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 ≤ 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 ρ_A > ρ_B.
Extended Data Fig. 2 | 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.
Extended Data Fig. 3 | 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.
Extended Data Fig. 4 | Heterogeneous graphs allow efficient evolution of prosocial behaviour. 

**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.
Extended Data Fig. 5 | 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.
Extended Data Fig. 6 | Graphs of size $N \leq 10$ 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.
Extended Data Fig. 7 | 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\(^97\), \(a\), and the subsequent split of the karate club into two disjoint groups\(^98\), \(b\) 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, \((b/c)_r\), is lowered by the split.

97. Zachary, W. W. An information flow model for conflict and fission in small groups. *J. Anthropol. Res.* **33**, 452–473 (1977).
98. Girvan, M. & Newman, M. E. J. Community structure in social and biological networks. *Proc. Natl Acad. Sci. USA* **99**, 7821–7826 (2002).
**Extended Data Fig. 8 | 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.

| Social good | 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
Extended Data Fig. 9 | Contributions by producers to a public pool can ameliorate payoff inequality. For ff-goods, suppose that each producer (blue) donates $\theta b$ 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 $\theta$, 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, $\theta$, when $b = 2$ and $c = 1$. This payoff is negative when $\theta \leq 1/2$, which means that 99% of the population is better off in the asocial (all-B) state. However, when $\theta \geq 1/2$, all individuals are better off when producers proliferate.
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted

Give P values as exact values whenever suitable.

For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings

For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes

Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection Numerical simulations were performed in MATLAB R2019a.

Data analysis Numerical simulations were performed in MATLAB R2019a.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The numerical simulations are described in Methods and SI of the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
- Behavioural & social sciences
- Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf
### Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

| Study description | This study is theoretical in nature and is based primarily on analytical calculations (described in SI). |
|-------------------|-----------------------------------------------------------------------------------------------------|
| Research sample   | N/A                                                                                                  |
| Sampling strategy | N/A                                                                                                  |
| Data collection   | N/A                                                                                                  |
| Timing and spatial scale | N/A                                                                                               |
| Data exclusions   | N/A                                                                                                  |
| Reproducibility   | N/A                                                                                                  |
| Randomization     | N/A                                                                                                  |
| Blinding          | N/A                                                                                                  |

Did the study involve field work?  [ ] Yes  [x] No

### Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

#### Materials & experimental systems

| n/a | Involved in the study |
|-----|-----------------------|
| [x] | Antibodies            |
| [x] | Eukaryotic cell lines |
| [x] | Palaeontology         |
| [x] | Animals and other organisms |
| [x] | Human research participants |
| [x] | Clinical data         |

#### Methods

| n/a | Involved in the study |
|-----|-----------------------|
| [x] | ChIP-seq              |
| [x] | Flow cytometry        |
| [x] | MRI-based neuroimaging |