Social evolution in structured populations

F. Débarre[1,†], C. Hauert[2] & M. Doebeli[1,2]

Understanding the evolution of social behaviours such as altruism and spite is a long-standing problem that has generated thousands of articles and heated debates. Previous theoretical studies showed that whether altruism and spite evolve may be contingent on seemingly artificial model features, such as which rule is chosen to update the population (for example, birth–death or death–birth), and whether the benefits and costs of sociality affect fecundity or survival. Here we unify these features in a single comprehensive framework. We derive a general condition for social behaviour to be favoured over non-social behaviour, which is applicable in a large class of models for structured populations of fixed size. We recover previous results as special cases, and we are able to evaluate the relative effects of benefits and costs of social interactions on fecundity and survival. Our results highlight the crucial importance of identifying the relative scale at which competition occurs.

[1] Department of Zoology and Biodiversity Research Centre, University of British Columbia, 6270 University Boulevard, Vancouver, British Columbia, Canada V6T 1Z4. [2] Department of Mathematics, University of British Columbia, 1984 Mathematics Road, Vancouver, British Columbia, Canada V6T 1Z2. † Present address: Centre for Ecology & Conservation, University of Exeter, Cornwall Campus, Tremough, Penryn TR10 9EZ, UK. Correspondence and requests for materials should be addressed to F.D. (email: florence.debarre@normalesup.org).

Why would an individual help or harm others if this reduces its fitness relative to individuals that do not exhibit the social behaviour? Because such behaviours are widespread in nature, numerous theoretical studies have sought to understand the conditions for the evolution of altruism or spite. Although different theoretical frameworks and vocabularies coexist, which sometimes generated heated debates between the proponents of different schools of thought[1–3], they all yield the same kind of conclusions: whether described in terms of games, or direct and indirect fitness or among and within group interactions, altruism requires some form of assortment, so that altruists interact more often with altruists than defectors do[4], while the evolution of spite requires negative assortment[5]. Assortment occurs for instance through clustering in spatially structured (that is, 'viscous') populations[6], or through conditional behaviour when there is kin or type recognition[7–9]. However, the conditions for the evolution of altruism or spite often depend on specific and sometimes artificial model assumptions, such as whether generations are discrete (Wright–Fisher model) or continuous (Moran[10] model)[11,12]. In the latter case, the sequence of events within one time step (death followed by birth, versus birth followed by death) is of crucial importance for the evolutionary outcome[13,14], yet the choice itself seems arbitrary[15]. Also, in most previous models, costs and benefits of altruism or spite (that is, the 'payoffs') are assumed to affect the fecundity of individuals, and costs and benefits for survival have received much less attention[16–18], despite the fact that such effects are equally plausible and should hence be incorporated in general models of social evolution.

Here we present a comprehensive modelling framework that applies to a large class of population structures and not only unites the various assumptions in a single model, but identifies the crucial elements which support the evolution of social behaviour. In structured populations, social behaviour evolves if, for social individuals, the net social benefit of living next to other social individuals outweighs the costs of competing against them. We show that the latter depends on the way the population is updated, the type of social game that is played, and on how social interactions affect individual fecundity and survival.

Results

Dispersal and social structures of the population. We consider a population of fixed size N whose structure is described by two graphs: a dispersal graph $\mathcal{D}$ and an interaction graph $\mathcal{E}$, where each node, also called site, corresponds to one individual of the population. Individuals reproduce clonally, and $\mathcal{D}$ determines dispersal patterns in the population: d_{ij} is the fraction of the offspring produced by the individual at site i that disperse to and compete to colonize site j ($\sum_j d_{ij} = 1$). We assume that the pattern of dispersal is similar for all individuals (technically, $\mathcal{D}$ is transitive, see Supplementary Methods)[14,18,19], and is symmetrical ($d_{ij} = d_{ji}$), and we denote by $d_{self} = d_{ii}$ the fraction of offspring that remain at their parent's site. Most classical population structures fall into this category[19]: metapopulations, regular lattices, stepping stones, but also groups of groups, and so on (see Fig. 1). Social interactions are reflected in the second graph $\mathcal{E}$, in which e_{ij} measures the strength of the social interaction between the individual at i and the individual at j (scaled so that $\sum_i e_{ij} = 1$). We denote by e_{self} the average strength of social interactions with oneself ($e_{self} = \sum_i e_{ii}/N$). Finally, the structural average $\overline{ed} = \sum_i \sum_i e_{ij} d_{ji}/N$ can be interpreted as the average over all pairs of sites (i, j) of the chance of receiving benefits (e_{ij}) from a site where offspring have been sent (d_{ji}), but also as a measure of the relatedness of an individual to its social
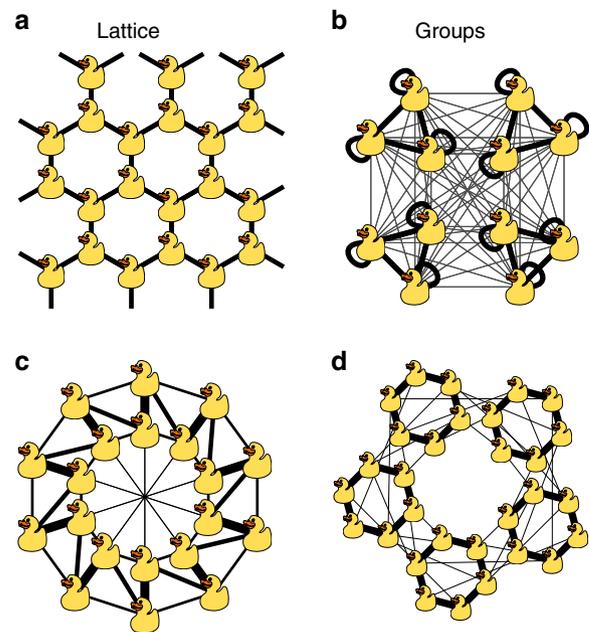
Figure 1 | Transitive dispersal graphs $\mathcal{D}$. Each duck corresponds to one individual living on one site. The thickness of a link between sites i and j is proportional to d_{ij}. (**a**) Lattice: $d_{self} = 0$; $d_{ij} = 1/k$ between i and its k neighbours, and $d_{ij} = 0$ for non neighbours; here $k = 3$. (**b**) Group-structured population: $d_{ij} = (1 - m)/n$ (where m is the chance of dispersing out of the group and n is the size of the groups; here $n = 3$) when i and j belong to the same group (thick links; the loops represent d_{self}); $d_{ij} = m/(N - n)$ when i and j belong to different groups (thin links). Our results are not limited to these two structures and apply to transitive $\mathcal{D}$ graphs in general. (**c**) and (**d**) are examples of other transitive dispersal graphs $\mathcal{D}$.

interaction partners, that is, as a measure of assortment. Hence, four parameters (N, d_{self}, e_{self}, $\overline{ed}$) summarize the dispersal and social structure of the population.

A two-step life-cycle and two payoff matrices. The evolution of the population is modelled using a Moran process. Each individual is either social (S) or non-social (NS). Between two time steps, exactly one individual dies and one individual reproduces, so that the size of the population remains N. The identities of the individuals who die and reproduce depend on the individuals' fecundity and survival potential, both being affected by social interactions and by the rules according to which the population is updated. We consider two classic updating rules[13]: death–birth (DB) and birth–death (BD) (these two rules are not restricted to specific population structures, unlike some others, such as budding[20], that are limited to deme-structured populations). In both cases, the first step (death in DB, birth in BD) involves choosing a first individual from all individuals in the population, while the second step (birth in DB, death in BD) involves only those individuals that are connected by dispersal to the site chosen in the first step (dispersal patterns being given by the $\mathcal{D}$ graph). While previous studies only considered effects of social interactions on one of the two steps, the framework we use here allows us to consider the general case where costs and benefits of social interactions can affect both steps, that is, both fecundity and survival. Since effects on fecundity with a BD updating are equivalent to effects on survival with a DB updating in the set of population structures that we consider[19], and *vice versa*, we will give all our heuristic explanations in terms of a DB updating.

The fecundity and survival potential of a given individual living at site i depend on its identity (social or non-social) and on the identities of the individuals it interacts with. To which extent an individual at site j interacts with the individual at i is determined by the e_{ji} term of the $\mathcal{E}$ graph. The effects of this interaction on the recipient's fecundity and survival potential are given by two general payoff matrices:

$$\mathbf{A}_{[1]} = \begin{pmatrix} a_{[1]} & b_{[1]} \\ c_{[1]} & d_{[1]} \end{pmatrix}$$

for the first step (that is, effects on survival with DB), and

$$\mathbf{A}_{[2]} = \begin{pmatrix} a_{[2]} & b_{[2]} \\ c_{[2]} & d_{[2]} \end{pmatrix}$$

for the second step (that is, effects on fecundity with DB).

In both matrices, the a (respectively b) terms refer to payoffs received by a social individual when interacting with another social individual (respectively a non-social), and the d (respectively c) terms refer to payoffs received by non-social individuals, when interacting with a non-social (respectively a social) individual. We use a dynamical system analysis based on moments (singlets, pairs, triplets of individuals and so on) of the distribution of social individuals in a structured population, and we assume weak selection, such that the fitness effects of interactions are small, but individuals are not necessarily phenotypically close[21]. We assume that mutations from one type to the other are rare; a new mutation only occurs after the previous one has been fixed or lost.

The scales of competition. The essential difference between the two steps of the process is the scale of competition (see equation (11), and Supplementary Methods), or, using a kin selection terminology, the identity of the secondary recipients[22] at each step. In a DB updating, an increase in the survival of an individual at site i indirectly harms all the individuals who can send offspring to site i, and the magnitude of this indirect effect on j is determined by d_{ji}. However, an increase in the fecundity of an individual i indirectly harms all other individuals j who would be competing with i for an empty site k, and the magnitude of this indirect effect on j is determined by $\sum_{k=1}^{N} d_{ik} d_{jk}$. Hence, competition in the first step is among all individuals that are one dispersal step away, while competition in the second step is among all individuals that are two dispersal steps away. These two different competition neighbourhoods are illustrated in Supplementary Fig. 1, in the case of a lattice-structured population. In other words, for both DB and BD updating rules, the first step, which involves choosing a first individual globally among all individuals of the population, results in a narrower competitive radius than the second step, in which another individual is chosen locally among the neighbours of the first individual[23]. Thus, whether social interactions affect the first or the second step results in a difference in the spatial scale over which social interactions affect competition. This difference turns out to be crucial for social evolution.

The condition for the evolution of social behaviour. We say that social behaviour evolves when the long-term frequency of social individuals is higher than the frequency of non-social individuals; or, equivalently in the limit of rare and symmetric mutations, when the probability of fixation of initially one social individual in a non-social population is higher than the converse probability of fixation: $\rho_S > \rho_{NS}$. Under our assumptions, we find

that this condition is satisfied when

$$
\begin{aligned}
\Big\{ &\left(1 + e_{\text{self}} - \tfrac{2}{N}\right)\left(a_{[1]} + a_{[2]} - d_{[1]} - d_{[2]}\right) \\
&+ \left(1 - e_{\text{self}}\right)\left(b_{[1]} + b_{[2]} - c_{[1]} - c_{[2]}\right) \\
&+ \left(d_{\text{self}} + \overline{\overline{ed}} - \tfrac{2}{N}\right)\left(a_{[2]} - d_{[2]}\right) + \left(d_{\text{self}} - \overline{\overline{ed}}\right)\left(b_{[2]} - c_{[2]}\right) \Big\} > 0
\end{aligned}
$$
(1)

This condition generalizes previous results[14,18] in two important ways: first, the effects of benefits and costs are not limited to fecundity but may also affect survival at the same time; second, interactions are not restricted to games with equal-gains-from-switching[24] where $a - c = b - d$. Equal-gains-from-switching occurs for instance in the Prisoner's Dilemma (PD, see below) but also, to first order, in any game with small phenotypic differences between individuals[25]. With such payoffs, it is possible to find explicit expressions for ρ_S and ρ_{NS} (see Fig. 2). In contrast, our equation (1) also includes synergistic enhancement or discounting.

Additional insights can be gained by defining an equivalent payoff matrix, $\tilde{\mathbf{A}}$, such that condition (1) can be rewritten as $(1 \; -1) \cdot \tilde{\mathbf{A}} \cdot (1 \; 1)^{\text{T}} > 0$. The equivalent matrix $\tilde{\mathbf{A}}$ can be written as

$$\tilde{\mathbf{A}} = \begin{pmatrix} \sigma_{[1]} a_{[1]} & b_{[1]} \\ c_{[1]} & \sigma_{[1]} d_{[1]} \end{pmatrix} + \xi \begin{pmatrix} \sigma_{[2]} a_{[2]} & b_{[2]} \\ c_{[2]} & \sigma_{[2]} d_{[2]} \end{pmatrix}.$$
(2)

For each step i of the life-cycle, $\sigma_{[i]}$ measures the amount of assortment due to the population structure, that is, how much individuals of the same type interact relative to individuals of different types[26] compared with how much they do in large
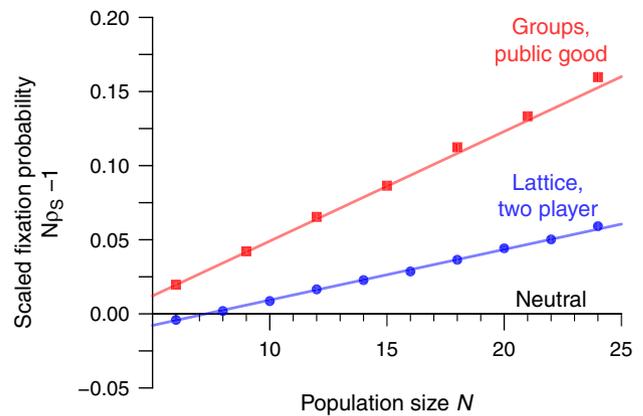
Figure 2 | Fixation probabilities. Comparison of the analytical fixation probability ρ_S with equal-gains-from-switching (solid coloured lines) to the frequency of fixation in numerical simulations (circles and squares), for different values of the total population size N and two classic population structures; results are scaled relative to $1/N$, the neutral expectation $((\rho_S - 1/N)/(1/N) = N\rho_S - 1)$. Two population structures were simulated: in blue, a lattice with $k = 3$ neighbours, two-player games $\left[e_{\text{self}} = d_{\text{self}} = 0, \overline{\overline{ed}} = 1/k \right]$; in red, groups of $n = 3$ individuals with public-good games $\left[e_{\text{self}} = 1/n \text{ and } d_{\text{self}} = (1-m)/n = \overline{\overline{ed}}; m = 0.1 \right]$. Fixation probability: denoting $B_{[i]} = c_{[i]} - d_{[i]}$ and $C_{[i]} = d_{[i]} - b_{[i]}$, the approximate fixation probability is $\big[B_{[1]}(N e_{\text{self}} - 1) - C_{[1]}(N - 1) + B_{[2]}\left(N e_{\text{self}} + N\overline{\overline{ed}} - 2\right) - C_{[2]}(N + N d_{\text{self}} - 2) \big]/(2N)$. Simulation data: 10^7 runs, $\pm 95\%$ confidence interval (CI) (behind the dots). Parameters: $b_{[1]} = 1\omega$, $c_{[1]} = 0.1\omega$, $b_{[2]} = 10\omega$, $c_{[2]} = 0.5\omega$, $\omega = 0.0025$.

well-mixed populations (in which $\sigma_{[1]} = \sigma_{[2]} = 1$):

$$\sigma_{[1]} = \frac{1 + e_{\text{self}} - 2/N}{1 - e_{\text{self}}},$$

$$\sigma_{[2]} = \frac{1 + e_{\text{self}} + d_{\text{self}} + \overline{\overline{ed}} - 4/N}{1 - e_{\text{self}} + d_{\text{self}} - \overline{\overline{ed}}}. \tag{3}$$

The quantity ξ in equation (2a) measures the relative importance of the second step of the life-cycle, compared with the first:

$$\xi = \frac{1 - e_{\text{self}} + d_{\text{self}} - \overline{\overline{ed}}}{1 - e_{\text{self}}}. \tag{4}$$

In a large well-mixed population, $\xi = 1$.

Contrasting the conditions on each step. By definition of the dispersal graph $\mathcal{D}$, the offspring of an individual are located one dispersal step away, which happens to correspond to the competitive radius during the first step of the Moran process. Individuals are therefore directly competing against their offspring, and the detrimental effects of kin competition exactly cancel the social benefits of living next to related individuals[27,28]. As a result, population structure barely has any effect on the evolution of social behaviour if social interactions affect the first step of the process (see $\sigma_{[1]}$ in equation (3)), compared with large well-mixed

populations. In this case, the relative effect of population structure is limited to the benefits a social individual provides to itself (e_{self}), while small population sizes reduce same-type interactions: $\sigma_{[1]}$ is smaller when N is smaller. This also confirms that even with synergistic effects in the first step ($a_{[1]} - c_{[1]} \neq b_{[1]} - d_{[1]}$), spatial structure does not facilitate the evolution of altruism[29] compared with well-mixed populations. Equation (1) also confirms that in the absence of kin discrimination or synergistic effects ($a_{[1]} - c_{[1]} = b_{[1]} - d_{[1]}$), the evolution of spite ($c_{[1]} - d_{[1]} < 0$) requires small population sizes (small N), and limited self-interactions ($e_{\text{self}} \to 0$).

In contrast, population structure is of crucial importance for the evolution of social behaviour whenever social interactions affect the second step of the process ($\sigma_{[2]}$ in equation (3) and ξ in equation (4)). This is because the radius of the competitive circle is wider at the second step (two dispersal steps away): individuals are therefore competing against less related individuals, on average, than at the first step[23]. This observation had been made with specific models. For instance, conditions for general games and regular graphs of degree k (and large N) were derived by Ohtsuki *et al.*[13]; these conditions reduced to $(c_{[2]} - d_{[2]})/(d_{[2]} - b_{[2]}) > k$ when $a_{[2]} - c_{[2]} = b_{[2]} - d_{[2]}$. For this restricted set of games, Taylor *et al.*[14] derived a generalization that extended to small populations sizes and covered a broad range of structures, including weighted graphs and distinct interaction $\mathcal{E}$
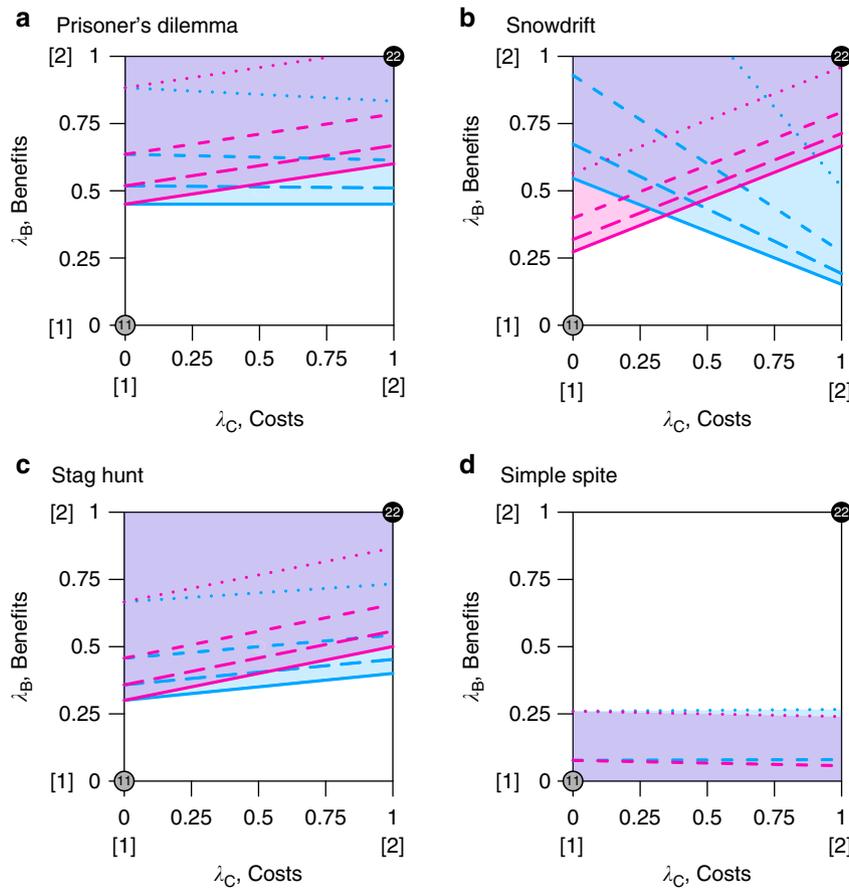
Figure 3 | Evolution of social behaviour under different allocations of benefits and costs. Comparing the conditions for the evolution of social behaviour (altruism in (**a**–**c**), spite in (**d**)), depending on whether benefits and costs affect the first or second step of the process, for four classical games. Results are shown for different population sizes (full line: $N \to \infty$, large dashed: $N = 60$, dashed: $N = 24$, dotted: $N = 12$) and two structures: in magenta, $d_{\text{self}} = 1/3$, in cyan $d_{\text{self}} = 0$; in both cases $e_{\text{self}} = 0$ and $\overline{\overline{ed}} = 1/3$. Social individuals are favoured ($\rho_S > \rho_{NS}$) in the shaded areas. The grey (benefits and costs on the first step exclusively) and black dots (benefits and costs on the second step exclusively) in the corners correspond to the only parameter combinations that have been analysed previously in this context[13,14,18,19,30]. Parameters: (**a**) Prisoner's dilemma, $c'/b' = 0.15$; (**b**) Snowdrift, $c'/(2b' - c') = 0.65$; (**b**) Stag hunt, $(b' + c')/2b' = 0.6$; (**d**) Simple spite, $c'/b' = 0.02$.

and dispersal $\mathcal{D}$ graphs. In particular, their derivation also included the $(c_{[2]} - d_{[2]})/(d_{[2]} - b_{[2]}) > hg/l$ rule for regular graphs[30], where g and h are the number of neighbours on the g and h graphs, respectively, and l the number of overlapping edges. Our formula (1) not only confirms and generalizes these findings but also offers additional insights, and we illustrate the implications of our results in Fig. 3.

Combining the two steps. We now consider four classical games that can be expressed in terms of benefits and costs (that is, two-parameter games), and we assume that social individuals can allocate these benefits and costs in different proportions to fecundity and survival effects, or more generally to the first and second step in the life-cycle. Then the benefits and costs affecting the first step (fecundity in BD, survival in DB) are $B_{[1]} = b'(1 - \lambda_b)$ and $C_{[1]} = c'(1 - \lambda_c)$, respectively, while the benefits and costs affecting the second step (survival in BD, fecundity in DB) are $B_{[2]} = b'(1 - \lambda_b)$ and $C_{[2]} = c'(1 - \lambda_c)$, respectively. Two of the games that we consider have equal-gains-from-switching: a Prisoner's dilemma (PD) with payoff matrix on step k given by $M_{[k]}^{\mathrm{PD}} = \begin{pmatrix} B_{[k]} - C_{[k]} & -C_{[k]} \\ B_{[k]} & 0 \end{pmatrix}$, and the equivalent spite version with negative benefits (SS), whose payoff matrices are $M_{[k]}^{\mathrm{SS}} = \begin{pmatrix} -B_{[k]} - C_{[k]} & -C_{[k]} \\ -B_{[k]} & 0 \end{pmatrix}$. We also consider a Snowdrift game (SD) with payoff matrices $M_{[k]}^{\mathrm{SD}} = \begin{pmatrix} B_{[k]} - C_{[k]}/2 & -C_{[k]} \\ B_{[k]} & 0 \end{pmatrix}$ and a Stag hunt game (SH) $M_{[k]}^{\mathrm{SH}} = \begin{pmatrix} 2B_{[k]} - C_{[k]} & -C_{[k]} \\ B_{[k]} & B_{[k]} \end{pmatrix}$. For each game, we assess how the allocation of benefits and costs on either step of the life-cycle affects the evolution of social behaviour, and do so for two population structures that differ by whether an individual can be replaced by their offspring or not (that is, whether $d_{\mathrm{self}} > 0$ or $d_{\mathrm{self}} = 0$). The results are illustrated in Fig. 3. We find that the optimal allocation of the benefits only depends on whether these correspond to altruism (positive benefits) or spite (negative benefits). Altruism (Fig. 3a–c) is most favoured if benefits are allocated to the second step of the process, which gives more weight to interactions of individuals of the same type ($\sigma_{[2]} > \sigma_{[1]}$). Spite, on the contrary, is more likely to evolve when the (negative) benefits affect the first step rather than the second one (Fig. 3d), and requires small population sizes. Let us now consider costs. In our examples, the optimal cost allocation depends on both the type of game, the size of the population (N) and on the fraction of an individual's propagules that remain at the exact same site (d_{self}). For example, for a sufficiently large d_{self} (magenta lines in Fig. 3), altruism is most favoured if costs affect the first step of the process. With a low d_{self}, however ($d_{\mathrm{self}} = 0$ for the cyan lines in Fig. 3), the allocation of the cost may not matter (for example, Fig. 3a,d, Prisoner's dilemma with large (N), or better be on the second step (for example, Fig. 3b, Snowdrift), or finally better be on the first step (for example, Fig. 3c, Stag hunt).

While we just described cases where social interactions are of the same type on both steps, our framework also allows for the consideration of mixed cases. Focusing on games with equal-gains–from-switching, with $e_{\mathrm{self}} = 0$ (like in Fig. 3a,d), our results suggest that, behaviours that are spiteful in the first step and altruistic in the second will be favoured by selection.

Discussion

The simplicity and generality of our condition for the evolution of social behaviour hinges on two standard and widely used

assumptions of weak selection and constant population size. Extending our results to unsaturated populations[17,31], as well as to all possible network structures, is an important future challenge. Meanwhile, our results unify and generalize a great number of existing studies on the evolution of altruism or spite in structured populations, and yield new insights when social behaviour can affect both fecundity and survival. In particular, the results highlight the crucial feature determining the outcome of social evolution: the evolution of social behaviour is determined by the scale at which social interactions affect competition.

Methods

Notation. The population lives in an environment with N sites, labelled $\{1, \ldots, N\}$. An indicator variable $X_i(t)$ gives the occupation of site i at time t: 1 (respectively 0) means that the site is occupied by a social individual (respectively non-social). Notation '$^-$' denotes a population average, and $\mathrm{var}_{\mathrm{S}}$ a population variance: $\bar{X}(t) = \sum_{i=1}^{N} X_i(t)/N$ and $\mathrm{var}_{\mathrm{S}}(X(t)) = \sum_{i=1}^{N} (X_i(t) - \bar{X}(t))^2/N = \overline{X^2}(t) - \bar{X}^2(t)$. The index $_{\mathrm{S}}$ in the population variance $\mathrm{var}_{\mathrm{S}}$ is here to distinguish this variance from the variance of the state of sites ($\mathrm{var}[X_i(t)] = \mathbb{E}_t[X_i^2] - \mathbb{E}_t[X_i]^2$); in the same way there is a distinction between average ($^-$) and expectation ($\mathbb{E}_t[]$). We denote by

$$p_i(t) = \mathbb{E}_t[X_i] = \mathbb{P}(X_i(t) = 1) \qquad (5)$$

the expectation of the state of the individual at site i and at time t; a vector $\mathbf{p}(t)$ groups the expected state of all sites, and $p(t) = \overline{p_i(t)}$ is the expected frequency of social individuals in the population at time t. Because of population structure, we also have to take into account the dynamics of pairs and triplets of individuals. We denote by

$$P_{ij}(t) = \mathbb{E}_t[X_i X_j] = \mathbb{P}(X_i(t) = 1, X_j(t) = 1) \qquad (6)$$

the expectation of the state of pairs of sites, and group them in a matrix $\mathbf{P}(t)$. We note that the notion of relatedness classically used in kin selection studies is different and involves probabilities of identity in state or probabilities of identity by descent, without conditioning on the type of the individuals: $G_{ij} = \mathbb{P}(X_i(t) = X_j(t))$ Relatedness R_{ij} is often defined as a standardized measure of identity[19]:

$$R_{ij} = \frac{G_{ij} - \overline{\overline{G}}}{1 - \overline{\overline{G}}}, \qquad (7)$$

where $\overline{\overline{G}} = \sum_{i=1}^{N} \sum_{j=1}^{N} G_{ij}/N^2$. Our measure of assortment $P_{ij}(t)$ can therefore be expressed as a function of this measure of relatedness R_{ij}:

$$P_{ij}(t) = \mathbb{E}_t[X_i X_j] = (R_{ij}(t) - 1)\mathbb{E}_t[\mathrm{var}_{\mathrm{S}} X] + p(t). \qquad (8)$$

We also need to account for associations between triplets of individuals, and we denote by

$$\Pi_{ijk}(t) = \mathbb{E}_t[X_i X_j X_k] = \mathbb{P}(X_i(t) = 1, X_j(t) = 1, X_k(t) = 1) \qquad (9)$$

the expectation of the state of triplets of sites, and group them in a three-dimensional array $\mathbf{\Pi}(t)$.

The dispersal and social structures of the population are represented by two graphs $\mathcal{D}$ and $\mathcal{E}$. We denote by d_{ij} the fraction of the offspring of the individual living at site i that is sent to site j, and by e_{ij} the strength of the interaction from i to j; all these parameters are grouped in two matrices $\mathbf{D}$ and $\mathbf{E}$. The dispersal graph is assumed to be symmetric, so that $d_{ij} = d_{ji}$ (or equivalently $\mathbf{D} = \mathbf{D}^{\mathrm{T}}$, where $^{\mathrm{T}}$ denotes transposition); the dispersal graph is also assumed to be transitive, so that the dispersal structure looks the same from every site[14] (note that the transitivity assumption is not required to derive equation (11), but it is needed to obtain explicit expressions for the expected state of pairs of individuals (equation (12))).

The fecundity and survival of individuals in the population are affected by pairwise interactions with other individuals, described by the interaction graph $\mathcal{E}$. For our derivation, it is more convenient to rewrite the payoff matrices $\mathbf{A}_{[1]}$ and $\mathbf{A}_{[2]}$ (whose expressions are given in the main text) as follows:

$$\mathbf{A}_{[i]} = \begin{pmatrix} b'_{[i]} - c'_{[i]} + d'_{[i]} & -c'_{[i]} \\ b'_{[i]} & 0 \end{pmatrix}, \qquad (10)$$

with $b'_{[i]} = c_{[i]} - d_{[i]}$, $c'_{[i]} = d_{[i]} - b_{[i]}$, and $d'_{[i]} = a_{[i]} - b_{[i]} - c_{[i]} + d_{[i]}$. The two formulations are equivalent[32]. The effects of pairwise interactions add up, and we assume that their effects on fecundity and survival are weak, of order $\omega \ll 1$: in other words, selection is weak. Note that this type of weak selection differs from weak selection due to small phenotypic differences ('δ-weak selection'[21]) classically used in kin selection models.

Expected change in the frequency of social individuals. We first derive a general equation for the change in the frequency of social individuals, under weak selection and for a symmetric dispersal graph ($\mathbf{D} = \mathbf{D}^{\mathrm{T}}$). We show in the Supplementary Methods how this expression relates to the Price equation[33]. Denoting by $\mathrm{Tr}(\mathbf{M})$ the trace, that is, the sum of diagonal elements, of a matrix $\mathbf{M}$,

the expected change in the frequency of social individuals in the population is (details of the calculations are presented in the Supplementary Methods):

$$
\Delta p(t) = \frac{\omega}{N^2} \Big[\overbrace{(\mathrm{Tr}(\mathbf{E}^{\mathrm{T}} \cdot \mathbf{P})}^{\text{direct effects}} \quad \overbrace{- \mathrm{Tr}(\mathbf{E}^{\mathrm{T}} \cdot \mathbf{P} \cdot \mathbf{D}))}^{\text{competition = secondary effects}} b'_{[1]}
$$
$$
- (\mathrm{Tr}(\mathbf{P}) \qquad - \mathrm{Tr}(\mathbf{P} \cdot \mathbf{D})) c'_{[1]}
$$
$$
+ (\mathrm{Tr}(\mathbf{E}^{\mathrm{T}} \cdot \mathbf{P}) \quad - \mathrm{Tr}(\mathbf{E}^{\mathrm{T}} \cdot \Pi \cdot \mathbf{D})) d'_{[1]}
$$
$$
+ (\mathrm{Tr}(\mathbf{E}^{\mathrm{T}} \cdot \mathbf{P}) \quad - \mathrm{Tr}(\mathbf{E}^{\mathrm{T}} \cdot \Pi \cdot \mathbf{D} \cdot \mathbf{D})) b'_{[2]}
$$
$$
+ (\mathrm{Tr}(\mathbf{P}) \qquad - \mathrm{Tr}(\mathbf{P} \cdot \mathbf{D} \cdot \mathbf{D})) c'_{[2]}
$$
$$
+ (\mathrm{Tr}(\mathbf{E}^{\mathrm{T}} \cdot \mathbf{P}) \quad - \mathrm{Tr}(\mathbf{E}^{\mathrm{T}} \cdot \Pi \cdot \mathbf{D} \cdot \mathbf{D})) d'_{[2]} \Big] \tag{11}
$$

Supplementary Figure 2 (direct effects) and Supplementary Fig. 3 (secondary effects) illustrate the different terms of equation (11). The first three lines correspond to the first step (fecundity effects under BD, survival effects under DB), the last three lines to the second step (survival effects under BD, fecundity effects under DB). For each step, we can distinguish between direct effects (first column) and competition terms (second column). These competition terms correspond to secondary effects in kin selection models[22] or to circles of compensation[23], and differ among the two steps: the competitive radius includes individuals one dispersal step away in the first step ($\mathbf{P} \cdot \mathbf{D}$ terms), and two dispersal steps away in the second step ($\mathbf{P} \cdot \mathbf{D} \cdot \mathbf{D}$ terms). This equation is dynamically not closed, because it depends on higher moments such as pairs ($\mathbf{P}$) and triplets (Π) of social individuals.

Evaluating the moments. We show that the dynamics of pairs and triplets occur at a much faster time scale than the dynamics of the average frequency, so that we can evaluate them using a separation of time scales[34] or quasi-equilibrium approximation.

Using the fact that the dispersal graph $\mathcal{D}$ is transitive[14,23] and that selection is weak, we obtain the following equalities for the pairs ($\mathbf{1}_{N,N}$ is a N-by-N matrix containing only 1s, and $\mathbf{I}_N$ is the identity matrix):

$$
\begin{aligned}
\mathbf{P} - \mathbf{P} \cdot \mathbf{D} &= -\phi(t) \mathbf{1}_{N,N} + N\phi(t) \mathbf{I}_N, \\
\mathbf{P} - \mathbf{P} \cdot \mathbf{D} \cdot \mathbf{D} &= -2\phi(t) \mathbf{1}_{N,N} + N\phi(t)(\mathbf{I}_N + \mathbf{D}),
\end{aligned} \tag{12}
$$

where $\phi(t) = (N/(N-1)) \mathbb{E}_t[\mathrm{var}_S X]$ (details of the calculations are in the Supplementary Methods). Note that we did not use the pair approximation[35] to derive equation (12).

For triplets of individuals, we use recent results[26,32] showing how to express terms with triplets as functions of pairs plus a frequency-dependent term scaled by a factor $\alpha_{[i]}$ for each step i (see Supplementary Methods for details). The $\alpha_{[i]}$ factors will remain implicit, but they will vanish in the final condition for the evolution of social behaviour.

Explicit dynamics. We use the expressions we derived for pairs (equation (12)) and triplets of individuals back in the equation for the frequency dynamics (equation (11)), thus arriving at a closed dynamical system, given below in equation (16). We denote by $d_{\mathrm{self}} = d_{ii}$, the fraction of propagules that remain at their parent's site (the same for all sites i on a transitive graph), by $e_{\mathrm{self}} = (1/N) \sum_{l=1}^{N} e_{ll}$ the average interaction with oneself, and finally a compound parameter summarizing the dispersal and interaction graphs (recall that $d_{kl} = d_{lk}$),

$$
\overline{\overline{ed}} = \frac{1}{N} \sum_{l=1}^{N} \sum_{k=1}^{N} e_{lk} d_{kl} = \frac{1}{N} \sum_{l=1}^{N} \sum_{k=1}^{N} e_{lk} d_{lk} = \frac{1}{N} \mathrm{Tr}(\mathbf{E}^{\mathrm{T}} \cdot \mathbf{D}) \tag{13}
$$

that can be interpreted as the average chance of receiving benefits (e_{lk}) from a site where offspring have been sent (d_{kl}). We then define two compound parameters, $\sigma_{\mathcal{DE}}$ and $\tau_{\mathcal{DE}}$, that depend on the population structure and the different payoffs:

$$
\begin{aligned}
\sigma_{\mathcal{DE}} = \omega \Big(&+ b'_{[1]} \frac{N e_{\mathrm{self}} - 1}{N-1} - c'_{[1]} \\
&+ d'_{[1]} \Big(\frac{N + N e_{\mathrm{self}} - 2}{2(N-1)} - \frac{\alpha_{[1]}}{2} \Big) \\
&+ b'_{[2]} \frac{N e_{\mathrm{self}} + N\overline{\overline{ed}} - 2}{N-1} - c'_{[2]} \frac{N + N d_{\mathrm{self}} - 2}{N-1} \\
&+ d'_{[2]} \Big(\frac{N + N d_{\mathrm{self}} + N e_{\mathrm{self}} + N\overline{\overline{ed}} - 4}{2(N-1)} - \frac{\alpha_{[2]}}{2} \Big) \Big),
\end{aligned} \tag{14}
$$

and

$$
\tau_{\mathcal{DE}} = \omega \Big(d'_{[1]} \alpha_{[1]} + d'_{[2]} \alpha_{[2]} \Big). \tag{15}
$$

With these definitions, we find that the expected change in the frequency of social individuals can be written as follows:

$$
\Delta p(t) = \underbrace{(\sigma_{\mathcal{DE}} + \tau_{\mathcal{DE}} p(t))}_{s_{\mathcal{DE}}(p(t))} \times \mathbb{E}[\mathrm{var}_S(p(t))] \tag{16}
$$

In other words, the structured population behaves as a well-mixed one, under linear frequency-dependent selection given by $s_{\mathcal{DE}}(p(t))$.

Fixation probabilities. The fixation probability of a single mutant under a Moran process with a linear frequency-dependent selection coefficient is a classical result[36–38]. Accordingly, the fixation probability of a social mutant in a population of non-social individuals, when selection is weak, can be approximated as

$$
\rho_S \approx \frac{1}{N} + \frac{N-1}{2N} \Big(\sigma_{\mathcal{DE}} + \tau_{\mathcal{DE}} \frac{N+1}{3N} \Big) \tag{17}
$$

Reciprocally, the fixation probability of initially one non-social individual in a population of social individuals is

$$
\rho_{NS} \approx \frac{1}{N} + \frac{N-1}{2N} \Big(-(\sigma_{\mathcal{DE}} + \tau_{\mathcal{DE}}) + \tau_{\mathcal{DE}} \frac{N+1}{3N} \Big) \tag{18}
$$

We define the evolutionary success of social individuals by the condition

$$
\rho_S > \rho_{NS} \tag{19}
$$

Using equations (17) and (18), this condition becomes $2\sigma_{\mathcal{DE}} + \tau_{\mathcal{DE}} > 0$. Finally, using the definitions (14) and (15) we obtain equation (1) in the main text.

Simulations. Stochastic simulations were coded in C; the population was updated following a Moran process as described in the main text, and the simulation stopped when the social trait was either fixed or lost. Fixation probabilities, with initially only one social individual in the population, were estimated after 10^7 runs for each parameter combination. The simulation scripts are available on Dryad: doi:10.5061/dryad.r28qk.

References

1. Nowak, M. A., Tarnita, C. E. & Wilson, E. O. The evolution of eusociality. *Nature* **466,** 1057–1062 (2010).
2. Abbot, P. *et al.* Inclusive fitness theory and eusociality. *Nature* **471,** E1–E4 (2011).
3. Lion, S., Jansen, V. A. & Day, T. Evolution in structured populations: beyond the kin versus group debate. *Trends. Ecol. Evol.* **26,** 193–201 (2011).
4. Fletcher, J. A. & Doebeli, M. A simple and general explanation for the evolution of altruism. *Proc. R. Soc. B* **276,** 13–19 (2009).
5. Gardner, A. & West, S. A. Spite and the scale of competition. *J. Evol. Biol.* **17,** 1195–1203 (2004).
6. Hamilton, W. The genetical evolution of social behaviour. i. *J. Theor. Biol.* **7,** 1–16 (1964).
7. Jansen, V. A. A. & van Baalen, M. Altruism through beard chromodynamics. *Nature* **440,** 663–666 (2006).
8. Nowak, M. A. Five rules for the evolution of cooperation. *Science* **314,** 1560–1563 (2006).
9. Lehmann, L. & Keller, L. The evolution of cooperation and altruism-a general framework and a classification of models. *J. Evol. Biol.* **19,** 1365–1376 (2006).
10. Moran, P. A. P. Random processes in genetics. *Math. Proc. Cambridge Phil. Soc.* **54,** 60–71 (1958).
11. Taylor, P. D. & Irwin, A. J. Overlapping generations can promote altruistic behavior. *Evolution* **54,** 1135–1141 (2000).
12. Lehmann, L., Keller, L. & Sumpter, D. J. T. The evolution of helping and harming on graphs: the return of the inclusive fitness effect. *J. Evol. Biol.* **20,** 2284–2295 (2007).
13. Ohtsuki, H., Hauert, C., Lieberman, E. & Nowak, M. A. A simple rule for the evolution of cooperation on graphs and social networks. *Nature* **441,** 502–505 (2006).
14. Taylor, P. D., Day, T. & Wild, G. Evolution of cooperation in a finite homogeneous graph. *Nature* **447,** 469–472 (2007).
15. Zukewich, J., Kurella, V., Doebeli, M. & Hauert, C. Consolidating birth-death and death-birth processes in structured populations. *PLoS One* **8,** e54639 (2013).
16. Nakamaru, M. & Iwasa, Y. The coevolution of altruism and punishment: role of the selfish punisher. *J. Theor. Biol.* **240,** 475–488 (2006).
17. Lion, S. & Gandon, S. Life history, habitat saturation and the evolution of fecundity and survival altruism. *Evolution* **64,** 1594–1606 (2010).
18. Taylor, P. Birth-death symmetry in the evolution of a social trait. *J. Evol. Biol.* **23,** 2569–2578 (2010).
19. Taylor, P., Lillicrap, T. & Cownden, D. Inclusive fitness analysis on mathematical groups. *Evolution* **65,** 849–859 (2011).
20. Lehmann, L. & Rousset, F. How life history and demography promote or inhibit the evolution of helping behaviours. *Phil. Trans. R. Soc. B* **365,** 2599–2617 (2010).
21. Wild, G. & Traulsen, A. The different limits of weak selection and the evolutionary dynamics of finite populations. *J. Theor. Biol.* **247,** 382–390 (2007).
22. West, S. A. & Gardner, A. Altruism, spite, and greenbeards. *Science* **327,** 1341–1344 (2010).
23. Grafen, A. & Archetti, M. Natural selection of altruism in inelastic viscous homogeneous populations. *J. Theor. Biol.* **252,** 694–710 (2008).

24. Nowak, M. & Sigmund, K. The evolution of stochastic strategies in the prisoner's dilemma. *Acta. Appl. Math.* **20,** 247–265 (1990).
25. Taylor, P. D., Day, T. & Wild, G. From inclusive fitness to fixation probability in homogeneous structured populations. *J. Theor. Biol.* **249,** 101–110 (2007).
26. Tarnita, C. E., Ohtsuki, H., Antal, T., Fu, F. & Nowak, M. A. Strategy selection in structured populations. *J. Theor. Biol.* **259,** 570–581 (2009).
27. Taylor, P. D. Inclusive fitness in a homogeneous environment. *Proc. Rl. Soc. Lond. B* **249,** 299–302 (1992).
28. Rodrigues, A. M. & Gardner, A. Evolution of helping and harming in heterogeneous populations. *Evolution* **66,** 2065–2079 (2012).
29. Ohtsuki, H. Does synergy rescue the evolution of cooperation? an analysis for homogeneous populations with non-overlapping generations. *J. Theor. Biol.* **307,** 20–28 (2012).
30. Ohtsuki, H., Nowak, M. A. & Pacheco, J. M. Breaking the symmetry between interaction and replacement in evolutionary dynamics on graphs. *Phys. Rev. Lett.* **98,** 108106 (2007).
31. Alizon, S. & Taylor, P. Empty sites can promote altruistic behavior. *Evolution* **62,** 1335–1344 (2008).
32. Taylor, P. & Maciejewski, W. An inclusive fitness analysis of synergistic interactions in structured populations. *Proc. R. Soc. B* **279,** 4596–4603 (2012).
33. Price, G. R. Selection and covariance. *Nature* **227,** 520–521 (1970).
34. Law, R. & Dieckmann, U. *The Geometry of Ecological Interactions: Simplifying Spatial Complexity* 252–270 (Cambridge University Press, 2000).
35. Matsuda, H., Ogita, N., Sasaki, A. & Satō, K. Statistical mechanics of population. *Prog. Theor. Phys.* **88,** 1035–1049 (1992).
36. Nowak, M., Sasaki, A., Taylor, C. & Fudenberg, D. Emergence of cooperation and evolutionary stability in finite populations. *Nature* **428,** 646–650 (2004).
37. Durrett, R. *Probability Models for DNA Sequence Evolution, vol. XII of Probability and Its Applications* 2nd edn, Ch. 1 (Springer, New York, 2008).
38. Huang, W. & Traulsen, A. Fixation probabilities of random mutants under frequency dependent selection. *J. Theor. Biol.* **263,** 262–268 (2010).

Acknowledgements

Author contributions

F.D. designed and analysed the model; all authors wrote the manuscript; F.D. wrote the Supplementary Information.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Débarre, F. *et al.* Social evolution in structured populations. *Nat. Commun.* 5:3409 doi: 10.1038/ncomms4409 (2014).

SUPPLEMENTARY FIGURES
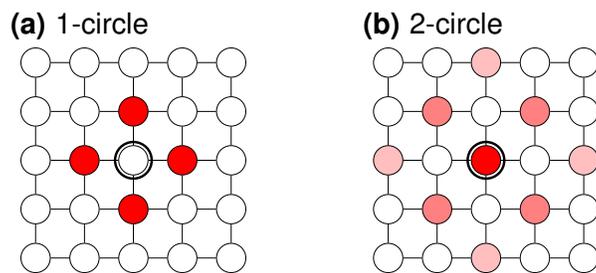
(a) 1-circle   **(b)** 2-circle

Supplementary Figure 1: The different competition neighbourhoods (1-circle in (a), 2-circle in (b)), as seen by a focal individual (with the double outline), in the specific example of a lattice-structured population where $k = 4$ neighbours. The shaded dots are the individuals that the focal is competing against, and the intensity of the shading reflects the strength of competition (white: no competition).

$$\beta_S = \frac{\omega}{N^2} \sum_{ijl} \left[\quad - \quad \right]$$

$$\gamma_S = \frac{\omega}{N^2} \sum_{ijl} \left[\quad + \quad \right]$$

$$\delta_S = \frac{\omega}{N^2} \sum_{ijl} \left[\quad \right]$$

Supplementary Figure 2: Schematic description of the different terms for the first step (survival) of the process. Orange nodes are occupied by social individuals, light gray nodes by non-social individuals. The node with the thick edge is the node that is potentially going to be replaced. Full arrows represent dispersal patterns (d), dashed arrows represent social interactions (e)

$$\beta_S = \frac{\omega}{N^2} \sum_{ijkl} \left[\quad - \quad \right]$$

$$\gamma_F = \frac{\omega}{N^2} \sum_{ijkl} \left[\quad + \quad \right]$$

$$\delta_F = \frac{\omega}{N^2} \sum_{ijkl} \left[\quad \right]$$

Supplementary Figure 3: Schematic description of the different terms for the second step of the process. The legend is the same as in Supplementary Fig. 2, and white nodes correspond to any type of individual (social or non social).

SUPPLEMENTARY METHODS

We first introduce notation and main assumptions. Then, we derive an equation for $\Delta p(t)$, the change in the expected frequency of social individuals (*i.e.*, helpers) in the population. $\Delta p(t)$ depends on the frequency of social individuals in the population, $p(t)$, but also on higher moments such as pairs and triplets of social individuals. We then evaluate these higher moments, where we write the equations for their dynamics and show that these occur on a faster time scale than the dynamics of the global frequency of social individuals. Using a separation of time scales and a mean-field relaxation approximation, we are able to express pairs and triplets of social individuals as functions of the global frequency of social individuals. We then use these expressions and manage to express the change in the expected frequency of social individuals $\Delta p(t)$ as a function of $p(t)$ only; the system is now closed, because we have as many equations as variables. Then, we estimate the fixation probability of a single "social" mutation. Finally, we illustrate our results with four specific games.

Notation and assumptions

Population

We model a population living in an environment with exactly N sites, each site hosting at most one individual. We assume that more offspring are produced than there are sites that could host them, so that the population is saturated. As a result, there are no empty sites, and each site contains exactly one individual (*i.e.*, the population is inelastic). There are two types of individuals in the population: social individuals, also referred to as helpers (S) and non social individuals, also referred to as defectors (NS). Social individuals provide benefits to others (potentially including themselves; the benefits can be negative, in the case of spiteful behaviour), but suffer from costs.

We denote by $X_i(t)$ the random variable describing the type of individual present at the ith site ($i \in \{1, \ldots, N\}$) at time t. For all i, $X_i(t)$ is an indicator variable, and takes values:

$$x_i = \begin{cases} 1 & \text{if at } t, \text{ the individual at site } i \text{ is of type S} \\ 0 & \text{if at } t, \text{ the individual at site } i \text{ is of type NS} \end{cases}. \tag{1}$$

We denote by $p_i(t) = \mathbb{E}_t[X_i]$ the expectation of the state of the individual at site i and at time t. Because $X_i(t)$ is an indicator variable, we have

$$p_i(t) = \mathbb{E}_t[X_i] = \mathbb{P}(X_i(t) = 1). \tag{2}$$

In the course of our calculations, we will also need the expectation of the states of pairs and triplets of sites, which we will denote as follows:

$$P_{ij}(t) = \mathbb{E}_t \left[X_i X_j \right] = \mathbb{P} \left(X_i(t) = 1, X_j(t) = 1 \right), \tag{3}$$

$$\Pi_{ijk}(t) = \mathbb{E}_t \left[X_i X_j X_k \right] = \mathbb{P} \left(X_i(t) = 1, X_j(t) = 1, X_k(t) = 1 \right). \tag{4}$$

At t, the state of the population is described by a random vector $X(t) = \{X_1(t), \ldots, X_N(t)\}$, and we write $x = \{x_1, \ldots, x_N\}$ the value of $X(t)$ for one realization of the process. Note that at each time t, there are 2^N possible configurations of the population, with different probabilities.

The notation "‾" denotes an average across the whole population:

$$\overline{x} = \sum_{i=1}^{N} \frac{x_i}{N}; \tag{5}$$

The quantity $\overline{x}$ is also the frequency of individuals of type S in the population for one given realization of the process[1]. Finally, the expectation of the frequency of social individuals at time t is denoted by $p(t)$:

$$p(t) = \mathbb{E}_t \left[\overline{X} \right] = \sum_{i=1}^{N} \frac{p_i(t)}{N} \quad \text{[because of linearity of } \mathbb{E}\text{]}. \tag{6}$$

Interactions between individuals

Definitions

Two graphs, $\mathscr{D}$ and $\mathscr{E}$, describe the dispersal and interaction patterns between the N sites, respectively. Dispersal refers to where an individual's offspring can be sent, and interaction refers to to whom benefits are given. Using the same notations as in Ref.[1], we denote by d_{ij} the fraction of the offspring of the individual living at site i that is sent to site j; we denote by e_{ij} the strength of the interaction from i to j, measured relative to all interactions strengths received by j. With these definitions of d_{ij} and e_{ij}, we have:

$$\forall i, \quad \sum_{j=1}^{N} d_{ij} = 1, \tag{7}$$

$$\forall j, \quad \sum_{i=1}^{N} e_{ij} = 1. \tag{8}$$

---

[1] We will see in the course of the derivation of the results that indicator variables (that take values in $\{0, 1\}$) are very useful and simplify a lot our calculations. We note however that indicator variables can only be used when there are only two possible states in the population (here, S and NS), and that the situation would be more complicated with more states, for instance if there were empty sites.

These two conditions mean that in a monomorphic population (*i.e.*, when only one type of individual is present), all individuals produce the same number of offspring, and receive the same amount of benefits from other sites. We note that we can write both graphs as matrices:

$$\mathbf{D} = \{d_{ij}\}_{\substack{i \in \{1,\ldots,N\} \\ j \in \{1,\ldots,N\}}}, \tag{9}$$

$$\mathbf{E} = \{e_{ij}\}_{\substack{i \in \{1,\ldots,N\} \\ j \in \{1,\ldots,N\}}}. \tag{10}$$

This matrix notation will help us write expressions in a more compact way.

Assumptions on the $\mathscr{D}$ graph

Symmetric dispersal We further assume that, for an individual at node i, the chance of receiving propagules from node j is the same as the chance of sending propagules to j, so that:

$$\forall \{i, j\} \quad d_{ji} = d_{ij}, \tag{11}$$

which implies in particular that

$$\forall i, \quad \sum_j d_{ji} = 1. \tag{12}$$

This means that the $\mathbf{D}$ matrix describing the $\mathscr{D}$ graph is symmetric ($\mathbf{D}^T = \mathbf{D}$). This assumption is needed to simplify the general expression of the change in the global frequency of social individuals (section "*Equation for the change of the frequency of social individuals*"). When talking about graphs, the vertices are often referred to as "nodes"; each node corresponds to one site, and we use the two words interchangeably.

Transitivity To give explicit expressions for the dynamics of social individuals (sections "*Evaluating the moments*" and after), we need to evaluate formulas for pairs and triplets of individuals, and we will have to make further assumptions on the $\mathscr{D}$ graph. Let us first recall a few definitions. An isomorphism T of the graph $\mathscr{D}$ is a bijection of the node set that preserves the graph[1;2]. In other words, T is a permutation of the node set such that for all i and j in $\{1,\ldots,N\}$, $d_{ij} = d_{\mathsf{T}(i)\mathsf{T}(j)}$. We will assume that the $\mathscr{D}$ graph is transitive: it means that for all nodes i and j, there is an isomorphism T that transforms i into j:

$$\forall \{i, j\} \in \{1,\ldots,N\}, \quad \exists \mathsf{T}, \quad \mathsf{T}(i) = j. \tag{13}$$

This condition may seem restrictive, and it is indeed if we consider all possible dispersal graphs $\mathscr{D}$ (of which exists an infinity for any population of size N since

we consider weighted graphs). However, this condition actually corresponds to the large majority of structures that are considered in studies on structured populations. For instance, populations subdivided in demes of equal sizes, or living on a lattice, satisfy our assumptions (see examples given in the main text).

Payoffs

All individuals in the population have the same baseline fecundity (W_F^0) and propensity to die (W_S^0), to which are added relative payoffs due to interactions with other individuals. In game theory studies[3], payoffs are usually on fecundity; here we allow for payoffs both on birth rates (fecundity) and death rates (survival) and use the subscripts $_F$ and $_S$ to distinguish between them.

Linearity As classically assumed in this type of study, the payoffs add up in a linear manner: the benefits of an interaction with k individuals is the sum of the k benefits received, so that we only need to specify the outcome of pairwise interactions. These pairwise interactions are usually represented with payoff matrices. We consider general payoff matrices, which we reparametrize in the following way:

$$
\begin{array}{cc}
 & \begin{array}{cc} \text{S} & \text{NS} \end{array} \\
\begin{array}{c} \text{S} \\ \text{NS} \end{array} & \begin{pmatrix} a & b \\ c & d \end{pmatrix}
\end{array}
=
\begin{array}{cc}
 & \begin{array}{cc} \text{S} & \text{NS} \end{array} \\
\begin{array}{c} \text{S} \\ \text{NS} \end{array} & \begin{pmatrix} b-c+d & -c \\ b & 0 \end{pmatrix}.
\end{array}
\tag{14}
$$

The two parametrizations are equivalent, with

$$
\begin{cases}
b & = & c-d \\
c & = & d-b \\
d & = & a-b-c+d.
\end{cases}
\tag{15}
$$

The a, b, c, d parametrization is used to present the results in the main text, but our calculations are easier with the b, c, d notation. The parameter b corresponds to the benefits given by social individuals to any kind of individual, while d corresponds to the additional benefits given to others social individuals only (also called "synergy"[4]). Finally, c corresponds to the costs of being social. Since we consider payoffs on both survival and fecundity, we have two payoff matrices:

$$
A_S = \begin{pmatrix} b_S - c_S + d_S & -c_S \\ b_S & 0 \end{pmatrix} \quad \text{and} \quad A_F = \begin{pmatrix} b_F - c_F + d_F & -c_F \\ b_F & 0 \end{pmatrix}.
\tag{16}
$$

Weak selection   Finally, we will assume weak selection, *i.e.*, that the payoffs are similar for all individuals; they will be scaled by a factor $\omega \ll 1$ (note that in the main text, ω is directly included in the payoffs). The type of weak selection that we consider here corresponds to what has been called[5] w-weak selection (small fitness contributions from the game), as opposed to δ-weak selection (small distance in phenotype space), which is the type of weak selection usually considered in kin selection studies. While the two types of weak selection converge when the payoff matrices are such that there are equal gains from switching *i.e.*, d $= 0$[6], they otherwise usually yield different answers .

Life-cycles

The population is of fixed size N, so at each time step, the number of individuals who die is equal to the number of newly established individuals. Although the equation (37) derived below will actually be valid irrespective of the actual number of individual who die between two time steps, we will focus on a Moran process[7], where only one individual dies—and therefore, only one individual reproduces—between two time steps.

With this constraint of only one individual dying between two time steps, the number of possible life-cycles is quite restricted. Two different life-cycles are usually considered for the dynamics in saturated populations: Death-Birth (DB), and Birth-Death (BD). It has already been shown elsewhere with transitive population structures that these DB and BD processes are in fact symmetrical (DB with payoffs on fecundity is equivalent to BD with payoffs on survival[8;9]), so we focus on DB in this section. The DB life-cycle goes as follows. At time t, the fecundities and death propensities of all individuals in the population are computed according to the interaction graph $\mathscr{E}$; each individual produces (a large number of) propagules, which disperse along the competition graph $\mathscr{D}$: each propagule produced by the individual at node i is sent to node j with a probability d_{ij}. Then, individuals die, each with a probability equal to their death propensity; in a Moran process, only one individual dies between two time steps. Immediately after this, one propagule is chosen among all propagules present in the node where death occurred and becomes a new adult at this node—so potentially, if $d_{ii} \neq 0$, the new individual can be the offspring of the individual who just died. All newborns are adults in the next step and can potentially reproduce (*i.e.*, there is no age structure in the population).

Note that in both cases, the fecundities and death propensities are recomputed at each time step; the process is therefore without memory: the state of the population at time $t + 1$ only depends on its state at time t.

Mutation   We assume that mutations from one type to the other (S → NS and NS → S) are so rare, that they actually only occur once one of the two types (S or NS) is already fixed in the population. Except for these rare mutation events, individuals breed true: the offspring of each reproducing individual have the same state as their parent. This is why there will not be terms accounting for mutation in our equations describing the change of the frequency of social individuals from one time step to the next.

Criteria for the evolution of social behaviour

When does natural selection favour the evolution of social behaviour? The answer to this question depends on the chosen time scale to evaluate the evolutionary success of social individuals. There are at least three different criteria:

1. *The frequency of social individuals is expected to increase from one time step to the next.* This criterion depends on the current state of the population and corresponds to a very short time scale.

2. *The probability of fixation of a single "social" mutation appearing in a non-social population (ρ_S) is greater than the probability of fixation of a neutral mutation ($1/N$).* This criterion corresponds to a long time scale.

3. *The probability of fixation of a single "social" mutation appearing in a non-social population (ρ_S) is greater than the probability of fixation of a "non-social" mutation in a population of social individuals (ρ_{NS}).* This criterion corresponds to a very long time scale. Assuming symmetric and rare mutation ($\mu_{S \to NS} = \mu_{NS \to S}$), this condition is equivalent to saying that the population as a whole spends more time in the S state than in the NS state.

The three criteria are the same only in a limited number of cases, where both d_S and d_F, the synergistic terms, are nought —this corresponds to the particular case of a Prisoner's Dilemma, or equal-gains-from-switching types of payoff matrices[6]—, and, at the first order, under δ-weak selection (small phenotypic distances; see above)[10]. In the general case, the three criteria are different, and they will still be under the type of weak selection that we consider here (w-weak selection, small fitness contributions from the game); we will define success using the third criterion ($\rho_S > \rho_{NS}$), for which explicit solutions can be obtained.

Equation for the change of the frequency of social individuals

In this section, we derive an equation for the change in the frequency of social individuals from one generation to the other. The frequency of social individuals in the population at time $t+1$ for one realization of the process depends on the configuration of the whole population at time t, $X(t) = x$. We will keep its expression as general as possible, making clear when different assumptions are needed.

Change in the frequency of social individuals

A first derivation

To show that our equations are compatible with other frameworks, and in particular with the Price equation, we introduce a new notation. For one realization of the process, we denote by $\xi_i(t)$ the number of individuals in the population at time $t+1$ that are the offspring of the individual who was in site i at time t. Note that "offspring" is meant in a general sense: an individual who survives until the next time step is considered here as its own offspring. Because the population size is constant and equal to N, we have

$$\forall t, \quad \sum_{i=1}^{N} \xi_i(t) = N \quad \text{so that} \quad \overline{\xi(t)} = \sum_{i=1}^{N} \frac{\xi_i(t)}{N} = 1. \tag{17}$$

In the general case, $\xi_i(t)$ can take any value between 0 and N (with the constraint (17)). In a Moran process, only one individual dies and one individual reproduces between t and $t+1$; if we denote by i_d the site where death occurred, and i_r the site in which the individual who reproduced was, we have,

$$\xi_i(t) = \begin{cases} 2 & \text{for } i = i_r \\ 0 & \text{for } i = i_d \\ 1 & \text{for } i \neq i_r \text{ and } i \neq i_d \end{cases} \quad \text{when } i_d \neq i_r, \text{ and} \tag{18}$$

$$\forall i, \xi_i(t) = 1 \qquad \text{when } i_d = i_r \text{ (self replacement).}$$

We denote by $\mathcal{W}_i(t)$ the expected number of offspring at $t+1$ of the individual living at site i at t, given that the population is in state $X(t) = x$:

$$\mathcal{W}_i(t) = \mathbb{E}_t\left[\xi_i(t)|_{X(t)=x}\right]. \tag{19}$$

With this definition, the expected frequency of social individuals at time $t+1$ given that the population is in state $X(t) = x$ at time t is

$$\mathbb{E}\left[\left.\overline{X(t+1)}\right|_{X(t)=x}\right] = \frac{1}{N}\sum_{i=1}^{N}\mathscr{W}_i(t)x_i(t) = \overline{\mathscr{W}x}, \tag{20}$$

so that the expected change, which we will write Δ_x is,

$$\begin{aligned}\Delta_x &= \mathbb{E}\left[\left.\overline{X(t+1)}\right|_{X(t)=x}\right] - \overline{x(t)}\\ &= \frac{1}{N}\sum_{i=1}^{N}\left(\mathscr{W}_i(t)-1\right)x_i.\end{aligned} \tag{21}$$

We define spatial (hence the $_S$ subscript) covariance as follows:

$$\mathrm{cov}_S\left(y,z\right) = \overline{y\,z} - \overline{y}\,\overline{z}. \tag{22}$$

We have seen in equation (17) that $\overline{\xi(t)} = 1$; then we also have $\overline{\mathscr{W}(t)} = 1$, and we can rewrite (21) as

$$\Delta_x = \mathrm{cov}_S\left(\mathscr{W},x\right). \tag{23}$$

Equation (23) is a version of the Price equation[11], and it directly follows from our definition of $\xi_i(t)$. This equation, however, gives us the expected frequency of social individuals at $t+1$, *given that* the population is in state $X(t) = x$ at time t. To be able to compute fixation probabilities, we will need unconditional expected changes in the frequency of social individuals. Using the law of total expectation, the expected frequency of social individuals at time $t+1$ can be written as

$$p(t+1) = \mathbb{E}_{t+1}\left[\overline{X}\right] = \mathbb{E}\left[\mathbb{E}\left[\left.\overline{X(t+1)}\right|_{X(t)}\right]\right] \tag{24}$$

so that the expected change in the frequency of social individuals between t and $t+1$ is

$$\begin{aligned}\Delta p(t) &= p(t+1) - p(t)\\ &= \mathbb{E}_t\left[\overline{\mathscr{W}X} - \overline{X}\,\overline{\mathscr{W}}\right] \quad \text{[because } \overline{\mathscr{W}} = 1\text{]}.\end{aligned} \tag{25}$$

Using the covariance notation defined in (22), we obtain

$$\boxed{\Delta p(t) = \mathbb{E}_t\left[\mathrm{cov}_S\left(\mathscr{W},X\right)\right]}. \tag{26}$$

Again, this result is a consequence of our definition of $\mathscr{W}_i(t)$, the expected number of offspring (*sensu lato*, that is, including the parent if it survives) of the individual living at site i at time t, given that the population is in state $X(t) = x$.

11

We now need to specify $\mathcal{W}_i(t)$. With a Death-Birth (DB) updating, symbol $\mathbb{P}$ denoting a probability, we have

$$
\begin{aligned}
\mathcal{W}_i(t) &= \sum_{j=1}^{N} \mathbb{P}\left(\text{Individual at } j \text{ dies}\right) \times \mathbb{P}\left(\text{An offspring from } i \text{ wins site } j\right) \\
&\quad + \mathbb{P}\left(\text{Individual at } i \text{ survives}\right) \\
&= \sum_{j=1}^{N} \mathbb{P}\left(\text{Individual at } j \text{ dies}\right) \times \mathbb{P}\left(\text{An offspring from } i \text{ wins site } j\right) \\
&\quad + \left(1 - \mathbb{P}\left(\text{Individual at } i \text{ dies}\right)\right).
\end{aligned}
\tag{27}
$$

We now have to specify the probabilities of dying and reproducing in a given site. We define the probability that the individual at site j dies as in a Moran process: it is the relative death propensity of this individual, compared to the whole population.

$$
\forall j, \quad \mathbb{P}\left(\text{Individual at } j \text{ dies}\right) = \frac{W_S^0 \left(1 - \omega w_{Sj}\right)}{\sum_{k=1}^{N} W_S^0 \left(1 - \omega w_{Sk}\right)}
\tag{28}
$$

$$
= \frac{1}{N} \frac{1 - \omega w_{Sj}}{1 - \omega \overline{w_S}},
\tag{29}
$$

where w_{Sk} represents the effects of social interactions on the death propensity of the individual living at site k, given $X(t) = x$. As introduced previously, ω scales the strength of selection. We note that when there is no selection on survival ($\forall k, w_{Sk} = 0$), the probability that the individual at j dies reduces to $1/N$. We call this step the *global step* because the probability of dying depends on an individual's death propensity compared to the death propensities of all other individuals in the population (see the denominator in (28)).

The probability that the offspring of the individual at site i wins site j depends on the effects of sociality on fecundity, but also on the dispersal graph $\mathcal{D}$: the propagules have to be sent to site j, which occurs with probability d_{ij}, and competition is against all propagules arriving at site j:

$$
\forall \{i, j\}, \quad \mathbb{P}\left(\text{An offspring from } i \text{ wins site } j\right) = \frac{W_F^0 \left(1 + \omega w_{Fi}\right) d_{ij}}{\sum_{k=1}^{N} W_F^0 \left(1 - \omega w_{Fk}\right) d_{kj}}
\tag{30}
$$

$$
= \frac{d_{ij}}{\hat{d}_j} \frac{1 + \omega w_{Fi}}{1 + \omega \sum_{k=1}^{N} w_{Fk} d_{kj}/\hat{d}_j},
\tag{31}
$$

where $\hat{d}_j$ measures how many links go to site j on the $\mathcal{D}$ graph:

$$
\hat{d}_j = \sum_{k=1}^{N} d_{kj};
\tag{32}
$$

in the particular case where the links are symmetrical on the $\mathscr{D}$ graph, that is, when $d_{kj} = d_{jk}$, then we have $\hat{d}_j = 1$, because by construction $\sum_{j=1}^{N} d_{kj} = 1$. In the absence of selection on fecundity ($\forall k, w_{Fk} = 0$), the probability that the offspring from i wins the emptied site j is $d_{ij}/\hat{d}_j$. We call this step the *local step* because the probability of giving birth depends on an individual's fecundity compared to the fecundities of the individuals in the neighbourhood, neighbourhood being defined through the dispersal graph $\mathscr{D}$ (see the denominator in (30)).

So replacing the probabilities of dying and giving birth in (27) by their expressions (29) and (31), we get

$$\mathscr{W}_i(t) = \sum_{j=1}^{N} \frac{1}{N} \frac{1 - \omega w_{Sj}}{1 - \omega \overline{w_S}} \times \frac{d_{ij}}{\hat{d}_j} \frac{1 + \omega w_{Fi}}{1 + \omega \sum_{k=1}^{N} w_{Fk} d_{kj}/\hat{d}_j} + 1 - \frac{1}{N} \frac{1 - \omega w_{Si}}{1 - \omega \overline{w_S}}. \quad (33)$$

We now use the fact that we assume weak selection ($\omega \ll 1$):

$$\mathscr{W}_i(t) = 1 + \sum_{j=1}^{N} \frac{1}{N} \left(1 - \omega w_{Sj} + \omega \overline{w_S} + \omega w_{Fi} - \omega \sum_{k=1}^{N} w_{Fk} \frac{d_{kj}}{\hat{d}_j} \right) \times \frac{d_{ij}}{\hat{d}_j}$$
$$- \frac{1}{N} \left(1 - \omega w_{Si} + \omega \overline{w_S} \right) + O(\omega^2). \quad (34)$$

We define $\hat{\hat{d}}_i$ as follows

$$\hat{\hat{d}}_i = \sum_{j=1}^{N} \frac{d_{ij}}{\hat{d}_j}, \quad (35)$$

so that we can rewrite $\mathscr{W}_i(t)$:

$$\mathscr{W}_i(t) = 1 + \sum_{j=1}^{N} \frac{1}{N} \left(1 - \omega w_{Sj} + \omega \overline{w_S} + \omega w_{Fi} - \omega \sum_{k=1}^{N} w_{Fk} \frac{d_{kj}}{\hat{d}_j} \right) \times \frac{d_{ij}}{\hat{d}_j} \quad (36)$$
$$- \sum_{j=1}^{N} \frac{1}{N} \left(1 - \omega w_{Si} + \omega \overline{w_S} \right) \times \frac{d_{ij}}{\hat{d}_j \hat{\hat{d}}_i} + O(\omega^2)$$

$$\mathscr{W}_i(t) = 1 + \sum_{j=1}^{N} \frac{1}{N} \left(\left(1 - \frac{1}{\hat{\hat{d}}_i} \right) \left(1 + \omega \overline{w_S} \right) + \omega \left(\frac{w_{Si}}{\hat{\hat{d}}_i} - w_{Sj} \right) + \omega \left(w_{Fi} - \sum_{k=1}^{N} w_{Fk} \frac{d_{kj}}{\hat{d}_j} \right) \right) \times \frac{d_{ij}}{\hat{d}_j} + O(\omega^2). \quad (37)$$

Equation (37) is valid when selection is weak, and for any dispersal graph $\mathscr{D}$ satisfying (7). From now on, we are going to focus on graphs that are such that for all sites i and j, $d_{ij} = d_{ji}$, which implies that $\hat{d}_j = 1 = \hat{\hat{d}}_j$. In this case, equation (37) simplifies into

$$\boxed{\mathscr{W}_i(t) = 1 + \omega \sum_{j=1}^{N} \frac{1}{N} \left((w_{Si} - w_{Sj}) + \left(w_{Fi} - \sum_{k=1}^{N} w_{Fk} d_{kj} \right) \right) \times d_{ij} + O(\omega^2)}, \quad (38)$$

and the expected change in the global frequency of social individuals, given the state $X(t) = x(t)$, is, using (25),

$$\Delta p(t) = \mathbb{E}_t \left[\frac{1}{N} \sum_{i=1}^{N} \left(\frac{\omega}{N} \sum_{j=1}^{N} \left(w_{Si} - w_{Sj} + w_{Fi} - \sum_{k=1}^{N} w_{Fk} d_{kj} \right) d_{ij} + O(\omega^2) \right) X_i \right].$$
(39)

Thanks to the assumption of weak selection and the Taylor-expansion made in equation (34), we have transformed products and ratios into sums and differences, and we can therefore study separately the effects of social behaviour on fecundity and survival. We define $\Delta_S(t)$ and $\Delta_F(t)$ as follows:

$$\Delta_S(t) = \mathbb{E}_t \left[\frac{\omega}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(w_{Si} - w_{Sj} \right) d_{ij} X_i \right]$$
(40)

$$\Delta_F(t) = \mathbb{E}_t \left[\frac{\omega}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(w_{Fi} - \sum_{k=1}^{N} w_{Fk} d_{kj} \right) d_{ij} X_i \right],$$
(41)

so that we have

$$\boxed{\Delta p(t) = \Delta_S(t) + \Delta_F(t) + O(\omega^2)}.$$
(42)

BD life-cycle   With a Birth-Death life-cycle, we simply have to switch the $_F$ and $_S$ labels.

Another derivation

The derivation we used in section "*A first derivation*" allowed us to draw a link with the Price equation. This derivation is however not classical under a Moran process. In this section, we briefly outline another derivation. For the sake of brevity, we focus as previously on dispersal graphs $\mathscr{D}$ that are symmetric, *i.e.*, such that $\forall \{i, j\}$, $d_{ij} = d_{ji}$, an assumption made from equation (38) above, and used throughout the remainder of the article.

As mentioned previously, under a Moran process, there is only one death and one birth (as previously, we focus on a DB updating, BD being obtained by switching the $_F$ and $_S$ subscripts). At time t, the number of social individuals in the population is $\sum_{k=1}^{N} x_k$, given that the population is in state $X(t) = x$. We denote by T_t^+ the probability that there is one more social individual at time $t + 1$, and T_t^- the probability that there is one less social individual at $t + 1$, given $X(t) = x$. The number of social individuals increases if a non-social dies between t and $t + 1$ and is replaced by a social individual:

$$T_t^+ = \sum_{i=1}^{N} (1 - x_i) \times \frac{W_S^0 (1 - \omega w_{Si})}{\sum_{k=1}^{N} W_0^S (1 - \omega w_{Sk})} \times \frac{\sum_{j=1}^{N} x_j W_F^0 (1 + \omega w_{Fj}) d_{ji}}{\sum_{k=1}^{N} W_F^0 (1 + \omega w_{Fk}) d_{ki}}.$$
(43)

14

With our weak selection assumption, and using the fact that since $d_{ij} = d_{ji}$, we have $\sum_{j=1}^{N} d_{ji} = 1$, this expression becomes

$$T_t^+ = \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{N} \left(1 + \omega \left[-w_{Si} + \sum_{k=1}^{N} \frac{w_{Sk}}{N} + w_{Fj} - \sum_{k=1}^{N} w_{Fk} d_{ki} \right] \right) (1 - x_i) \, x_j \, d_{ji} + O(\omega^2)$$

$$T_t^+ = \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{N} \left(1 + \omega \left[-w_{Sj} + \sum_{k=1}^{N} \frac{w_{Sk}}{N} + w_{Fi} - \sum_{k=1}^{N} w_{Fk} d_{kj} \right] \right) (1 - x_j) \, x_i \, d_{ji} + O(\omega^2)$$

$$\text{(44)}$$

[switching the i and j indices, and using the fact that $d_{ij} = d_{ji}$].

Similarly, given that the population is in state $X(t) = x$, the number of social individuals decreases with probability

$$T_t^- = \sum_{i=1}^{N} x_i(t) \times \frac{W_S^0 (1 - \omega \, w_{Si})}{\sum_{k=1}^{N} W_0^S (1 - \omega \, w_{Sk})} \times \frac{\sum_{j=1}^{N} (1 - x_j) \, W_F^0 (1 + \omega \, w_{Fj}) d_{ji}}{\sum_{k=1}^{N} W_F^0 (1 + \omega \, w_{Fk}) d_{ki}}$$

$$T_t^- = \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{N} \left(1 + \omega \left[-w_{Si} + \sum_{k=1}^{N} \frac{w_{Sk}}{N} + w_{Fj} - \sum_{k=1}^{N} w_{Fk} d_{ki} \right] \right) x_i \, (1 - x_j) \, d_{ji} + O(\omega^2),$$

$$\text{(45)}$$

and we can check that these two expressions are compatible our previous derivation, by noting that the expected change in the frequency of social individuals at time $t + 1$ given $X(t) = x$ can be written as

$$\Delta_x = +\frac{1}{N} T_t^+ - \frac{1}{N} T_t^-. \tag{46}$$

With the expressions of T_t^+ and T_t^- given in equations (44) and (45), we get

$$\Delta_x = \frac{1}{N} \sum_{i=1}^{N} \left[\frac{\omega}{N} \sum_{j=1}^{N} \left(w_{Si} - w_{Sj} + w_{Fi} - w_{Fj} + \sum_{k=1}^{N} w_{Fk} d_{ki} - \sum_{k=1}^{N} w_{Fk} d_{kj} \right) d_{ji} + O(\omega^2) \right] x_i$$

$$= \frac{1}{N} \sum_{i=1}^{N} \left[\frac{\omega}{N} \sum_{j=1}^{N} \left(w_{Si} - w_{Sj} + w_{Fi} - \sum_{k=1}^{N} w_{Fk} d_{kj} \right) d_{ji} + O(\omega^2) \right] x_i, \tag{47}$$

because

$$\sum_{i=1}^{N} \sum_{j=1}^{N} \left(w_{Fj} - \sum_{k=1}^{N} w_{Fk} d_{ki} \right) x_i \, d_{ji} = \sum_{i=1}^{N} \sum_{j=1}^{N} w_{Fj} x_i \, d_{ji} - \sum_{i=1}^{N} \sum_{k=1}^{N} w_{Fk} d_{ki} x_i \quad \text{[using (11)].}$$

$$= 0;$$

and taking the expectation, (47) is the same as (39).

Remark

Because the population average of $\mathscr{W}_i(t)$ is 1 since the population remains of size N, an individual at site i produces more offspring than the average when $\mathscr{W}_i(t) > 1$, which occurs when, at the first order in ω, and using a general notation (The subscripts ([1],[2]) being (S,F) for DB, and (F, S) for BD):

$$\sum_{j=1}^{N} \left(w_{[1]i} - w_{[1]j} + w_{[2]i} - \sum_{k=1}^{N} w_{[2]k}\, d_{kj} \right) \times d_{ij} > 0. \tag{48}$$

This equation simplifies greatly when selection only affects one step of the process only—and previous studies only considered the two cases below.

1. When selection affects the first step only, equation (48) simplifies into

$$w_{[1]i} - \sum_{j=1}^{N} w_{[1]j}\, d_{ij} > 0, \tag{49}$$

where we used the fact that, by definition, $\sum_j d_{ij} = 1$. The dispersal graph $\mathscr{D}$ defining neighbourhood, equation (49) means that the individual at site i is expected to have more offspring than the average individual in the population ($\mathscr{W}_i(t) > 1$) if its fitness $w_{[1]i}$ is greater than the average fitness of its neighbours, $\sum_{j=1}^{N} w_{[1]j}\, d_{ij}$. There is therefore direct competition between the individual at site i and its neighbours, the strength of this competition being scaled by $d_{ij}(= d_{ji})$. This is linked to what has been called the 1-circle[2]. This 1-circle is illustrated in Supplementary Fig. 1(a) for the specific case of a lattice structured population with $k = 4$ neighbours for each individual.

2. When selection affects the second step only, equation (48) simplifies into

$$w_{[2]i} - \sum_{j=1}^{N} \sum_{k=1}^{N} w_{[2]k}\, d_{kj}\, d_{ij} > 0. \tag{50}$$

Equation (50) means that an individual at site i is expected to have more offspring than the average individual if its fitness $w_{[2]i}$ is greater than the average fitness of all its neighbours' neighbours, $\sum_{j=1}^{N} \sum_{k=1}^{N} w_{[2]k}\, d_{kj}\, d_{ij}$. This is linked to what has been called the 2-circle[2], which is illustrated in Supplementary Fig. 1(b).

Hence, the inclusive fitness of an individual living at focal site i ($\mathscr{W}_i(t)$) is differently affected at each step of the process. The first step of the process involves

16

competition with the focal's neighbours, while the second step involves competition with the neighbours' neighbours; the two effects sum up (thanks to our assumption of weak selection).

Specifying the fitnesses

We next specify the w terms. Since the fecundity and survival expressions will be equivalent, we leave the S and F subscripts aside for the moment. We will need to recall that the w terms are conditional to the state $X(t)$ of the population.

General expressions

For an individual at site j and at time t, the state of the population being given by $X(t) = x$, we have

$$w_j = \sum_{l=1}^{N} e_{lj} \left(\mathsf{b}\, x_l - \mathsf{c}\, x_j + \mathsf{d}\, x_l\, x_j \right)$$

$$\boxed{w_j = \sum_{l=1}^{N} e_{lj} \left(\mathsf{b}\, x_l + \mathsf{d}\, x_l\, x_j \right) - \mathsf{c}\, x_j} \quad \text{[because of (8)].} \tag{51}$$

Summing over all the other sites l, we count the benefits and costs received by the individual at j; the interaction between l and j is scaled by e_{lj} (interaction graph $\mathscr{E}$). Unconditional benefits (b) are received from site l if this site is occupied by a social individual (x_l), while additional benefits (synergistic terms d) are received if both individuals at l and j are social individuals ($x_l x_j$). Finally, costs (c) are paid only if the individual at j is a social individual (x_j).

Survival terms

We now plug the expression for w_S derived above into the expression of $\Delta_S(t)$ given in equation (40). We first note that

$$\sum_{i=1}^{N} \sum_{j=1}^{N} w_{Sj}\, d_{ij}\, x_i = \sum_{i=1}^{N} \sum_{j=1}^{N} w_{Si}\, d_{ji}\, x_j \quad \text{[exchanging the } i \text{ and } j \text{ indices]}$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} w_{Si}\, d_{ij}\, x_j, \tag{52}$$

17

where we used the fact that $d_{ij} = d_{ji}$. We can hence rewrite $\Delta_S(t)$ as follows

$$\Delta_S(t) = \mathbb{E}_t \left[\frac{\omega}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(w_{Si} X_i(t) - w_{Si} X_j(t) \right) d_{ij} \right]$$

$$= \frac{\omega}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{l=1}^{N} e_{li} \Bigg(\mathsf{b}_S \left(\mathbb{E}_t \left[X_l X_i \right] - \mathbb{E}_t \left[X_l X_j \right] \right) \tag{53}$$

$$+ \mathsf{d}_S \left(\mathbb{E}_t \left[X_l X_i \right] - \mathbb{E}_t \left[X_l X_i X_j \right] \right)$$

$$- \mathsf{c}_S \left(\mathbb{E}_t \left[X_i \right] - \mathbb{E}_t \left[X_i X_j \right] \right) \Bigg) d_{ij},$$

where we used the fact that for all sites i $X_i(t)^2 = X_i(t)$, because X_i takes values 0 and 1. We can rewrite $\Delta_S(t)$ in a more compact way:

$$\Delta_S(t) = \beta_S \mathsf{b}_S + \delta_S \mathsf{d}_S - \gamma_S \mathsf{c}_S, \tag{54}$$

where the β_S, γ_S and δ_S are defined below.

Survival benefits   The first term, β_S, is as follows:

$$\beta_S = \frac{\omega}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{l=1}^{N} \left(\mathbb{E}_t \left[X_l X_i \right] - \mathbb{E}_t \left[X_l X_j \right] \right) e_{li} d_{ji}. \tag{55}$$

So far, the focus was on the benefits received by i, but we can now rewrite this equation from another perspective, looking at the benefits given by l:

$$\beta_S = \frac{\omega}{N^2} \sum_{l=1}^{N} \sum_{i=1}^{N} e_{li} \sum_{j=1}^{N} \left(\mathbb{E}_t \left[X_l X_i \right] - \mathbb{E}_t \left[X_l X_j \right] \right) d_{ji}, \tag{56}$$

and this formulation is similar an inclusive fitness formulation[1], which we did not assume, but instead derived. Note that this derivation required one assumption only on the dispersal $\mathcal{D}$ graph, namely the fact that dispersal is symmetrical ($d_{ij} = d_{ji}$), as well as the assumption of weak selection. We can also write equation (56) as follows:

$$\beta_S = \frac{\omega}{N^2} \sum_{l=1}^{N} \sum_{i=1}^{N} e_{li} \sum_{j=1}^{N} \left(\mathbb{E}_t \left[X_l X_i (1 - X_j) \right] - \mathbb{E}_t \left[X_l (1 - X_i) X_j \right] \right) d_{ji}. \tag{57}$$

This expression is used in Supplementary Fig. 2, a figure that provides a schematic description of the β, γ and δ terms. It will also be useful in section "*Simplifying the δ_S and δ_F terms*", to simplify δ_S.

We can also rewrite the expression for β_S using matrices. We denote by $\mathbf{P}$ the matrix such that $P_{ij} = \mathbb{E}_t \left[X_i X_j \right]$; in particular, we have $\mathbf{P}^T = \mathbf{P}$. Then, (56) becomes

$$\boxed{\beta_S = \frac{\omega}{N^2} \, \text{Tr} \left(\mathbf{E}^T . (\mathbf{P} - \mathbf{P}.\mathbf{D}) \right)}, \tag{58}$$

where Tr is the trace (sum of diagonal elements).

Survival costs   For the costs on survival, we have

$$\gamma_S = \frac{\omega}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\mathbb{E}_t \left[X_i \right] - \mathbb{E}_t \left[X_i X_j \right] \right) d_{ji}. \tag{59}$$

Using the fact that $\sum_l e_{li} = 1$ (equation (8)), we can rewrite the expression for γ_S as follows:

$$\gamma_S = \frac{\omega}{N^2} \sum_{l=1}^{N} \sum_{i=1}^{N} e_{li} \sum_{j=1}^{N} \left(\mathbb{E}_t \left[X_l X_i (1 - X_j) \right] + \mathbb{E}_t \left[(1 - X_l) X_i (1 - X_j) \right] \right) d_{ji}. \tag{60}$$

This expanded notation is used in Supplementary Fig. 2, and in section "*Simplifying the δ_S and δ_F terms*". Using matrix notation, we get

$$\boxed{\gamma_S = \frac{\omega}{N^2} \, \text{Tr} \left(\mathbf{P} - \mathbf{P}.\mathbf{D} \right)}. \tag{61}$$

Survival synergistic effects   For the synergy terms, we have

$$\delta_S = \frac{\omega}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{l=1}^{N} e_{li} \left(\mathbb{E}_t \left[X_l X_i \right] - \mathbb{E}_t \left[X_l X_i X_j \right] \right) d_{ji}; \tag{62}$$

after reorganizing the sums, we obtain the following expression:

$$\delta_S = \frac{\omega}{N^2} \sum_{l=1}^{N} \sum_{i=1}^{N} e_{li} \sum_{j=1}^{N} \left(\mathbb{E}_t \left[X_l X_i \right] - \mathbb{E}_t \left[X_l X_i X_j \right] \right) d_{ji}. \tag{63}$$

This expression can be rewritten as follows (see Supplementary Fig. 2):

$$\delta_S = \frac{\omega}{N^2} \sum_{l=1}^{N} \sum_{i=1}^{N} e_{li} \sum_{j=1}^{N} \left(\mathbb{E}_t \left[X_l X_i (1 - X_j) \right] \right) d_{ji}. \tag{64}$$

If we define Π the tensor of order 3 (or three-dimensional array), such that $\Pi_{ijk} = \mathbb{E}_t \left[X_i X_j X_k \right]$, using a matrix notation, we get

$$\boxed{\delta_S = \frac{\omega}{N^2} \, \text{Tr} \left(\mathbf{E}^T . (\mathbf{P} - \Pi.\mathbf{D}) \right)}. \tag{65}$$

Fecundity terms

We now plug the expression for w_S derived above into the expression of $\Delta_S(t)$ given in equation (40). We note that, just by switching summation indices,

$$\sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} w_{Fk}\, d_{kj} d_{ij}\, x_i = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} w_{Fi}\, d_{ij} d_{kj}\, x_k \tag{66}$$

We can hence rewrite $\Delta_F(t)$ as follows:

$$\Delta_F(t) = \mathbb{E}_t\left[\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{N} \omega \left(w_{Fi} X_i - \sum_{k=1}^{N} w_{Fi} X_k\, d_{kj} \right) \times d_{ij} \right]$$

$$= \mathbb{E}_t\left[\frac{\omega}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \left(w_{Fi} X_i - \sum_{k=1}^{N} w_{Fi} X_k \right) d_{kj}\, d_{ij} \right] \text{ [since } d_{kj} = d_{jk},\, \sum_{k} d_{kj} = 1\text{]}$$

$$= \frac{\omega}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{l=1}^{N} e_{li} \bigg(\mathsf{b}_F\, (\mathbb{E}_t\,[X_l X_i] - \mathbb{E}_t\,[X_l X_k]) \tag{67}$$

$$+ \mathsf{d}_F\, (\mathbb{E}_t\,[X_l X_i] - \mathbb{E}_t\,[X_l X_i X_k])$$

$$- \mathsf{c}_F\, (\mathbb{E}_t\,[X_i] - \mathbb{E}_t\,[X_i X_k]) \bigg) d_{kj}\, d_{ij},$$

As previously with the survival terms, we can rewrite $\Delta_F(t)$ in a more compact way:

$$\Delta_F(t) = \beta_F\, \mathsf{b}_F + \delta_F\, \mathsf{d}_F - \gamma_F\, \mathsf{c}_F, \tag{68}$$

where the β_F, γ_F and δ_F are defined below.

Fecundity benefits The first term, β_F, is as follows:

$$\beta_F = \frac{\omega}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{l=1}^{N} e_{li}\, (\mathbb{E}_t\,[X_l X_i] - \mathbb{E}_t\,[X_l X_k])\, d_{kj}\, d_{ij}. \tag{69}$$

Again, we can change the focus and rewrite this equation from another perspective, looking at the benefits given by l:

$$\beta_F = \frac{\omega}{N^2} \sum_{l=1}^{N} \sum_{i=1}^{N} e_{li} \sum_{j=1}^{N} \sum_{k=1}^{N} (\mathbb{E}_t\,[X_l X_i] - \mathbb{E}_t\,[X_l X_k])\, d_{kj}\, d_{ij}, \tag{70}$$

an expression that can also be rewritten as follows:

$$\beta_F = \frac{\omega}{N^2} \sum_{l=1}^{N} \sum_{i=1}^{N} e_{li} \sum_{j=1}^{N} \sum_{k=1}^{N} (\mathbb{E}_t\,[X_l X_i (1 - X_k)] - \mathbb{E}_t\,[X_l (1 - X_i) X_k])\, d_{kj}\, d_{ij}. \tag{71}$$

This expression is depicted in Supplementary Fig. 3, and will be used in section "*Simplifying the δ_S and δ_F terms*" to simplify δ_F. Using matrices, we obtain the following compact expression:

$$\beta_F = \frac{\omega}{N^2} \operatorname{Tr}\left(\mathbf{E}^T.(\mathbf{P} - \mathbf{P}.\mathbf{D}.\mathbf{D})\right).$$ (72)

Fecundity costs   For the costs on fecundity, we have

$$\gamma_F = \frac{\omega}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} (\mathbb{E}_t[X_i] - \mathbb{E}_t[X_i X_k]) \, d_{kj} \, d_{ij}.$$ (73)

As for the survival costs, we can use the fact that $\sum_l e_{li} = 1$, and rewrite the expression as follows (see Supplementary Fig. 3):

$$\gamma_F = \frac{\omega}{N^2} \sum_{l=1}^{N} \sum_{i=1}^{N} e_{li} \sum_{j=1}^{N} \sum_{k=1}^{N} (\mathbb{E}_t[X_l X_i(1 - X_k)] + \mathbb{E}_t[(1 - X_l) X_i(1 - X_k)]) \, d_{kj} \, d_{ij}.$$ (74)

We can also rewrite γ_F with matrices:

$$\gamma_F = \frac{\omega}{N^2} \operatorname{Tr}(\mathbf{P} - \mathbf{P}.\mathbf{D}.\mathbf{D}).$$ (75)

Fecundity synergistic effects   For the synergy terms, we have

$$\delta_F = \frac{\omega}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{l=1}^{N} e_{li} (\mathbb{E}_t[X_l X_i] - \mathbb{E}_t[X_l X_i X_k]) \, d_{kj} \, d_{ij};$$ (76)

after reorganizing the terms, we obtain the following expression:

$$\delta_F = \frac{\omega}{N^2} \sum_{l=1}^{N} \sum_{i=1}^{N} e_{li} \sum_{j=1}^{N} \sum_{k=1}^{N} (\mathbb{E}_t[X_l X_i] - \mathbb{E}_t[X_l X_i X_k]) \, d_{kj} \, d_{ij},$$ (77)

that we can also write as follows (see Supplementary Fig. 3):

$$\delta_F = \frac{\omega}{N^2} \sum_{l=1}^{N} \sum_{i=1}^{N} e_{li} \sum_{j=1}^{N} \sum_{k=1}^{N} (\mathbb{E}_t[X_l X_i(1 - X_k)]) \, d_{kj} \, d_{ij}.$$ (78)

Finally, with matrices, we can rewrite δ_F as

$$\delta_F = \frac{\omega}{N^2} \operatorname{Tr}\left(\mathbf{E}^T.(\mathbf{P} - \Pi.\mathbf{D}.\mathbf{D})\right).$$ (79)

Together

If we go back to equation (42), we can rewrite it as

$$\Delta p(t) = \beta_S \, b_S + \delta_S \, d_S - \gamma_S \, c_S + \beta_F \, b_F + \delta_F \, d_F - \gamma_F \, c_F + O(\omega^2) \qquad (80)$$

for a Death-Birth process, or, more generally,

$$\boxed{\Delta p(t) = \beta_{[1]} \, b_{[1]} + \delta_{[1]} \, d_{[1]} - \gamma_{[1]} \, c_{[1]} + \beta_{[2]} \, b_{[2]} + \delta_{[2]} \, d_{[2]} - \gamma_{[2]} \, c_{[2]} + O(\omega^2)} \qquad (81)$$

To derive equation (80), we have had to assume weak selection ($\omega \ll 1$), that all nodes produce the same number of offspring when the population is monomorphic ($\sum_j d_{ij} = 1$), and the fact that the dispersal graph $\mathscr{D}$ is symmetrical ($\forall i, j, d_{ij} = d_{ji}$, or, in matrix notation, $\mathbf{D} = \mathbf{D}^T$), but this equation is valid even if $\mathscr{D}$ is not transitive, and no assumption is necessary (and no will be in the following), besides (8) ($\sum_l e_{li} = 1$), for the interaction graph $\mathscr{E}$.

All the expressions that we have derived so far depend not only on the expected state of each site i, $\mathbb{E}_t[X_i]$, but also on higher spatial moments such as the expected state of pairs ($\mathbb{E}_t[X_i X_j]$) and triplets of sites ($\mathbb{E}_t[X_i X_j X_k]$, for the δ_S and δ_F terms only). We would like to express the change in the frequency of social individuals as a function of the frequency of social individuals at time t only. To do so, we have to evaluate spatial moments, which we will do in the following section. This will require a separation of time scales, which is made possible by the assumption of weak selection.

Link with relatedness

In our derivation, for the β and γ terms, we need to evaluate $\mathbb{E}_t[X_i X_j]$ pairs; as observed before,

$$\mathbb{E}_t[X_i X_j] = \mathbb{P}(X_i(t) = 1, X_j(t) = 1). \qquad (82)$$

Inclusive fitness derivations often use instead probabilities of identity in state or probabilities of identity by descent, without conditioning on the type of the individuals:

$$G_{ij} = \mathbb{P}(X_i(t) = X_j(t)); \qquad (83)$$

relatedness is often defined as [12]

$$R_{ij} = \frac{G_{ij} - \overline{\overline{G}}}{1 - \overline{\overline{G}}}, \qquad (84)$$

where

$$\overline{\overline{G}} = \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{G_{ij}}{N^2}. \qquad (85)$$

While our derivation does not require the use of these relatedness coefficients, a link can be made by noticing that

$$G_{ij}(t) = 1 - \mathbb{E}_t[X_i] - \mathbb{E}_t[X_j] + 2\mathbb{E}_t[X_i X_j]. \tag{86}$$

To leading order ω^0, on a transitive dispersal graph, $\mathbb{E}_t[X_i] = \mathbb{E}_t[X_j] = \mathbb{E}_t[\overline{X}] = p(t)$. Noting that since X_i is an indicator variable, $\overline{X} = \overline{X^2}$, using the same definition of covariance as in (22), and writing $\text{var}_S(x) = \text{cov}_S(x, x)$, we obtain after simplifying

$$R_{ij}(t) = 1 + \frac{\mathbb{E}_t[X_i X_j] - p(t)}{\mathbb{E}_t[\text{var}_S(X)]} \tag{87}$$

or, equivalently,

$$\mathbb{E}_t[X_i X_j] = (R_{ij}(t) - 1)\,\mathbb{E}_t[\text{var}_S(X)] + p(t). \tag{88}$$

Evaluating the moments

We need to find expressions for the dynamics of pairs $(X_i X_j(t))$ and triplets $(X_i X_j X_k(t))$, in order to give explicit expressions of the β_S, β_F, γ_S, γ_F, δ_S and δ_F terms in equation (80). We derive these equations neglecting terms in ω and higher (recall that ω scales the strength of selection). We start with the dynamics of pairs.

Dynamics of pairs

Let us consider two sites, i and $j \neq i$. We first derive an expression for the probability that sites i and j are both occupied by social individuals at time $t + 1$, $\mathbb{P}(X_i(t+1) = 1, X_j(t+1) = 1) = \mathbb{E}[X_i(t+1)X_j(t+1)]$ given that the population is in state $X(t) = x$ at time t. As mentioned in section "*Equation for the change of the frequency of social individuals*", we focus on symmetric graphs, *i.e.*, graphs such that $\forall i, j, d_{ij} = d_{ji}$. Since we consider a Moran process, at most one individual changes between t and $t + 1$. To the leading order ω^0, we have:

$$\forall i, j \neq i, \quad \mathbb{E}[X_i(t+1)X_j(t+1)|_{X(t)=x}] = x_i \frac{1}{N} \sum_{k=1}^{N} x_k d_{kj} + \frac{1}{N} \sum_{k=1}^{N} x_k d_{ki} x_j$$
$$+ \left(1 - \frac{2}{N}\right) x_i x_j + O(\omega). \tag{89}$$

The first term in equation (89) means that the individual who died during t and $t + 1$ was the individual at i (whichever its status was) $(1/N)$ and it is replaced

by a social individual coming from site k ($\sum_k x_k d_{ki}$), and while site j, whose occupancy is unchanged, is occupied by a social individual (x_j); the second term is similar, except for the fact that j is the site where death occurs. Finally, the third and last term corresponds to the situation where death occurred in a site which is neither i nor j ($1 - 2/N$), and these two sites are occupied by social individuals ($x_i x_j$).

The expected change between t and $t + 1$, given the state $X(t) = x$ of the population at t is, for all i and $j \neq i$,

$$\Delta_{xy} = \mathbb{E}\left[X_i(t+1)X_j(t+1)|_{X(t)=x}\right] - x_i x_j$$
$$= \frac{1}{N}\left[\sum_{k=1}^{N} x_k x_i d_{kj} + \sum_{k=1}^{N} x_k x_j d_{ki} - 2 x_i x_j\right] + O(\omega), \tag{90}$$

and if we take the expectation over all population configurations at time t, we get

$$\forall i, j \neq i, \quad \Delta P_{ij}(t) = \mathbb{E}\left[\mathbb{E}\left[X_i(t+1)X_j(t+1)|_{X(t)}\right]\right] - \mathbb{E}_t\left[X_i X_j\right] \tag{91}$$
$$= \frac{1}{N}\left[\sum_{k=1}^{N} \mathbb{E}_t\left[X_k X_i\right] d_{kj} + \sum_{k=1}^{N} \mathbb{E}_t\left[X_k X_j\right] d_{ki} - 2\mathbb{E}_t\left[X_i X_j\right]\right] + O(\omega),$$

Hence, the dynamics of pairs are of order $\omega^0 = 1$, and so will be the dynamics of triplets. We have seen previously (equation (40)–(41)) that the dynamics of singlets are of order ω, and with weak selection $\omega \ll 1$. In order words, at the zeroth order in ω, we have $\Delta p(t) = O(\omega)$. This means that, while the dynamics of singlets are extremely slow, the dynamics of pairs and triplets occur on a much faster time scale; we can therefore decompose the two time scales.

Separation of time scales for the pairs

Quasi equilibrium $\neq$ equilibrium The dynamics of pairs are much faster than the dynamics of singlets, so we can separate time scales and use a quasi-equilibrium approximation (a technique also called relaxation projection[13]). Note that a quasi-equilibrium is not an equilibrium: if, for all pairs of sites i and $j \neq i$, the expectation of the change is 0 (as will be the case when an equilibrium is reached), then nothing happens any more in the population, and the global dynamics do not change either: we will have $\Delta p(t) = 0$ too[2]. We are concerned here with population dynamics, not equilibria. Our quasi-equilibrium approximation will be such that once the quasi equilibrium is reached, the expected

---

[2] Population geneticists may note here that a similar reasoning is made with quasi-linkage approximations (QLE).

24

change for pairs of individuals only depends on lower moments (singlets); we denote by $f(t)$ this expected change at the quasi equilibrium.

Mean field relaxation projection   We will assume that once the quasi equilibrium is reached, the dynamics of pairs and triplets become equivalent to their dynamics in an unstructured population with the same frequency of social individuals $\overline{X}$; in other words, we will do a mean field relaxation projection.

Getting back to the equation for the pairs, (91), once the quasi equilibrium is reached, we have $\Delta P_{ij} = f(t)$, that is,

$$\frac{1}{N}\left[\sum_{k=1}^{N}\mathbb{E}_t\left[X_k X_i\right]d_{kj} + \sum_{k=1}^{N}\mathbb{E}_t\left[X_k X_j\right]d_{ki} - 2\mathbb{E}_t\left[X_i X_j\right]\right] = f(t), \qquad (92)$$

which is equivalently written as

$$\boxed{\forall i, j \neq i, \quad \mathbb{E}_t\left[X_i X_j\right] = \frac{1}{2}\left(\sum_{k=1}^{N}\mathbb{E}_t\left[X_j X_k\right]d_{ki} + \sum_{k=1}^{N}\mathbb{E}_t\left[X_i X_k\right]d_{kj}\right) - \phi(t)}, \qquad (93)$$

where

$$\phi(t) = \frac{N f(t)}{2}. \qquad (94)$$

When the $\mathscr{D}$ graph is transitive...   calculations become easier. As described in section "*Assumptions on the $\mathscr{D}$ graph*", we say that $\mathscr{D}$ is transitive if for all sites i and j, there is an isomorphism T (*i.e.*, a transformation that leaves the graph structure unchanged) such that $\mathsf{T}(i) = j$ (recall that classical structures like deme-structured populations, or lattices, fall into this category). Denoting by $\mathbf{T}$ the matrix corresponding to the transformation T, and $\mathbf{D}$ being the matrix representing the dispersal graph $\mathscr{D}$ (defined in (9)), this means that

$$\mathbf{T}^T.\mathbf{D}.\mathbf{T} = \mathbf{D}, \qquad (95)$$

and since $\mathbf{T}$ corresponds to a permutation, $\mathbf{T}^T = \mathbf{T}^{-1}$. We can follow a previous method[2] (their Appendix A) and rewrite (91) with matrices, noting that for all sites, $P_{ii}(t+1) = P_{ii}(t) + O(\omega)$:

$$\mathbf{P}(t+1) - \mathbf{P}(t) = \frac{1}{N}\left[\mathbf{D}.\mathbf{P}(t) + \mathbf{P}(t).\mathbf{D} - 2\mathbf{P}(t)\right] + \mathbf{L}(t) + O(\omega), \qquad (96)$$

where $\mathbf{L}(t)$ is a diagonal matrix, whose i^{th} diagonal element is

$$L_i(t) = -\frac{2}{N}\left[(\mathbf{D}.\mathbf{P}(t))_{ii} - P_{ii}\right]. \qquad (97)$$

25

Initially, there is one social individual who can be anywhere in the population, so that $\mathbf{P}(0) = \frac{1}{N}\mathbf{I}_N$ ($\mathbf{I}_N$ being the identity matrix). The transitivity of $\mathbf{D}$ implies the transitivity of $\mathbf{P}$ (at all times), which, in turn, implies the transitivity of $\mathbf{L}(t)$. Since $\mathbf{L}(t)$ is a diagonal matrix, this means that all its diagonal elements are equal: $\mathbf{L}(t) = c(t)\mathbf{I}_N$. Hence, we have $\mathbf{D}.\mathbf{P}(0) = \mathbf{P}(0).\mathbf{D}$ and also for all t, $\mathbf{L}(t).\mathbf{D} = \mathbf{D}.\mathbf{L}(t)$. Then, by induction, equation (96) implies that at all times, the dispersal matrix $\mathbf{D}$ and the matrix $\mathbf{P}$ commute[2]:

$$\mathbf{D}.\mathbf{P} = \mathbf{P}.\mathbf{D}, \tag{98}$$

where we dropped the time dependency of $\mathbf{P}$ for notational simplicity. This means that the two sums in the right hand side of equation (93) are equal[1;9]. So with a transitive $\mathscr{D}$ graph, equation (93) reads now:

$$\forall i, j \neq i, \quad \mathbb{E}_t\left[X_i X_j\right] = \sum_{k=1}^{N} \mathbb{E}_t\left[X_j X_k\right] d_{ki} - \phi(t)$$
$$= \sum_{k=1}^{N} \mathbb{E}_t\left[X_i X_k\right] d_{kj} - \phi(t) \tag{99}$$

We can sum equation (99) over $j \neq i$:

$$\sum_{j\neq i} \mathbb{E}_t\left[X_i X_j\right] = \sum_{j\neq i}\sum_{k=1}^{N} \mathbb{E}_t\left[X_i X_k\right] d_{kj} - \sum_{j\neq i}\phi(t)$$
$$= \sum_{j=1}^{N} \mathbb{E}_t\left[X_i X_j\right] - \mathbb{E}_t\left[X_i\right] = \sum_{k=1}^{N} \mathbb{E}_t\left[X_i X_k\right](1 - d_{ki}) - (N-1)\phi(t),$$

where we used the fact that $X_i(t)^2 = X_i(t)$ (because $X_i(t)$ is an indicator variable), and that $\sum_{j=1}^{N} d_{kj} = 1$ (by definition). Hence, on a transitive graph, we have:

$$\boxed{\forall i, \quad \mathbb{E}_t\left[X_i\right] = \sum_{k=1}^{N} \mathbb{E}_t\left[X_i X_k\right] d_{ki} + (N-1)\phi(t)}. \tag{100}$$

We denote by $\mathbf{1}_{N,N}$ the N by N matrix that contains only ones; and $\mathbf{I}_N$ the identity matrix; we can combine the results (99) and (100) as follows:

$$\boxed{\mathbf{P} = \mathbf{P}.\mathbf{D} - \phi(t)\mathbf{1}_{N,N} + N\phi(t)\mathbf{I}_N}. \tag{101}$$

Right-multiplying both sides by $\mathbf{D}$, and simplifying, we also obtain

$$\boxed{\mathbf{P} = \mathbf{P}.\mathbf{D}.\mathbf{D} - 2\phi(t)\mathbf{1}_{N,N} + N\phi(t)\,(\mathbf{I}_N + \mathbf{D})}. \tag{102}$$

26

Evaluating $\phi(t)$

Our mean-field relaxation approximation means that at the quasi equilibrium, the population behaves like an unstructured one. In an unstructured population, for all i and j, $d_{ij} = 1/N$. We note that, in an unstructured population of fixed size, we have

$$
\begin{aligned}
\forall i, j \neq i \quad \mathbb{E}_t \left[X_i X_j \right] &= \mathbb{E} \left[\mathbb{E} \left[X_i(t) X_j(t) |_{X(t)=x(t)} \right] \right] \quad \text{[law of total expectation]} \\
&= \mathbb{E} \left[\mathbb{P} \left(X_i(t) = 1, X_j(t) = 1 |_{X(t)=x(t)} \right) \right] \\
&= \mathbb{E}_t \left[\overline{X} \frac{N\overline{X} - 1}{N - 1} \right].
\end{aligned} \tag{103}
$$

We plug this expression in equation (100), and after rearranging we get

$$
\phi(t) = \frac{N}{N-1} \mathbb{E}_t \left[\overline{X} - \overline{X}^2 \right]; \tag{104}
$$

we note that since X_i takes values 0 and 1,

$$
\overline{X^2} = \sum_{i=1}^{N} \frac{X_i^2}{N} = \sum_{i=1}^{N} \frac{X_i}{N} = \overline{X}, \tag{105}
$$

so that, using the same definition of covariance as in (22), we obtain

$$
\boxed{\phi(t) = \frac{N}{N-1} \mathbb{E}_t \left[\mathrm{cov}_S \left(X, X \right) \right] = \frac{N}{N-1} \mathbb{E}_t \left[\mathrm{var}_S \left(X \right) \right]}. \tag{106}
$$

Simplifying the δ_S and δ_F terms

The δ factors in (80) appear in front of the synergistic terms d, and contain triplets of individuals (see (62) and (76)), while the expressions of the β (factor in front on the unconditional benefits b) and γ (in front of the costs c) terms only involve at most pairs of individuals. We show below how our relaxation approximation allows us to write the δ terms as functions of the β and γ terms and the expected global frequency of social individuals, $p(t)$ [14;15].

Survival terms

The expression of δ_S (see equation (62)) appears within the expressions of β_S (55) and γ_S (59), as can be seen below:

$$\beta_S - \delta_S = -\frac{\omega}{N^2} \sum_{l=1}^{N} \sum_{i=1}^{N} e_{li} \sum_{j=1}^{N} \mathbb{E}_t \left[(1 - X_i) X_j X_l \right] d_{ji}$$

$$= -\mathbb{E}_t \left[F_S(X) \right] \text{[hereby defining } F_S\text{]}; \tag{107}$$

$$\gamma_S - \delta_S = \frac{\omega}{N^2} \sum_{l=1}^{N} \sum_{i=1}^{N} e_{li} \sum_{j=1}^{N} \mathbb{E}_t \left[X_i(1 - X_j)(1 - X_l) \right] d_{ji}$$

$$= \mathbb{E}_t \left[F_S(1 - X) \right]. \tag{108}$$

We therefore have

$$\delta_S = \frac{\beta_S + \gamma_S}{2} + \frac{\mathbb{E}_t \left[F_S(X) \right] - \mathbb{E}_t \left[F_S(1 - X) \right]}{2}. \tag{109}$$

We will now use again the mean-field relaxation approximation that we used in section "*Evaluating* $\phi(t)$", which will allow us to express triplets as a function of lower moments (pairs and singlets). We note that in a well-mixed population, $F_S(X) - F_S(1 - X) = 0$ when $\overline{X} = 1/2$. Secondly, we note that $F_S(X) = 0$ when $\overline{X} = 0$ or $\overline{X} = 1$. We assume that taken together, these two considerations mean that there exist a constant α_S such that we can write [15] (recalling that $\mathbb{E}_t \left[\text{var}_S(X) \right] = \mathbb{E}_t \left[\overline{X}(1 - \overline{X}) \right]$)

$$\boxed{\delta_S = \frac{\beta_S + \gamma_S}{2} + \left(p(t) - \frac{1}{2} \right) \omega \, \alpha_S \times \mathbb{E}_t \left[\text{var}_S(X) \right]}. \tag{110}$$

We note that in a well-mixed population, we have

$$\delta_S = \frac{\beta_S + \gamma_S}{2} + \left(p(t) - \frac{1}{2} \right) \left(\gamma_S - \beta_S \right). \tag{111}$$

Fecundity terms

We are going to use a similar argument for the fecundity terms. The expression of δ_F (see equation (76)) also appears within the expressions of β_F (69) and γ_F

(73), as can be seen below:

$$\beta_F - \delta_F = -\frac{\omega}{N^2} \sum_{l=1}^{N} \sum_{i=1}^{N} e_{li} \sum_{j=1}^{N} \sum_{k=1}^{N} \mathbb{E}_t \left[(1 - X_i) X_k X_l \right] d_{kj} d_{ij}$$

$$= -\mathbb{E}_t \left[F_F(X) \right] \text{ [hereby defining } F_F\text{]};} \tag{112}$$

$$\gamma_F - \delta_F = \frac{\omega}{N^2} \sum_{l=1}^{N} \sum_{i=1}^{N} e_{li} \sum_{j=1}^{N} \sum_{k=1}^{N} \mathbb{E}_t \left[X_i (1 - X_k) (1 - X_l) \right] d_{kj} d_{ij}$$

$$= \mathbb{E}_t \left[F_F(1 - X) \right]. \tag{113}$$

We therefore have, once the quasi-equilibrium is reached and provided the mean-field relaxation approximation holds,

$$\delta_F = \frac{\beta_F + \gamma_F}{2} + \frac{\mathbb{E}_t \left[F_F(X) - F_F(1 - X) \right]}{2}. \tag{114}$$

Same as for the survival terms, we note that in an unstructured population, $F_F(X) = F_F(1 - X)$ when $\overline{X} = 1/2$, and that that $F_S(X) = 0$ when $\overline{X} = 0$ or $\overline{X} = 1$, so that we obtain a similar formula as (110):

$$\boxed{\delta_F = \frac{\beta_F + \gamma_F}{2} + \omega \left(p(t) - \frac{1}{2} \right) \alpha_F \times \mathbb{E}_t \left[\mathrm{var}_S (X) \right]}, \tag{115}$$

and we note that in a well-mixed population,

$$\delta_F = \frac{\beta_F + \gamma_F}{2} + \left(p(t) - \frac{1}{2} \right) \left(\gamma_F - \beta_F \right). \tag{116}$$

Dynamics on a transitive graph

We have seen in equation (42) that thanks to the assumption of weak selection, we can look at the payoffs on survival and fecundity independently. We will start with the effects on survival, and will then turn to effects on fecundity.

Survival terms (first step)

Costs on survival

Using the expression of γ_S derived in equation (61), and the relaxation solution (101):

$$\gamma_S = \frac{\omega}{N^2} \mathrm{Tr} \left(\mathbf{P} - \mathbf{P.D} \right)$$

$$= \frac{\omega}{N^2} \mathrm{Tr} \left(-\phi(t) \mathbf{1}_{N,N} + N \phi(t) \mathbf{I}_N \right), \tag{117}$$

29

and we obtain

$$\gamma_S = \frac{\omega}{N}(N-1)\,\phi(t). \tag{118}$$

Benefits on survival

For the benefits, we have, using (58), then the relaxation solution (101):

$$\begin{aligned}
\beta_S &= \frac{\omega}{N^2}\,\mathrm{Tr}\left(\mathbf{E}^T.(\mathbf{P}-\mathbf{P}.\mathbf{D})\right) \\
&= \frac{\omega}{N^2}\,\mathrm{Tr}\left(\mathbf{E}^T.\left(-\phi(t)\,\mathbf{1}_{N,N}+N\phi(t)\,\mathbf{I}_N\right)\right).
\end{aligned} \tag{119}$$

Because of (8), $\mathrm{Tr}\left(\mathbf{E}^T.\mathbf{1}_{N,N}\right) = N$, and we denote by e_{self} the average interaction with oneself:

$$e_{\mathrm{self}} = \frac{1}{N}\sum_{l=1}^{N} e_{ll}, \tag{120}$$

so that β_S simplifies into

$$\beta_S = \frac{\omega}{N}\,(Ne_{\mathrm{self}}-1)\,\phi(t). \tag{121}$$

Conclusion for the effects on the first step of the process

Collecting the different expressions, we obtain

$$\boxed{\begin{aligned}
\gamma_S &= \frac{\omega}{N}(N-1)\,\phi(t) \\
\beta_S &= \frac{\omega}{N}\,(Ne_{\mathrm{self}}-1)\,\phi(t) \\
\delta_S &= \frac{\beta_S+\gamma_S}{2}+\omega\left(p(t)-\frac{1}{2}\right)\alpha_S\times\mathbb{E}_t\,[\mathrm{var}_S\,(X)]
\end{aligned}} \tag{122}$$

System (122) reveals that, when the social interactions influence the first step of the process, population structure does not influence the dynamics of the frequency of social individuals in the population.

Fecundity terms (second step)

Costs on fecundity

We use the expression of γ_F derived in equation (75), then the relaxation solution (102):

$$\begin{aligned}
\gamma_F &= \frac{\omega}{N^2}\,\mathrm{Tr}\,(\mathbf{P}-\mathbf{P}.\mathbf{D}.\mathbf{D}) \\
&= \frac{\omega}{N^2}\,\mathrm{Tr}\left(-2\,\phi(t)\,\mathbf{1}_{N,N}+N\phi(t)\,(\mathbf{I}_N+\mathbf{D})\right).
\end{aligned} \tag{123}$$

Denoting by $d_{\text{self}} = d_{ii}$ the probability that offspring stay at their parent's site, which is the same for all i because the $\mathcal{D}$ graph is transitive, γ_F simplifies into

$$\gamma_F = \frac{\omega}{N} \left(N + N d_{\text{self}} - 2\right) \phi(t). \tag{124}$$

Benefits on fecundity

We now turn to the benefits on fecundity, whose expression is given in equation (72), and we again use the relaxation solution (102):

$$\begin{aligned}
\beta_F &= \frac{\omega}{N^2} \, \text{Tr}\left(\mathbf{E}^T.\left(\mathbf{P} - \mathbf{P}.\mathbf{D}.\mathbf{D}\right)\right) \\
&= \frac{\omega}{N^2} \, \text{Tr}\left(\mathbf{E}^T.\left(-2\,\phi(t)\,\mathbf{1}_{N,N} + N\phi(t)\,\left(\mathbf{I}_N + \mathbf{D}\right)\right)\right)
\end{aligned} \tag{125}$$

As previously, we note that $\text{Tr}\left(\mathbf{E}^T.\mathbf{1}_{N,N}\right) = N$, and that $\text{Tr}\left(\mathbf{E}^T.\mathbf{I}_N\right) = N\,e_{\text{self}}$ (see equation (120)). We now define

$$\overline{\overline{ed}} = \frac{1}{N} \sum_{l=1}^{N} \sum_{k=1}^{N} e_{lk} d_{kl} = \frac{1}{N} \, \text{Tr}\left(\mathbf{E}^T.\mathbf{D}\right), \tag{126}$$

and we obtain

$$\beta_F = \frac{\omega}{N} \left(N\,e_{\text{self}} + N\,\overline{\overline{ed}} - 2\right) \phi(t). \tag{127}$$

Conclusion for the effects on the first step of the process

Collecting the different expressions for the terms involved in the second step of the process, we get

$$\boxed{\begin{aligned}
\gamma_F &= \frac{\omega}{N} \left(N + N d_{\text{self}} - 2\right) \phi(t) \\
\beta_F &= \frac{\omega}{N} \left(N\,e_{\text{self}} + N\,\overline{\overline{ed}} - 2\right) \phi(t) \\
\delta_F &= \frac{\beta_F + \gamma_F}{2} + \omega \left(p(t) - \frac{1}{2}\right) \alpha_F \times \mathbb{E}_t \left[\text{var}_S\left(X\right)\right]
\end{aligned}} \tag{128}$$

Comparing these expressions to the expressions for the first step of the process, (122), we see that spatial structure does influence the evolution of social behaviour, when the effects of the behaviour affect the second step of the process.

Conclusion

Going back to equation (80), using the expressions derived in systems (122) and (128) and our evaluation of $\phi(t)$ derived in (106), we obtain the following change in the expected frequency of social individuals in the population:

$$
\begin{aligned}
\Delta p(t) = \Bigg(&+ \mathsf{b}_S \frac{N e_{\text{self}} - 1}{N-1} \\
&+ \mathsf{d}_S \left(\frac{N + N e_{\text{self}} - 2}{2(N-1)} - \frac{\alpha_S}{2} + \alpha_S\, p(t) \right) \\
&- \mathsf{c}_S \\
&+ \mathsf{b}_F \frac{N e_{\text{self}} + N \overline{\overline{ed}} - 2}{N-1} \\
&+ \mathsf{d}_F \left(\frac{N + N d_{\text{self}} + N e_{\text{self}} + N \overline{\overline{ed}} - 4}{2(N-1)} - \frac{\alpha_F}{2} + \alpha_F\, p(t) \right) \\
&- \mathsf{c}_F \frac{N + N d_{\text{self}} - 2}{N-1} \Bigg) \omega \times \mathbb{E}_t\left[\text{var}_S(X) \right]
\end{aligned} \tag{129}
$$

and we define $\sigma_{\mathscr{D}\mathscr{E}}$ as

$$
\sigma_{\mathscr{D}\mathscr{E}} = \Bigg(+ \mathsf{b}_S \frac{N e_{\text{self}} - 1}{N-1} + \mathsf{d}_S \left(\frac{N + N e_{\text{self}} - 2}{2(N-1)} - \frac{\alpha_S}{2} \right) - \mathsf{c}_S
$$
$$
+ \mathsf{b}_F \frac{N e_{\text{self}} + N \overline{\overline{ed}} - 2}{N-1} + \mathsf{d}_F \left(\frac{N + N d_{\text{self}} + N e_{\text{self}} + N \overline{\overline{ed}} - 4}{2(N-1)} - \frac{\alpha_F}{2} \right) - \mathsf{c}_F \frac{N + N d_{\text{self}} - 2}{N-1} \Bigg) \omega \tag{130}
$$

and $\tau_{\mathscr{D}\mathscr{E}}$ as

$$
\tau_{\mathscr{D}\mathscr{E}} = \left(\mathsf{d}_S\, \alpha_S + \mathsf{d}_F\, \alpha_F \right) \omega, \tag{131}
$$

and

$$
s_{\mathscr{D}\mathscr{E}}(p(t)) = \sigma_{\mathscr{D}\mathscr{E}} + p(t)\, \tau_{\mathscr{D}\mathscr{E}}, \tag{132}
$$

so that $s_{\mathscr{D}\mathscr{E}}(p(t))$ is a selection coefficient and we can rewrite (129) as

$$
\Delta p(t) = s_{\mathscr{D}\mathscr{E}}(p(t)) \times \mathbb{E}_t\left[\text{var}_S(X) \right]. \tag{133}
$$

A note for δ-weak selection

In this study, we focus on w-weak selection (weak effects of the interactions), but this section highlights the difference between w- and δ- weak selection[5]. Let us briefly go back to the payoff matrices, and denote by x' the expressed phenotype of a social individual (S), and x the expressed phenotype of a non-social (NS). We denote by $B(x_1, x_2)$ benefits given by a x_2 individual to a x_1 individual, and $C(x)$ the cost paid by a x individual. Under δ-weak selection, we have $x' = x + \delta$, where $\delta \ll 1$. A payoff matrix can therefore be written as follows:

$$\mathbf{A}(x', x) = \begin{pmatrix} B(x', x') - C(x') & B(x', x) - C(x') \\ B(x, x') - C(x) & B(x, x) - C(x) \end{pmatrix} = \mathbf{A_0}(x) + \delta \mathbf{A}'(x) + O(\delta^2), \quad (134)$$

where the matrix

$$\mathbf{A_0}(x) = (B(x, x) - C(x)) \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \quad (135)$$

corresponds to baseline payoffs that are the same for all individuals, while the difference among types is contained in

$$\mathbf{A}'(x) = \begin{pmatrix} \frac{\partial B}{\partial x_1}(x, x) + \frac{\partial B}{\partial x_2}(x, x) - C'(x) & \frac{\partial B}{\partial x_1}(x, x) - C'(x) \\ \frac{\partial B}{\partial x_2}(x, x) & 0 \end{pmatrix}. \quad (136)$$

We note that $\mathbf{A}'(x)$ corresponds to a payoff matrix without synergistic terms d: if we define

$$\begin{cases} \mathsf{b} & = & \frac{\partial B}{\partial x_2}(x, x) \\ \mathsf{c} & = & C'(x) - \frac{\partial B}{\partial x_1}(x, x) \end{cases}, \quad (137)$$

then we can rewrite $\mathbf{A}'(x)$ as

$$\mathbf{A}'(x) = \begin{pmatrix} \mathsf{b} - \mathsf{c} & -\mathsf{c} \\ \mathsf{b} & 0 \end{pmatrix}. \quad (138)$$

Hence, with δ-weak selection, the sign of the selection coefficient becomes frequency independent ($\tau_{\mathscr{D}\mathscr{E}}$, defined in (131), is zero), and the three criteria enumerated in section "*Criteria for the evolution of social behaviour*" are equivalent, as previously demonstrated[10]. But, as we have seen, the situation is different with w-weak selection.

Fixation probability

Fixation probability of a "social" mutation

Our choice of a mean-field relaxation means that we assume that once the relaxation manifold is reached, the dynamics of the structured population become

equivalent to the dynamics in a well-mixed population, with a frequency dependent selection coefficient $s_{\mathscr{DE}}(p(t))$ given by equation (132). Note that in the main text, we use a a, b, c, d parametrization of the payoff matrices, instead of b, c, d (see (15) for the equivalence between the two parametrizations), and also that the ω term is merged in the payoff terms, so that it does not appear.

In a Moran process, we can get explicit expressions of the fixation probability of a single mutant[16;17]. Writing $p(t) = y(t)/N = y/N$ (dropping the time dependency for the sake of notational simplicity), at the first order in ω, the fixation probability of initially one mutant social individual ($y(0) = 1$) with a selection coefficient $s_{\mathscr{DE}}(p(t))$ reads:

$$
\begin{aligned}
\rho_{\mathrm{S}} &\approx \frac{1}{1 + \sum_{K=1}^{N-1} \prod_{x=1}^{K} \left[1 - \left(\sigma_{\mathscr{DE}} + \tau_{\mathscr{DE}} \frac{y}{N}\right)\right]} \\
&\approx \frac{1}{1 + \sum_{K=1}^{N-1} \left(1 - \sum_{x=1}^{K} \left(\sigma_{\mathscr{DE}} + \tau_{\mathscr{DE}} \frac{y}{N}\right)\right)} \\
&\approx \frac{1}{1 + \sum_{K=1}^{N-1} \left(1 - K\sigma_{\mathscr{DE}} - \tau_{\mathscr{DE}} \frac{1}{N} \frac{K(K+1)}{2}\right)} \\
&\approx \frac{1}{1 + \left(N - 1 - \sigma_{\mathscr{DE}} \frac{(N-1)N}{2} - \tau_{\mathscr{DE}} \frac{1}{N} \frac{N(N-1)(N+1)}{6}\right)} \\
&\approx \frac{1}{N} \frac{1}{1 - \sigma_{\mathscr{DE}} \frac{N-1}{2} - \tau_{\mathscr{DE}} \frac{1}{N} \frac{(N-1)(N+1)}{6}},
\end{aligned}
\tag{139}
$$

so that we get:

$$
\boxed{\rho_{\mathrm{S}} \approx \frac{1}{N} + \frac{N-1}{2N}\left(\sigma_{\mathscr{DE}} + \tau_{\mathscr{DE}} \frac{N+1}{3N}\right)}.
\tag{140}
$$

We note that when N gets large, the term between the brackets in (140) becomes equivalent to $s_{\mathscr{DE}}(1/3)$. Social behaviours get fixed more often than a neutral mutation if $\rho > 1/N$. If we denote by $p^* = -\frac{\sigma_{\mathscr{DE}}}{\tau_{\mathscr{DE}}}$ the value of p at which the selection coefficient $s_{\mathscr{DE}}$ is zero, we note that, when N is large,

$$
\begin{aligned}
\rho > 1/N &\iff s_{\mathscr{DE}}(1/3) > 0 \\
&\iff \begin{cases} p^* < \frac{1}{3} & \text{if } \tau_{\mathscr{DE}} > 0 \\ p^* > \frac{1}{3} & \text{if } \tau_{\mathscr{DE}} < 0. \end{cases}
\end{aligned}
\tag{141}
$$

This is the generalization of the "one-third" law[18]. We also note that this result is valid for any kind of weak, linearly frequency dependent selection (*i.e.*, where the selection coefficient can be written as $s(p) = \sigma + \tau\, p$, with σ and τ small compared to 1), and is hence not restricted to social interactions.

Fixation probability of a "non-social" mutation

We write $q(t) = 1 - p(t)$ the expected frequency of non-social individuals in the population. We know that

$$\Delta q(t) = \Delta(1 - p(t)) = -\Delta p(t), \tag{142}$$

so that, following (133) we can write

$$\Delta q(t) = \left(-(\sigma_{\mathcal{DE}} + \tau_{\mathcal{DE}}) + q(t)\,\tau_{\mathcal{DE}}\right) \mathbb{E}_t\left[\mathrm{var}_S(X)\right]. \tag{143}$$

Following (140), the probability of fixation of a single non-social mutant in a social population is

$$\boxed{\rho_{\mathrm{NS}} \approx \frac{1}{N} + \frac{N-1}{2N}\left(-(\sigma_{\mathcal{DE}} + \tau_{\mathcal{DE}}) + \tau_{\mathcal{DE}}\frac{N+1}{3N}\right).} \tag{144}$$

Conclusion

Our evolutionary criterion described in section "*Criteria for the evolution of social behaviour*" states that social individuals are favoured by natural selection when

$$\rho_S > \rho_{\mathrm{NS}}, \tag{145}$$

which, given the formulas presented in equations (140) and (144), means that we need

$$2\,\sigma_{\mathcal{DE}} + \tau_{\mathcal{DE}} > 0. \tag{146}$$

Given the definitions of $\sigma_{\mathcal{DE}}$ and $\tau_{\mathcal{DE}}$ presented in (130) and (131), this is equivalent to

$$\left(+\, \mathsf{b}_S\frac{N e_{\mathrm{self}} - 1}{N-1} + \mathsf{d}_S\left(\frac{N + N e_{\mathrm{self}} - 2}{2(N-1)}\right) - \mathsf{c}_S\right.$$
$$\left. +\, \mathsf{b}_F\frac{N e_{\mathrm{self}} + N\overline{\overline{ed}} - 2}{N-1} + \mathsf{d}_F\left(\frac{N + N d_{\mathrm{self}} + N e_{\mathrm{self}} + N\overline{\overline{ed}} - 4}{2(N-1)}\right) - \mathsf{c}_F\frac{N + N d_{\mathrm{self}} - 2}{N-1}\right) > 0$$
$$\tag{147}$$

Going back to the a, b, c, d parametrization of the payoffs (see equation (15)), after simplifying, and using the general notation for the first and second step (for DB, the subscript [1] is S and [2] is F, and conversely for BD), condition (147)

becomes

$$
\left\{
\begin{array}{l}
\left(1 + e_{\text{self}} - \frac{2}{N}\right)\left(a_{[1]} + a_{[2]} - d_{[1]} - d_{[2]}\right) + (1 - e_{\text{self}})\left(b_{[1]} + b_{[2]} - c_{[1]} - c_{[2]}\right) \\[2mm]
+ \left(d_{\text{self}} + \overline{\overline{ed}} - \frac{2}{N}\right)\left(a_{[2]} - d_{[2]}\right) + \left(d_{\text{self}} - \overline{\overline{ed}}\right)\left(b_{[2]} - c_{[2]}\right)
\end{array}
\right\} > 0
$$

(148)

Illustrations

Parameters

In the following, we investigate the relative effect of allocating benefits and costs of social interactions to fecundity or survival, or, most generally, the first or second step of the process. We consider three canonical games (Prisoner's Dilemma, Snowdrift, Stag Hunt); we give expressions with general population structures, and illustrate the results with specific ones.

Allocation of Benefits and Costs

An individual receiving a benefit B allocates a fraction λ_B of it to the life-history trait that corresponds to the second step of the Moran process (*i.e.*, fecundity for a Death-Birth updating, or survival for a Birth-Death updating):

$$
B_{[1]} = (1 - \lambda_B) B, \quad \text{and} \quad B_{[2]} = \lambda_B B.
$$

(149)

Similarly, a fraction λ_C of the cost paid by social individuals affects the second step of the process:

$$
C_{[1]} = (1 - \lambda_C) C, \quad \text{and} \quad C_{[2]} = \lambda_C C.
$$

(150)

Population structures

We recapitulate here the values of the key parameters of the dispersal graph $\mathscr{D}$ and interaction graph $\mathscr{E}$ for specific population structures.

Well-mixed population Every individual interacts with every individual in the population, including itself, and its offspring can go anywhere, including staying at the parent's site:

$$
d_{\text{self}} = \frac{1}{N}, \quad e_{\text{self}} = \frac{1}{N}, \quad \overline{\overline{ed}} = \frac{1}{N}.
$$

(151)

36

Full graph Every individual interacts with every other individual in the population (excluding itself), and its offspring can go anywhere, except for staying at the parent's site:

$$d_{\text{self}} = 0, \quad e_{\text{self}} = 0, \quad \overline{\overline{ed}} = \frac{1}{N-1}. \tag{152}$$

Regular graph of degree k Each individual is connected to exactly k other individuals, interacts with them socially and its offspring disperses to one of these k sites.

$$d_{\text{self}} = 0, \quad e_{\text{self}} = 0, \quad \overline{\overline{ed}} = \frac{1}{k}. \tag{153}$$

Deme structured population 1 The population is subdivided into demes of n individuals exactly; each individual interacts socially with all the members of its deme (including itself); its offspring stays in the deme with probability $1 - m$ (landing at any of the n sites with equal chance) or disperses to a different deme with probability m.

$$d_{\text{self}} = \frac{1-m}{n}, \quad e_{\text{self}} = \frac{1}{n}, \quad \overline{\overline{ed}} = \frac{1-m}{n}. \tag{154}$$

Deme structured population 2 The dispersal structure is the same as in the first deme structured population, but now individuals interact with the other members of their deme, excluding themselves:

$$d_{\text{self}} = \frac{1-m}{n}, \quad e_{\text{self}} = 0, \quad \overline{\overline{ed}} = \frac{1-m}{n}. \tag{155}$$

Prisoner's Dilemma

The payoffs at step i read

$$M_{[i]}^{PD} = \begin{pmatrix} B_{[i]} - C_{[i]} & -C_{[i]} \\ B_{[i]} & 0 \end{pmatrix}. \tag{156}$$

Defining the cost to benefit ratio $r = \frac{C}{B}$ ($r > 1$), social individuals are favoured whenever condition (148) is satisfied—note that this is the particular case where the three conditions detailed in section "*Criteria for the evolution of social behaviour*" are equivalent—:

$$\left(e_{\text{self}} - \frac{1}{N} \right) - r \left(1 - \frac{1}{N} \right) + \lambda_B \left(\overline{\overline{ed}} - \frac{1}{N} \right) + r \lambda_C \left(\frac{1}{N} - d_{\text{self}} \right) > 0. \tag{157}$$

In most cases (unless the $\mathcal{D}$ and $\mathcal{E}$ graphs do not overlap), $\overline{\overline{ed}} > 1/N$, so that strategies allocating the benefits to the second step (higher λ_B) are favoured. The effect of the cost depends on whether an individual can replace itself (d_{self}). If it cannot, as is the case on a lattice, then allocating the cost to the second step is favourable (higher λ_C), but this effect vanishes when the population becomes large, and allocating the cost on fecundity or survival does not matter any more. If an individual can replace itself, *i.e.*, if some of its offspring stay on the same site ($d_{\text{self}} > 1/N$) as is the case in a population structured in demes, then it is better to allocate the costs to the first step of the process.

Snowdrift

The payoffs at step i read

$$M^{SD}_{[i]} = \begin{pmatrix} B_{[i]} - C_{[i]}/2 & B_{[i]} - C_{[i]} \\ B_{[i]} & 0 \end{pmatrix}. \tag{158}$$

Defining the cost to benefit ratio $r' = \frac{C}{2B-C}$ ($0 < r' < 1$), condition (148) is satisfied when

$$\left(1 + e_{\text{self}} - \frac{2}{N}\right) - 2r'(1 - e_{\text{self}}) + \lambda_B(1+r')\left(\overline{\overline{ed}} + d_{\text{self}} - \frac{2}{N}\right) + r'\lambda_C\left(\frac{2}{N} + \overline{\overline{ed}} - 3d_{\text{self}}\right) > 0. \tag{159}$$

In general, allocating the benefits to the second step (higher λ_B) is more conducive to the evolution of social behaviour, except in the particular case of a full graph ($d_{\text{self}} = 0$, $\overline{\overline{ed}} = \frac{1}{N-1}$). Here again, the effect of the allocation of the cost depends on whether individuals can replace themselves. When they do not, allocation of the costs to the second step is favoured, while when they do substantially, as for instance in a deme-structured population, it is better to allocate the costs to the first step.

Stag Hunt

The payoffs at step i read

$$M^{SH}_{[i]} = \begin{pmatrix} 2B_{[i]} - C_{[i]} & 0 \\ B_{[i]} & B_{[i]} \end{pmatrix}. \tag{160}$$

Using the cost to benefit ratio $r'' = \frac{B+C}{2B}$ ($\frac{1}{2} < r'' < 1$), condition (148) is satisfied when

$$\left(1 + 3e_{\text{self}} - \frac{4}{N}\right) - 2r''\left(1 + e_{\text{self}} - \frac{2}{N}\right) + \lambda_B \times 2\left(\overline{\overline{ed}} - \frac{1}{N}\right) + \lambda_C(2r''-1)\left(\frac{2}{N} - d_{\text{self}} - \overline{\overline{ed}}\right) > 0. \tag{161}$$

The term in front of λ_B has the same sign as in the Prisoner's Dilemma game: benefits on the second step are favoured. The effect of the cost is such that (except for a full graph), allocation to the first step (lower λ_C) is more conducive to the evolution of social behaviour.

Simple Spite

The payoffs at step i read

$$M^{SS}_{[i]} = \begin{pmatrix} -B_{[i]} - C_{[i]} & -C_{[i]} \\ -B_{[i]} & 0 \end{pmatrix}. \tag{162}$$

We can use the same cost to benefit ratio $r = \frac{C}{B}$ ($r > 1$) as for the Prisoner's Dilemma. Social individuals are favoured whenever condition (148) is satisfied (and here too, this is the particular case where the three conditions detailed in section "*Criteria for the evolution of social behaviour*" are equivalent):

$$-\left(e_{\text{self}} - \frac{1}{N}\right) - r\left(1 - \frac{1}{N}\right) - \lambda_B\left(\overline{\overline{ed}} - \frac{1}{N}\right) + r\,\lambda_C\left(\frac{1}{N} - d_{\text{self}}\right) > 0. \tag{163}$$

Since the cost terms are equivalent to the ones in the Prisoner's Dilemma game, the effects of λ_C are the same. The effect of λ_B differs however, because of the minus sign in front of the benefit terms in the payoff matrix $M^{SS}_{[i]}$. As mentioned previously, in most structured populations, $\overline{\overline{ed}} > 1/N$, so that strategies allocating the benefits to the first step (lower λ_B) are favoured.

References

[1] Taylor, P. D., Day, T. & Wild, G. Evolution of cooperation in a finite homogeneous graph. *Nature* **447**, 469–472 (2007).

[2] Grafen, A. & Archetti, M. Natural selection of altruism in inelastic viscous homogeneous populations. *Journal of Theoretical Biology* **252**, 694 – 710 (2008).

[3] Ohtsuki, H., Hauert, C., Lieberman, E. & Nowak, M. A. A simple rule for the evolution of cooperation on graphs and social networks. *Nature* **441**, 502–505 (2006).

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

[5] Wild, G. & Traulsen, A. The different limits of weak selection and the evolutionary dynamics of finite populations. *Journal of Theoretical Biology* **247**, 382 – 390 (2007).

[6] Nowak, M. & Sigmund, K. The evolution of stochastic strategies in the prisoner's dilemma. *Acta Applicandae Mathematicae* **20**, 247–265 (1990). 10.1007/BF00049570.

[7] Moran, P. A. P. Random processes in genetics. *Mathematical Proceedings of the Cambridge Philosophical Society* **54**, 60–71 (1958).

[8] Nakamaru, M. & Iwasa, Y. The coevolution of altruism and punishment: Role of the selfish punisher. *Journal of Theoretical Biology* **240**, 475 – 488 (2006).

[9] Taylor, P. Birth–death symmetry in the evolution of a social trait. *Journal of Evolutionary Biology* **23**, 2569–2578 (2010).

[10] Taylor, P. D., Day, T. & Wild, G. From inclusive fitness to fixation probability in homogeneous structured populations. *Journal of Theoretical Biology* **249**, 101 – 110 (2007).

[11] Price, G. R. Selection and covariance. *Nature* **227**, 520–521 (1970).

[12] Taylor, P., Lillicrap, T. & Cownden, D. Inclusive fitness analysis on mathematical groups. *Evolution* **65**, 849–859 (2011).

[13] Law, R. & Dieckmann, U. *The Geometry of Ecological Interactions: Simplifying Spatial Complexity*, chap. Relaxation Projections and the Method of Moments, 252–270 (Cambridge University Press, 2000).

[14] Tarnita, C. E., Ohtsuki, H., Antal, T., Fu, F. & Nowak, M. A. Strategy selection in structured populations. *Journal of Theoretical Biology* **259**, 570 – 581 (2009).

[15] Taylor, P. & Maciejewski, W. An inclusive fitness analysis of synergistic interactions in structured populations. *Proceedings of the Royal Society B: Biological Sciences* (2012).

[16] Durrett, R. *Probability Models for DNA Sequence Evolution*, vol. XII of *Probability and Its Applications*, chap. 1. Basic Models (2008), 2nd edn.

[17] Huang, W. & Traulsen, A. Fixation probabilities of random mutants under frequency dependent selection. *Journal of Theoretical Biology* **263**, 262 – 268 (2010).

[18] Nowak, M., Sasaki, A., Taylor, C. & Fudenberg, D. Emergence of cooperation and evolutionary stability in finite populations. *Nature* **428**, 646–650 (2004).