A thermodynamic limit constrains complexity and primitive social function

Supporting Information

Philip J Gerrish[1,2]*, Claudia P Ferreira[3]

[1] School of Biological Sciences, Georgia Institute of Technology, 310 Ferst Dr, Atlanta, GA 30332, USA.
[2] Theoretical Biology & Biophysics, Los Alamos National Laboratory, Los Alamos, NM 87545, USA.
[3] Departamento de Bioestatística, Instituto de Biociências, Universidade Estadual Paulista 18618-000 Botucatu, São Paulo, Brazil.

*To whom correspondence should be addressed. E-mail: pgerrish@gatechs.edu

Our general aim is to determine the critical mutation rate above which cooperators (or "social" genotypes) are displaced by cheaters (or "asocial" genotypes) in a structured population. Our approach is to allow only forward mutation from social to asocial genotypes and determine the lowest possible mutation rate for which the social-free state is dynamically stable. The rationale behind this approach is that even when reversion to social genotypes is re-introduced into the model, these genotypes will be maintained at very low frequencies under conditions above the critical mutation rate. To determine how the different parameters affect this critical mutation rate, we take two approaches: 1) agent-based simulations, and 2) an analytical approximation to the structured-population dynamics, employing an extension of the "pair-approximation" technique. The agent-based simulations are outlined in the main text, and we focus here on the analytical approximation.

Analytical Approximations

Generally speaking, a structured population may be represented by a graph (network), in which each node represents an individual in the population. Individuals may interact along edges of the graph, meaning that the state of one node may be affected by the states of other nodes that are connected to it, called "neighboring nodes". The states of these neighbors, however, depend in turn on the states of *their* neighbors, and so on, until the state of the entire network

is taken into account. Computer simulations, such as those described above, take into account the state of an entire network – whether the updating scheme be synchronous or asynchronous – either explicitly through direct simulation of the entire network or implicitly through the use of von Neumann boundary conditions. It can often be difficult, however, to perform exhaustive exploration of the parameter space using simulations alone, and gains may therefore be made from analytical approaches.

General "pair approximation" framework. This technique makes the simplifying assumption that the state of any given node in a structured population is affected only by the immediate neighbors of that node. [Note: Such an assumption is to spatial dimensions what the Markov property is to the temporal dimension.] The name "pair approximation" indicates the strategy to be used: graphs ranging from regular lattices to networks with complex structure are modeled by considering only randomly-chosen *pairs* of neighboring nodes (Baalen, 2000; Matsuda et al., 1992).

Nodes in the graph may be in different states. For example, a node may be occupied (denoted by a 1) or vacant (denoted by a 0). To generalize, we let the state of a node be denoted by index variables such as *i* or *j*. (In the forgoing example, these index variables may assume the values 0 or 1.) The frequency of nodes that are in state *i* is denoted by p_i. Suppose we choose an individual node at random – call this the "focal" node – and then choose an immediately neighboring node at random. The probability that the focal node is in state *i* and its immediate neighbor is in state *j* is denoted by p_{ij}. Now suppose we pick a focal node at random and observe that it is in state *i*. Given that we know that this focal node is in state *i*, a randomly-chosen immediate neighbor of the focal node will be in state *j* with probability $q_{j|i}$. By keeping track of such *conditional probabilities*, the structure of the population is accounted for to some extent. For many purposes, this Markov-like simplifying assumption yields analytical models that reproduce the population dynamics of spatially explicit computer simulations with surprising accuracy.

The equations resulting from the "pair approximation" technique encompass dynamics of singleton frequencies as well as the dynamics of all possible neighboring pairs. If we model a system in which two states are possible (say, 1 and 0, as in the above example), then the system of equations is:

$$\dot{p}_1 = p_0 r_{0\to 1} - p_1 r_{1\to 0}$$
$$\dot{p}_0 = p_1 r_{1\to 0} - p_0 r_{0\to 1}$$
$$\dot{p}_{11} = 2p_{01} r_{01\to 11} - 2p_{11} r_{11\to 01}$$
$$\dot{p}_{01} = 2p_{11} r_{11\to 01} + 2p_{00} r_{00\to 01} - 2p_{01} r_{01\to 11} - 2p_{01} r_{01\to 00}$$
$$\dot{p}_{00} = 2p_{01} r_{01\to 00} - 2p_{00} r_{00\to 01}$$

where the superscript dots denote time derivatives, $r_{i\to j}$ is the rate at which nodes transition from state *i* to state *j* and $r_{ij\to kl}$ is the rate at which neighboring pairs of nodes transition from state *ij* to state *kl*. These rates may themselves be functions of the p_i's, p_{ij}'s and $q_{i|j}$'s. The system of equations is greatly reduced by the identities $p_{ij} = q_{i|j} p_j = q_{j|i} p_i$, and because of the fact that there are only two states in our example, implying: $p_1 = 1 - p_0$, $q_{1|0} = 1 - q_{0|0}$ and $q_{1|1} = 1 - q_{0|1}$. If we simplify notation by letting $p_1 = p$ and $q_{1|1} = q$, then all variables may be expressed in terms of *p* and *q*:

$$p_1 = p$$
$$p_0 = 1 - p$$
$$p_{11} = pq$$
$$p_{01} = (1-q)p$$
$$p_{00} = \left(1 - \frac{(1-q)p}{1-p}\right)p$$

And the system of equations is reduced to:

$$\dot{p} = (1-p) r_{0\to 1} + p r_{1\to 0}$$
$$\dot{q} = 2(1-q) r_{01\to 11} - 2q r_{11\to 01} - (q(1-p)/p) r_{0\to 1} + q r_{1\to 0}$$

For a more thorough introduction to the pair approximation technique, and for instruction on how to extrapolate our results to more complex networks, see Van Baalen (Baalen, 2000). For a brief but clearly-written synopsis of the technique, see the introduction in Ellner (Ellner, 2001).

Mutation and selection on a network of competing social and asocial genotypes.

Overview. We model our analytical approximation after the agent-based simulations described above, and many features are therefore the same. In our model, each node on the network represents an individual (a cell, for example) that can be in one of two states: social or asocial. Social individuals help their neighbors by increasing their neighbors' fitness. Each social neighbor increases an individual's fitness by a factor of $(1+\varepsilon)$. Asocial individuals do not help their neighbors and the energy they save by not helping their neighbors gives them an intrinsic growth advantage, thereby increasing their own fitness by a factor of $(1+\beta)$. The maximum possible fitness is conferred to an asocial individual surrounded only by social neighbors. If each individual in the network has exactly *n* neighbors, then this maximum possible fitness is equal to $(1+\beta)(1+\varepsilon)^n$. We are interested in how the frequencies of social vs asocial genotypes change over time in the population and what their long-term (equilibrium) values are.

Network structure. In constructing our model, we had in mind a regular lattice network structure (modeled after the computer simulations; see Fig. S1a) in which each node has exactly *n* immediate neighbors. Given the generality of the pair approximation technique, however, our findings should encompass regular networks in general, and simple modifications (Baalen, 2000) should extend our results to networks with qualitatively different structures.

Selection. With some probability, an individual may "die" and the vacant node left by this death is filled by reproduction of a neighboring individual. Which one of an individual's *n* neighbors gets to reproduce and fill the vacant node depends in part on their fitnesses: the most fit of the *n* neighbors is most likely to be the one to reproduce and fill the vacant node. As in the computer simulations, however, there is a role for chance in this analytical model, and the most fit neighbor is not always the lucky one. The degree to which chance influences the outcome of this competition is determined in our model by a sampling exponent which we denote by α.

In Fig S1b, the green node inside the blue square represents the "focal" node. After one generation (one updating), this individual leaves no offspring (it "dies") with probability e^{-w_f}, where w_f denotes the fitness of the "focal" individual. Upon this individual's "death", the immediate neighbors inside the blue circle compete for the newly vacant node left by this death. In Fig S1b, each internal node on the network has $n = 4$ neighbors. The blue node inside the red square is an immediate neighbor of the focal individual. Let's suppose that, of the focal individual's four immediate neighbors, the one inside the red square is the third one to be

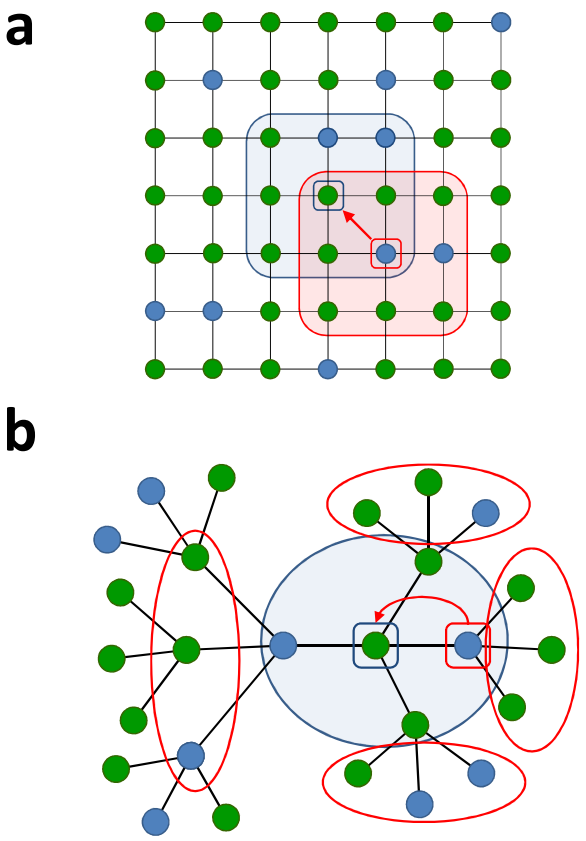

Figure S1 | **Two different regular network configurations**. Green nodes represent "social" individuals; blue nodes represent "asocial" individuals. Focal nodes are indicated by a small blue square; a candidate displacement comes from an immediate neighbor indicated by a small red square. Immediate neighbors of the focal individual are surrounded by a blue line; immediate neighbors of the candidate displacing neighbor are surrounded by a red line. **a**, a regular lattice network; **b**, a regular tree network.

considered. If we then let w_3 denote the fitness of this neighbor, then the probability that it will outcompete the other three immediate neighbors for the newly vacant focal node is $w_3^\alpha \Big/ \sum_{i=1}^{4} w_i^\alpha$, where w_1, w_2, and w_4 are the fitness of the other three immediate neighbors (inside the blue circle). In general, if we let w_k denote the fitness of the k^{th} of n immediate neighbor, then this neighbor will win the competition for the newly vacant node with probability $w_k^\alpha \Big/ \sum_{i=1}^{n} w_i^\alpha$.

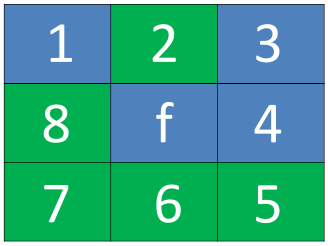

Figure S2 | **Example neighborhood of an asocial focal individual, indicated with an "f"**. A green square represents a "social" individual; a blue square represents an "asocial" individual. Neighbors 1, 3 and 4 are asocial; the others are social.

The governing equations. We let *s* denote "social" genotype and we let *a* denote "asocial" genotype. Because we have a two-state system as in the above example, the system may be reduced to two equations:

$$\dot{p}_s = p_a r(a \to s) - p_s r(s \to a)$$
$$\dot{p}_{ss} = 2p_{as} r(as \to ss) - 2p_{ss} r(ss \to as)$$

Singleton transition rates. The transition rate from asocial to social singletons is $r(a \to s)$. This rate is calculated as the probability that the asocial focal individual will 1) die and leave no offspring, and 2) be replaced by the offspring of a social neighbor. The probability that the focal individual leaves no offspring is e^{-w_f}. The fitness of the focal individual w_f is calculated by accounting for: 1) the growth advantage of being asocial ($1+\beta$), and 2) the help received from social neighbors $(1+\varepsilon)^X$, where *X* is the number of immediate neighbors that have the social genotype. Assuming multiplicative fitness, the fitness of the focal genotype is therefore: $w_f = (1+\beta)(1+\varepsilon)^X$ If we knew *a priori* how many neighbors were social, then we could simply substitute the *X* with that number to calculate w_f. In Fig. S2, for example, the fitness of the focal individual is $w_f = (1+\beta)(1+\varepsilon)^5$. Unfortunately, in our analytical model, we do not know *X* *a priori* but we do know what *X* is on average, and we can use this: $\bar{X} = nq_{s|a}$, where $q_{s|a}$ is the

probability that a randomly chosen neighbor of an asocial individual is social. The fitness of the focal individual is therefore calculated as: $w_f = (1+\beta)(1+\varepsilon)^{nq_{s|a}}$. To simplify the mathematics a bit, we assume that β and ε are small, and the fitness of the focal individual is therefore:

$$w_f \approx 1+\beta+\varepsilon n q_{s|a}$$

Next, in order to calculate the probability that the asocial focal individual will be replaced by the offspring of a social neighbor, we must know the fitnesses of all immediate neighbors. The trouble is, the fitness of an immediate neighbor depends on the states of *its* immediate neighbors. This is illustrated in Fig. S3, where the fitness of neighbor number 3 is calculated from the fact that it is asocial and has 4 social neighbors: $w_3 = (1+\beta)(1+\varepsilon)^4$. From this, the probability that neighbor number 3 displaces the focal individual is calculated as: $e^{-w_f} w_3^{\alpha} \Big/ \sum_{i=1}^{n} w_i^{\alpha}$. Again, however, in practice we do not know exactly how many neighbors of neighbor number 3 are social. But in this case, we do know the state of *one* neighbor: we know that one neighbor is the focal individual and we know that it is asocial. So that leaves only $n-1$ neighbors whose states are unknown. The *average* number of *social* neighbors of neighbor number 3 is therefore $(n-1)q_{s|a}$, and the fitness of neighbor number 3 may be calculated as $w_3 = (1+\beta)(1+\varepsilon)^{(n-1)q_{s|a}} \approx 1+\beta+\varepsilon(n-1)q_{s|a}$. This calculation will apply to all asocial neighbors of the focal individual, and we will therefore define: $w_a \approx 1+\beta+\varepsilon(n-1)q_{s|a}$. Similar logic applies to social neighbors of the focal individual, each of whose average fitness is defined as: $w_s \approx 1+\varepsilon(n-1)q_{s|s}$. On average, the asocial focal individual will be surrounded by $n_s = nq_{s|a}$ social neighbors each of whose fitness is, on average, $w_s \approx 1+\varepsilon(n-1)q_{s|s}$ and by $n_a = nq_{a|a}$ asocial neighbors each of whose fitness

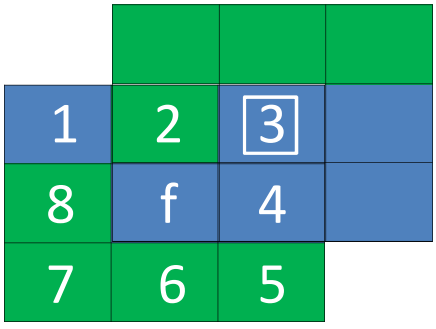

Figure S3 | Example neighborhood of an asocial focal individual, and the neighborhood of neighbor number 3. A green square represents a "social" individual; a blue square represents an "asocial" individual. Neighbor number 3 has four social and four asocial neighbors.

is, on average, $w_a \approx 1+\beta+\varepsilon(n-1)q_{s|a}$. Given that the asocial focal individual dies and leaves no offspring, therefore, the probability that the vacant node will be filled by a social individual is: $\frac{n_s w_s^\alpha}{n_s w_s^\alpha + n_a w_a^\alpha} = \frac{nq_{s|a} w_s^\alpha}{nq_{s|a} w_s^\alpha + nq_{a|a} w_a^\alpha} = \frac{q_{s|a} w_s^\alpha}{q_{s|a} w_s^\alpha + q_{a|a} w_a^\alpha}$. In addition to displacement of the asocial focal individual by the offspring of a social neighbor, there is another way that the individual on the focal node can transition from asocial to social, namely through mutation. Let $\mu_{a\to s}$ denote the per-generation rate of mutation from asocial to social genotype. When all of the above factors are accounted for, the transition rate becomes:

$$r(a \to s) = \frac{e^{-w_f} q_{s|a} w_s^{\alpha}}{q_{s|a} w_s^{\alpha} + q_{a|a} w_a^{\alpha}} + \mu_{a \to s}$$

where:

$$w_f = 1 + \beta + \varepsilon n q_{s|a}$$

$$w_s = 1 + \varepsilon (n-1) q_{s|s}$$

$$w_a = 1 + \beta + \varepsilon (n-1) q_{s|a}$$

(1.1)

Likewise, similar logic leads to:

$$r(s \to a) = \frac{e^{-w_f} q_{a|s} w_a^{\alpha}}{q_{a|s} w_a^{\alpha} + q_{s|s} w_s^{\alpha}} + \mu_{s \to a}$$

where:

$$w_f = 1 + \varepsilon n q_{s|s}$$

$$w_s = 1 + \varepsilon + \varepsilon (n-1) q_{s|s}$$

$$w_a = 1 + \beta + \varepsilon + \varepsilon (n-1) q_{s|a}$$

(1.2)

Pair transition rates. Calculation of these rates follows similar logic to the calculation of singleton rates. The principle difference is that now, in addition to the state of the focal individual, the state of one neighbor is also treated as known, i.e., the states of pairs are treated as known. This approach is depicted in Fig. S4, where the pair in question is indicated by a white square. This figure depicts the transition from an {*as*} pair to an {*ss*} pair. Here, we treat the *a* in the {*as*} pair as the focal individual. The fitness of this focal individual can be directly calculated from the explicit configuration depicted in Fig. S4: $w_f = (1+\beta)(1+\varepsilon)^5$. In practice, our

analytical model does not specify the explicit configuration; however, we do know that *one* neighbor of the asocial focal individual is social; we know this from the fact that the pair in question is an {*as*} pair. In our analytical model, the fitness of the focal individual is calculated as $w_f = (1+\beta)(1+\varepsilon)^{1+(n-1)q_{s|a}} \approx 1+\beta+\varepsilon+\varepsilon(n-1)q_{s|a}$. The other immediate neighbors of the focal individual (the neighbors outside of the white square) have fitness $w_a \approx 1+\beta+\varepsilon(n-1)q_{s|a}$ if they are asocial and $w_s \approx 1+\varepsilon(n-1)q_{s|s}$ if they are social.

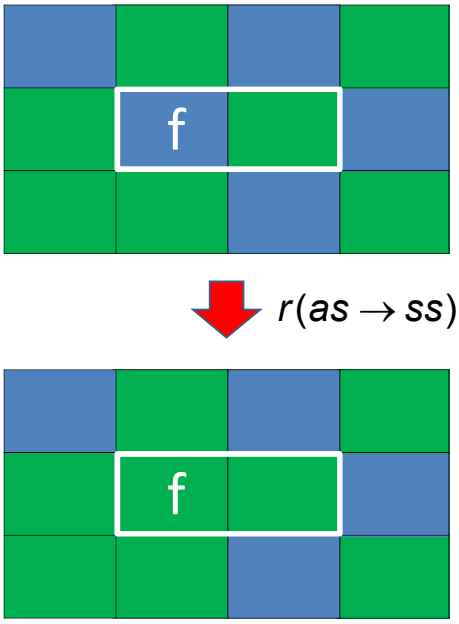

Figure S4 | Example neighborhood for calculation of pair transition rates. A green square represents a "social" individual; a blue square represents an "asocial" individual. The left-hand member of the pair is taken to be the "focal" individual because it is the one that changes.

$$r(as \to ss) = \frac{e^{-w_f}(1+(n-1)q_{s|a})w_s^{\alpha}}{(1+(n-1)q_{s|a})w_s^{\alpha} + (n-1)q_{a|a}w_a^{\alpha}} + \mu_{a\to s}$$

where:

$$w_f = 1+\beta+\varepsilon+\varepsilon(n-1)q_{s|a}$$

$$w_s = 1+\varepsilon(n-1)q_{s|s}$$

$$w_a = 1+\beta+\varepsilon(n-1)q_{s|a}$$

(1.3)

In the transition from {*ss*} pairs to {*as*} pairs, the left-hand side of the pair is treated as the focal node (i.e., the side that transitions from *s* to *a*). Logic similar to that used above leads to:

$$r(ss \to as) = \frac{e^{-w_f}(n-1)q_{a|s}w_a^{\alpha}}{(n-1)q_{a|s}w_a^{\alpha} + (1+(n-1)q_{s|s})w_s^{\alpha}} + \mu_{s\to a}$$

where: $w_f = 1+\varepsilon+\varepsilon(n-1)q_{s|s}$

$w_s = 1+\varepsilon+\varepsilon(n-1)q_{s|s}$

$w_a = 1+\beta+\varepsilon(n-1)q_{s|a}$

(1.4)

The system of equations is reduced by the identities $p_{as} = q_{a|s}p_s = q_{s|a}p_a$ and $p_{ss} = q_{s|s}p_s$, and because of the fact that there are only two states in our model, implying: $p_a = 1-p_s$, $q_{s|a} = 1-q_{a|a}$ and $q_{s|s} = 1-q_{a|s}$. If we simplify notation by letting $p_s = p$ and $q_{s|s} = q$, then all variables may be expressed in terms of *p* and *q*:

$p_s = p$ $\quad$ $q_{s|s} = q$

$p_a = 1-p$ $\quad$ $q_{a|s} = 1-q$

$p_{ss} = pq$ $\quad$ and $\quad$ $q_{s|a} = \frac{(1-q)p}{1-p}$

$p_{as} = (1-q)p$ $\quad$ $q_{a|a} = 1-\frac{(1-q)p}{1-p}$

$p_{aa} = \left(1-\frac{(1-q)p}{1-p}\right)(1-p)$

(1.5)

And the system of equations is reduced to:

The governing equations:

$$\dot{p} = (1-p)r(a \to s) - pr(s \to a)$$
$$\dot{q} = 2(1-q)r(as \to ss) - 2qr(ss \to as) - (q(1-p)/p)r(a \to s) + qr(s \to a)$$

(1.6)

Solution procedure.

The dynamics of "social" genotype frequency, p, are determined by solution of the governing equations (1.6). This is achieved by substituting into these governing equations the expressions for the rate terms – equations (1.1) through (1.4) – and then making the substitutions listed in (1.5) to put everything in terms of p and q.

Example numerical solution. If we start with a population of cooperators with very few isolated cheaters, then our initial conditions would be something like this: $p(0) = 1 - 10^{-4}$ and $q(0) = 1 - 10^{-4}$. For the parameters of our model, we will use the values: $\beta = 0.01$, $\varepsilon = 0.1$, $n = 8$ and $\alpha = 2$. Cooperators that become cheaters by mutation do so by losing a cooperation function, i.e., by receiving any one of hundreds of possible knockout mutations in a gene encoding for some aspect of cooperation. Cheaters that become cooperators by mutation, on the other hand, do so by either recuperating a lost function by receiving exactly the right mutation or acquiring a novel cooperating function. The mutation rate from cooperator to cheater, $\mu_{s\to a}$, is therefore likely to be much higher than the mutation rate from cheater to cooperator, $\mu_{a\to s}$. Initially, we will assume that $\mu_{a\to s} = 0$. (This assumption is convenient for analysis of the governing equations (see below), because looking for conditions under which cooperation is lost then reduces to the problem of finding the requirements for the absorbing state $\hat{p} = 0$ to be stable.) Numerical solution of the governing equations is plotted in Fig. S5 for $\mu_{s\to a} = 0.1$. This plot elucidates a characteristic of the spatial structure that will become instrumental in our analysis of the system

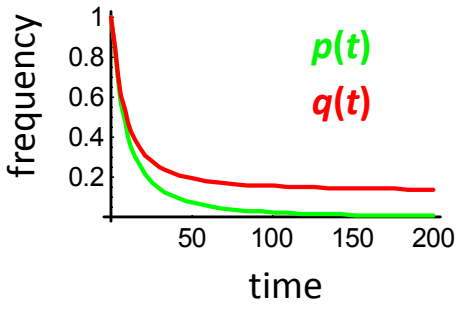

Figure S5a | **Numerical solution of the governing pair-approximation equations**. Cooperator frequencies. Mutation rate is above the threshold.

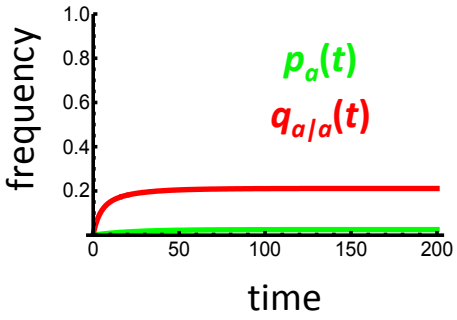

Figure S5b | **Numerical solution of the governing pair-approximation equations**. Cheater frequencies. Mutation rate is below the threshold.

and our understanding of the mutation rate threshold: when cooperators are common, there is little difference between cooperator frequency, $p(t)$, and the probability that a given cooperator has an immediate neighbor who is also a cooperator, $q(t)$; when cooperators are rare, however, there is a big difference between $p(t)$ and $q(t)$, indicating that while a cooperator might be hard to find, when you do find one you will likely find that it has a cooperating neighbor as well (following the adage "where there's one there's two"). Their very small presence, therefore, is due more to extremely rare clusters of cooperators and less to isolated individual cooperators. The same is true for cheaters, as shown in Fig S5b: their very small presence is due more to extremely rare clusters of cheaters and less to isolated individual cheaters. Clustering is disadvantageous to cheaters, however, and this helps to keep cheaters in check below the threshold mutation rate.

Cheater fitness.

Global relative fitness. On average, cooperators and cheaters will have absolute fitnesses given by

$$w_a = 1 + \beta + nq_{s|a}\varepsilon \quad \text{and} \quad w_s = 1 + nq_{s|s}\varepsilon ,$$

respectively. The global relative fitness of cheaters is therefore calculated as:

$$\tilde{w}_a = w_a / (p_a w_a + p_s w_s) = w_a / ((1-p)w_a + pw_s).$$

Figure S6 here and Fig. 3 in the main text plot this global relative fitness of cheaters as a function of cheater frequency for high and low mutation rates.

Local relative fitness. The absolute fitness of a cheater's immediate neighborhood is calculated as the sum of two terms: 1) the frequency of cooperators in the neighborhood, $q_{s|a}$, multiplied by

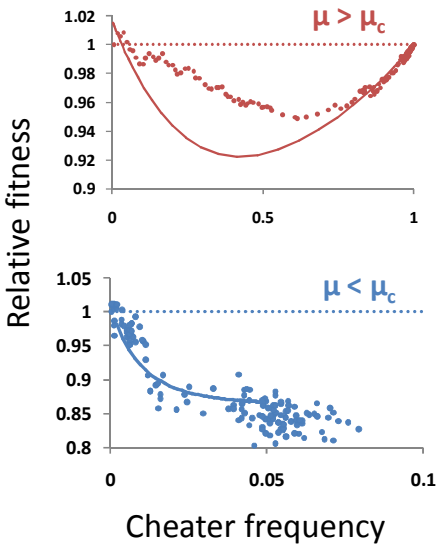

Figure S6 | Global relative fitness of cheaters as a function of cheater frequency. Curves plot numerical solution of pair-approximation equations, and points plot simulation results. Mutation rate is above the critical mutation rate in the top plot and below the critical mutation rate in the bottom plot.

their average fitness, $1+(n-1)q_{s|s}\varepsilon$, and 2) the frequency of cheaters in the neighborhood, $q_{a|a}$, multiplied by their average fitness, $1+\beta+(n-1)q_{s|a}\varepsilon$. The absolute fitness averaged over the cheater and its immediate neighbors is therefore:

$$\bar{w}_a = \left(nq_{s|a}(1+(n-1)q_{s|s}\varepsilon) + nq_{a|a}(1+\beta+(n-1)q_{s|a}\varepsilon) + w_a \right)/(n+1)$$

and the local relative fitness of cheaters is:

$$\lambda_a = w_a / \bar{w}_a .$$

Figure S7 here and Fig. 3 in the main text plot this local relative fitness of cheaters as a function of cheater frequency for high and low mutation rates.

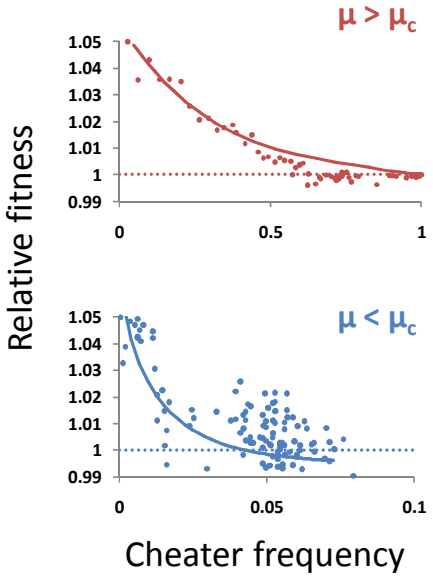

Figure S7 | Local relative fitness of cheaters as a function of cheater frequency. Curves plot numerical solution of pair-approximation equations, and points plot simulation results. Mutation rate is above the critical mutation rate in the top plot and below the critical mutation rate in the bottom plot.

Analysis.

The goal in our analysis is to determine the conditions under which cooperation is lost. Initially, we make the convenient assumption that $\mu_{a\to s}=0$, as discussed above, so that it is possible for cooperators to achieve a frequency of zero. If cooperation is lost completely, then the absorbing state $p(t)=0$ is reached and is stable (meaning biologically that if you re-introduce a small number of cooperators after they have been completely eliminated from the population, they will not subsequently grow in frequency but will be eliminated again). We therefore proceed with a standard stability analysis, defining the new functions f_p and f_q to be the right-hand-sides of the governing equations (1.6), so that these equations simply read: $\dot{p}=f_p$ and $\dot{q}=f_q$. The Jacobian is then computed as usual:

$J = \begin{pmatrix} \frac{\partial f_p}{\partial p} & \frac{\partial f_p}{\partial q} \\ \frac{\partial f_q}{\partial p} & \frac{\partial f_q}{\partial q} \end{pmatrix}$, and evaluated at the steady state defined by: $\hat{p} = 0$ and $\hat{q} = q_0$. We find that the eigenvalues of the resulting matrix are both negative, implying stability of this steady state, when mutation rate is greater than a critical value equal to:

$$\mu_c = (1-q_0)\left(e^{-(1+\beta)}\left(\frac{1+(n-1)q_0\varepsilon}{1+\beta}\right)^{\alpha} - e^{-(1+nq_0\varepsilon)}\left(1+q_0\left[\left(\frac{1+((n-1)q_0+1)\varepsilon}{1+\beta+\varepsilon}\right)^{\alpha}-1\right]\right)^{-1}\right) \quad (1.7)$$

By numerical methods, we generated parameter values at random with the ranges: $\varepsilon \in (0, 0.3)$, $\beta \in (0, \varepsilon)$, $n \in (2, 50)$, and $Log_{10}\mu \in (-5, -1)$. Using these randomly-generated parameter values, we solved the equation $f_q\big|_{p=0} = 0$ for q, thereby giving a computed value for q_0. We repeated this for 500 different randomly-generated sets of parameters and found that q_0 was sensitive almost exclusively to the number of neighbors, n, and a plot of $\ln q_0$ as a function of $\ln n$ is shown in Fig. S8. A linear regression of $\ln q_0$ on $\ln n$ yields the equation: $q_0 = k_1 n^{k_2}$, where regression constants were estimated to be $k_1 \cong 1.16$ and $k_2 \cong -1.06$ with small error, from which we make the reasonable approximation, $q_0 \approx 1/n$. Inserting this approximation into (1.7) and eliminating small terms gives an approximate expression for the critical mutation rate:

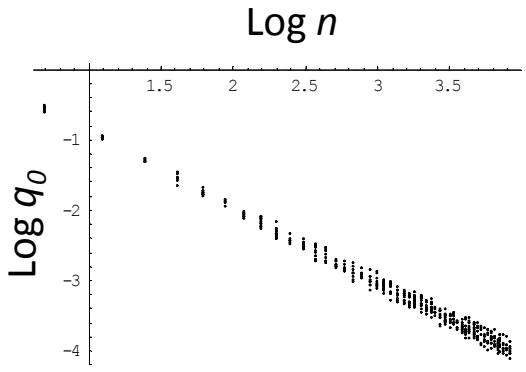

Figure S8 | Computed values for q_0 show a power-law relationship with mean number of immediate neighbors, n.

$$\mu_c \approx \frac{n-1}{n}\left(e^{\alpha\varepsilon-(1+\alpha)\beta-1} - e^{-\varepsilon-1}\right) \quad \xrightarrow{n} \quad e^{\alpha\varepsilon-(1+\alpha)\beta-1} - e^{-\varepsilon-1} \quad (1.8)$$

Inaccuracies of the pair approximation technique due to demographic stochasticity are well-documented (Baalen, 2000); in comparing this analytical result with those of simulations, we have found the analytical result was off by an apparently constant factor, such that:

$$\mu_c \propto \quad e^{\alpha\varepsilon-(1+\alpha)\beta-1} - e^{-\varepsilon-1},$$

where the proportionality constant was estimated by least squares to be 0.081 (see following section). For $k \approx 0.081$ and for small values of ε and β, therefore, we have:

$$\mu_c \approx ke^{-1}(\alpha+1)(\phi\varepsilon-\beta).$$

Estimating the constant, k. The constant *k* accounts for demographic stochasticity, i.e., it accounts for well-documented shortcomings of the "pair approximation" technique used in our analysis (Baalen, 2000), and is estimated by least-squares fitting to the results of simulations to be; this estimate is remarkably consistent across a wide variety of regular lattices, random networks and scale-free networks. Figure S9 plots values for μ_c calculated by routine of bisection(Press, 1992) from simulated populations on regular lattices, random networks and scale-free networks. The constant *k* was estimated by least squares fit of equation (1.8) to the values for μ_c computed from simulations on all the different networks.

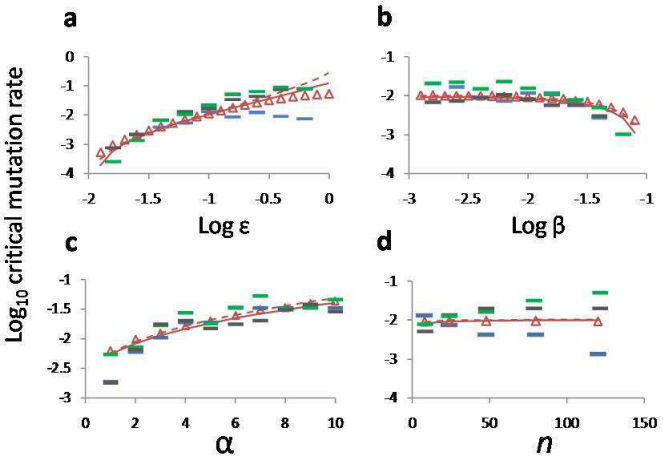

Figure S9 | Dependence of critical mutation rate μ_c on the various parameters. The red solid lines plot the full analytical expression for the equilibrium obtained by solving the simultaneous equations $f_p = 0$ and $f_q = 0$, the red dashed lines plot the approximate expression given in the main text, and the red triangles represent critical mutation rates derived from numerical solution of the dynamic equations $\dot{p} = f_p$ and $\dot{q} = f_q$. Critical mutation rates were determined from simulations and are plotted as: blue bars for a regular two-dimensional grid, grey bars for random networks, and green bars for scale-free networks. (In panel (**d**), the increasing discrepancy among results from the different network structures is due to the fact that the mean-shortest-path is affected differently by increasing *n*: it is drastically reduced in random and scale-free networks but is not changed in the regular grid.) We believe that the absolute mutation rate values plotted in this figure could in fact be quite conservatively high because of the large discrepancy between our default values of $\varepsilon = 0.1$ and $\beta = 0.01$; when these values are closer together, critical mutation rates can be much lower, as intuited from our analytical expression (main text). Other default parameters were: $\alpha = 2$ and $n = 8$.

Some indications that the loss of cooperation is a second-order phase transition. Figure 1 in the main text shows a sharp inflection separating the region in which cooperation can persist from the region where it is lost. The frequency of cooperators in the context of phase transitions is the order parameter. The sharp inflection in this order parameter is suggestive of second-order phase-transition behavior. Other evidence in support of this classification is the divergence of the "susceptibility" to increases in ε at the critical point μ_c -- determined numerically as an essentially vertical increase in the order parameter as ε increases when $\mu = \mu_c$. Further supporting evidence is found in Fig. 4 of the main text: the fraction of populations in which the largest cluster is a cheater cluster goes very quickly from zero to one, marking the transition to the loss of cooperation. Finally, when $\mu = \mu_c - \delta$, for small δ, the cluster size distribution has a power-law tail, as shown in Fig. S10.

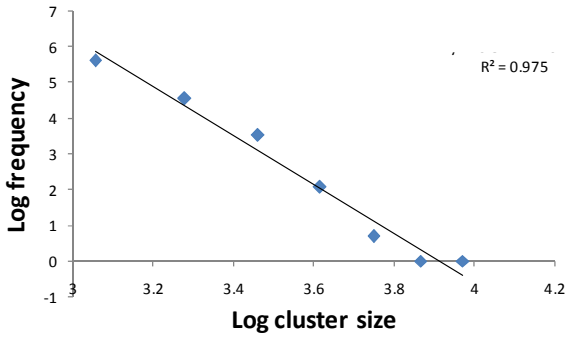

Figure S10 | Some evidence of a power-law tail on the cluster-size distribution.

Baalen, V. 2000 In The Geometry of Ecological Interactions: Simplifying Spatial Complexity(Eds, Dieckmann, U., Law, R. and Metz, J.A.J.) Cambridge University Press, pp. 359-387.

Ellner, S.P. 2001, J. Theor. Biol., 210(4), pp. 435-447.

Matsuda, H., Ogita, N., Sasaki, A., Sato, K. 1992, Prog. Theor. Phys., 88, pp. 1035-1049.

Press, W.H. 1992 Numerical recipes in C : the art of scientific computing, Cambridge University Press, pp. xxvi, 994 p.