The Evolution of Resilience

David K. Levine[1]

Abstract

I analyze a social evolutionary model in which there is a non-excludable public good that reduces the chances of catastrophe. I show that while resilient types that produce the public good cannot survive head-to-head competition with other more selfish types in a fixed population, they do if catastrophes reduce population.

Email address: david@dklevine.com (David K. Levine)

[1]Department of Economics, Royal Holloway University of London and WUSTL

Acknowledgements: First version: September 10, 2022. I would like to thank Rohan Dutta, John Mair, Andrea Mattozzi, Salvatore Modica and H. Peyton Young. I am particularly indebted to Bruno de Albuquerque Furtado for a careful reading of the paper and suggesting many corrections and improvements. I gratefully acknowledge support from the MIUR PRIN 2017 n. 2017H5KPLL_01 and the Leverhulme Trust.

1. Introduction

I analyze a social evolutionary model in which there is a non-excludable public good that reduces the chances of a catastrophe. The novel contribution of this paper is to demonstrate how resilient types that cannot survive head-to-head competition with other more selfish types, nevertheless survive if they provide protection against catastrophes that reduce the total population.

Models of stochastic social evolution as developed by Young (1993), Kandori, Mailath and Rob (1993) Ellison (2000), Cui and Zhai (2010), Peski (2010), Levine and Modica (2016a) and Newton (2021), among others, describe predominant behavior in the long run when individual group members adopt strategies that promise increased utility. These models have been used to show, for example, that if types can be observed by other members, then in a population matched to play repeated prisoner's dilemma games, cooperation supported by punishment strategies can emerge in the long run. Implicit in these results is that the public good is excludable, as it need only be provided to those of the same type. However, in practice we observe not only cooperation in individual interactions, but the production of public goods that benefit both cooperators and cheaters equally. If a non-excludable public good is neutral, in the sense of providing equal advantage to everyone, can evolution favor the production of such public goods?

In the standard setting of group selection with identifiable types and a fixed population, I show that under certain conditions the answer is no. As in the literature of the evolution of cooperation in individual matching games, I use a group selection model in which it is possible to distinguish the types that contribute from types that do not. Although the assumption of identifiable types is controversial, I indicate in Section 7 that in the setting of social evolution it can be considered reasonable. In such a model, a population of types that punish other types but not their own has good stability properties: it requires a large invasion of mutants to reduce punishment enough to acquire an advantage over the incumbents. Writers in the evolutionary literature such as Ellison (2000) have long been aware of this.

Among types that punish, with equal numbers of a nasty type that does not produce the public good, and a type that does produce the public good, the nasty type yields more utility. Consequently, under certain conditions this implies that the nasty types will predominate in the long run.

Why then are neutral non-excludable public goods produced? I introduce a novel element into the analysis: the possibility of catastrophic reductions in population that are mitigated by the public good. I refer to this as resilience. After a catastrophic reduction in population, the nasty types still have an advantage. However, with a much smaller population, evolution has a much higher degree of randomness. While frequently nasty types will reemerge, there is a good chance that types that do produce the public good will emerge instead. In this case the frequency with which types are present depends less on whether they have an immediate advantage, and more on whether they avoid catastrophes. This shifts the evolutionary advantage towards types that produce the public good.

In the formal analysis I show that if the public good is sufficiently effective in reducing the frequency of catastrophes, then indeed the predominant long run behavior will involve production of the public good. More strongly, I show that with a large enough population, the predominant type will be one that maximizes output of the public good subject to the constraint that punishment is adequate to prevent free-riding. As this theoretical argument involves taking limits, it is reasonable to ask if the range of parameters for which the result holds is unrealistically extreme. To answer this question, I conduct simulations showing that the theoretical results hold for plausible parameter values.

As indicated, I propose a specific mechanism through which types that provide the public good survive and thrive: they do so on account of catastrophes that reduce population and make it easier for new types to enter. Human catastrophes can take place in many ways: they may be ecological as in Diamond (2005), may involve losses in war as in Levine and Modica

(2013), or may simply involve civil strife, arising, for example, out of inequality. In this respect, the insurance schemes studied by Townsend (1994) may be seen to reduce the chances of catastrophe.

Catastrophes that reduce population are not uncommon. Some of the more notable ones are the severe glaciation of the Younger Dryas period(Peteet (1995)), the depopulation of Rome between the second and sixth centuries AD (Twine (1992)), and the Great Famine of the 1840s in Ireland (Ross (2002)). Notice that the theory does not say that new, more resilient institutions will emerge from these catastrophes. Rather, it says that institutional change is more likely after such a catastrophe. One example is the decline of serfdom in England after the Black Death plague (Bailey (2014)).

Finally, the idea that population reductions make it easier for new entrants to thrive is strongly supported in the biological record. A key fact (see for example Jablonski (2001)) is that mass extinctions caused by events such as asteroid strikes not only reduce the population of existing species, but subsequently lead to a great increase in the diversity of species.

Related Literature and Marginal Contribution

The idea that evolution favors functional institutions is scarcely new. For example, in a repeated game setting with enough structure, evolution favors the grim trigger strategy. This is the thrust of the work of Bowles and Choi (2013) and Choi and Bowles (2007) in their study of the emergence of institutions in the post-Dryas period. More abstract results can be found in Axelrod and Hamilton (1981), Binmore and Samuelson (1992), Johnson, Levine and Pesendorfer (2001), Dal Bo and Pujals (2015), and Juang and Sabourian (2021) among others. In the biological literature this model has been called that of parochial altruism.

In the repeated game framework, the public good is excludable and punishment takes place through repetition. By contrast, in this paper I consider a non-excludable public good and contemporaneous punishment. Even in the context of cartels, incentives are often provided through contemporaneous fines rather than future market action. More broadly, industry operates through contemporaneous rewards in the form of wages, and criminal justice operates through punishments, such as imprisonment, not through revenge-driven retaliation. The same is true of the ostracism and peer pressure that Ostrom (1990) documents are widely used for public goods provision.

In the contemporaneous punishment environment, there is a strong evolutionary force towards conformity. However, many social norms can be supported as conformist through punishment of deviators. Consider the following intuition. Suppose that there are two equal-sized groups, one of which contributes, and one of which does not, and that both punish each other for failing to conform. Then everyone benefits from the public good and suffers the same punishment. Here, it is the non-contributing group that does best, as they avoid paying for the public good. Put differently, the possibility of punishing deviators creates a strong evolutionary force towards conformity - but not obviously towards efficiency. Notice the key difference with prisoner's dilemma type models: here the free-riders are able to punish the contributors. Rusch (2014) observes that the inability of free-riders to punish is a key limitation of the parochial altruism model. That limitation is lifted here.

This pressure towards conformity connects the results here to the literature on conformity. Benassy (1998) and Akerlof and Kranton (2005) show how conformity can result in dysfunctionality. By contrast, Coase (1960) and Ostrom (1990) show that we often see the provision of non-excludable public goods. Here, I show that while evolution can lead to dysfunctionality when catastrophes are not important, they lead to the provision of non-excludable public goods that increase resilience to catastrophes, when those are important.

The presence of aggregate risk in the form of catastrophes is reminiscent of the work of Robson (1992) and Heller and Nehama (2023) on aggregate risk preferences and evolutionary growth, while the idea of collapse and rebuilding occurs in the form of large deviation events that upset self-confirming equilibrium in Cho and Kasa (2014). The literature on evolution under conflict also has related results. The theory that it leads to hegemony, as in Levine and Modica (2012), Levine and Modica (2013), Levine and Modica (2016a) Levine and Modica (2022), and Bilancini, Boncinelli and Marcos-Prieto (2022), has a similar flavor to the results here. In particular, the evolution of the ability to withstand "outside pressure" is similar to resilience in this paper. However, those papers consider neither punishment nor conformity. They also are limited to explaining how a public good can protect against being absorbed by another society, but have nothing to say about natural disasters.

The model here is one of group selection with identifiable types. These models are controversial because they rely on types being able to identify themselves and outsiders. Indeed, conformist types here are related to the biological notion of a "green-beard", who helps their own types and hurts other types,[2] but it differs in the sense that a conformist type does not necessarily help their own type, and if they do help, they help all types equally. It is similar to what Gardner and West (2010) call a "facultative harming green-beard." While the literature on biological evolution recognizes that green-beards promote cooperation, it is generally viewed that they are unsuccessful in doing so because they are subject to invasion by fake green-beards. In the current setting of human behavior, this problem is mitigated by the fact that it is generally possible to observe whether people conform to the social norm and whether they punish those who fail to do so. I elaborate on this point in section 7.

In the context of evolution favoring efficiency, the early work of Winter (1971) showed how the survival of more profitable firms leads to efficient competitive equilibrium, which has a more modern incarnation in Seranno and Volij (2008). There is also a literature that studies the evolution of altruism in an environment without the possibility of punishment. Positive assortative matching, as in Alger and Weibull (2013), and voluntary migra-

[2]See Hamilton (1964a), Hamilton (1964b) and Jansen and Van Baalen (2006).

tion, as in Ely (2003), for example, gives reason for the survival of altruism. As made explicit in Dutta, Levine and Modica (2022), this type of altruism may serve in a useful way as the "grease on the wheels of resilience."

2. The Model

In each of $t = 1, 2, \ldots$ periods there is a finite group whose members are assigned to a finite collection of types I. In period t there is an integer number $n_t^i > 0$ of members of type i and the vector n_t constitutes the *state*. A number $m^i > 0$ of each type are *residual* and these populations do not change over time,[3] while the remaining *variable* populations of $n_t^i - m^i \geq 0$ evolve according to an evolutionary process that is Markov on the state space and is described below. I denote by $M = \sum_i m^i$ the residual population, and by $N_t = \sum_i n_t^i$ the total population. It is convenient as well to record the population fractions $\phi_t^i = n_t^i / N_t$.

Each type $i \in I$ provides *resilience* $w^i \geq 0$ which is a public good, and imposes a utility punishment $P^i/N_t \geq 0$ on each member that is of a different type. Both resilience and punishment are costly, and type i incurs a private utility cost $c^i = \alpha w^i + \psi \max\{0, P^i - \underline{P}\}$ with $\alpha, \psi, \underline{P} > 0$. The total private cost, including punishment by other types, incurred by an individual of type i at t can then be computed as

$$C^i(\phi_t) = c^i + \sum_{j \neq i} \phi_t^j P^j.$$

Resilience provides no utility but, as explained below, reduces the chances of a catastrophic drop in population.

This formulation allows for a wide variety of types. A *selfish type* provides

[3]Having a residual population avoids the problem of types becoming extinct. An alternative would be to allow extinction but reintroduce extinct types through mutation. The basic results would not be changed if the probability of catastrophe is high relative to the mutation rate, but the details and computations are considerably more difficult due to the need to account for mutations even in the absence of a catastrophe.

no resilience and does not punish: $w^i = 0, P^i = 0$ so minimizes cost at $c^i = 0$. A *nasty type* also provides no resilience, but provides the maximum punishment consistent with no cost: $w^i = 0, P^i = \underline{P}$. An *altruistic type* provides resilience, but does not punish: $w^i > 0, P^i = 0$ so $c^i > 0$.

A *conformist type* i satisfies $c^i < c^j + P^i$ for every other type $j \neq i$. That is, when all others are conformists of the same type, the conformists get strictly more utility than any other type. A nasty type, at least, is conformist, and I assume that there is exactly one type that is nasty. By contrast, an altruistic type is not conformist: deviating to nastiness reduces cost without incurring punishment. An immediate consequence of the existence of the nasty type is that the conformism constraints reduce to $c^i < P^i$.

The evolutionary process

Two forces determine the evolution of the variable population: catastrophes and arrivals. Catastrophes are governed by $v > 0, \underline{N}, \overline{N}$, where v is a measure of the importance of resilience in preventing catastrophes, $\underline{N}$ is the population after a catastrophe, and $\overline{N} > \underline{N}$ is the largest sustainable population.

I study the case where catastrophes and arrivals of the unfit are rare events, and I use a single parameter $1 > \epsilon > 0$ of how rare these events are. Specifically, it will be convenient to specify probabilities in terms of their resistances: probabilities at time t will have the form

$$h(n_t) = H(n_t)\epsilon^{r(n_t)}$$

where $H(n_t)$ is uniformly bounded between zero and one, and the resistance $r(n_t) \geq 0$. Roughly speaking, higher resistance, as measured by $r(n_t)$, means an order of magnitude of lower probability. By contrast, $H(n_t)$ is a scaling factor that allows for variations in probabilities that are independent of the parameter ϵ, the measure of how rare, rare events are. As discussed below, a key assumption is that greater resilience results in greater resistance to catastrophes.

The Markov process governing the evolution of n_t over time is defined by transition probabilities, which in turn are determined by a sequence of random events.

1. A catastrophe occurs with resistance $r^0_\kappa(n_t)$. If a catastrophe occurs, then $N_t - \underline{N}$ members are randomly removed from the variable population.

2. If a catastrophe does not occur, then with some probability exactly one type i will receive one additional member. This is referred to as *arrival* to the type and has resistance $r^i_p(n_t)$.

3. If an arrival leads to a population that is greater than $\overline{N}$, then one member is randomly removed from the variable population.

Formally, denote by $\tilde{N}(n_t, K)[n_{t+1}]$ the probability of the vector n_{t+1} induced by removing K members randomly from the variable population in n_t. Let n_t^{+i} be the population derived from n_t by increasing the population of type i by one. Then the transition probabilities are given by

$$\Pr(n_{t+1}|n_t) = \left(H_\kappa(n_t)\epsilon^{r^0_\kappa(n_t)}\right) \tilde{N}(n_t, N_t - \underline{N})[n_{t+1}]$$

$$+ \left(1 - H_\kappa(n_t)\epsilon^{r^0_\kappa(n_t)}\right) \sum_{i \in I} \mathbf{1}\left(n_{t+1} = n_t\right) \left(1 - H^i_p(n_t)\epsilon^{r^i_p(n_t)}\right)$$

$$+ \left(1 - H_\kappa(n_t)\epsilon^{r^0_\kappa(n_t)}\right) \sum_{i \in I} \mathbf{1}\left(n_t < \overline{N}\right) \mathbf{1}\left(n_{t+1} = n_t^{+i}\right) H^i_p(n_t)\epsilon^{r^i_p(n_t)}$$

$$+ \left(1 - H_\kappa(n_t)\epsilon^{r^0_\kappa(n_t)}\right) \sum_{i \in I} \mathbf{1}\left(n_t = \overline{N}\right) H^i_p(n_t)\epsilon^{r^i_p(n_t)} \tilde{N}(n_t^{+i}, 1)[n_{t+1}]. \quad (2.1)$$

The first line corresponds to a catastrophe, and the final three lines to no catastrophe. The second line corresponds to no arrival taking place. The third line is for a population below the limit, in which case an arrival is added according to $H^i_p(n_t)\epsilon^{r^i_p(n_t)}$. The fourth line is for a population already at the limit, in which case an arrival is added according to $H^i_p(n_t)\epsilon^{r^i_p(n_t)}$, then one member is subtracted according to $\tilde{N}(n_t^{+i}, 1)[n_{t+1}]$. As indicated above, $H^i_p(n_t)$ and $H^i_p(n_t)$ are scaling factors for probabilities, but do not effect resistance.

Assumptions about the process

Define per capita resilience as $W_t = \sum_i \phi_t^i w^i$. I assume that resistance to a catastrophe is given by $r_\kappa^0(n_t) = r_\kappa(vW_t, N_t/\overline{N})$. The function r_κ is assumed to be strictly positive and is strictly increasing in its first argument with derivative bounded uniformly away from zero. That is, for fixed N_t, increasing resilience increases resistance to catastrophe. If $\overline{N}$ is determined by the geographical area occupied by the group, then $N_t/\overline{N}$ is a measure of population density. I am agnostic as to the dependence of resistance on the population density. If catastrophes are due to invasion by neighbors occupying a similar geographical area, then a large population density is likely to reduce their probability. If catastrophes are due to ecological collapse, a large population density is likely to increase their probability.

The resistance to arrival $r_p^i(n_t)$ is assumed to depend only on the private cost of type i relative to the cost of other types. Specifically, letting z be a $I-1$ vector of non-negative pairs (ϕ^j, C^j), there is a common arrival resistance function $r_p(C^i, z)$ and $r_p^i(n_t) = r_p\left(C^i(\phi_t), \left(\phi_t^j, C^j(\phi_t)\right)_{j\neq i}\right)$. The common arrival resistance function r_p is assumed weakly increasing in C^i and anonymous with respect to z in the sense that it is invariant to permuting the pairs.

Denote the average cost of all types by $\overline{z} = \sum_j \phi^j C^j$. I refer to above average cost types i, for which $C^i > \overline{z}$, as *unfit.* I assume that the resistance to arrival for these unfit types satisfies $r_p(C^i, z) \geq 1$. For a *least cost type* i, with $C^i \leq \min_j C^j$, the resistance to arrival $r_p^i(n_t)$ is assumed to be zero. That is, there is a positive probability that an arrival occurs to some type.

It is worth pointing out that the arrival rate depends not only on the resistance, but also the multiplier $H_p^i(n_t)$. This allows the possibility that the overall arrival rate is greater in larger populations, and in particular that it is proportional to the size of the population.

A simple example of an arrival resistance function is the noisy *best response dynamic.* Here the least cost types have zero resistance to arrival, and all other types have resistance one to arrival. The interpretation is that a new group member chooses a type that promises the highest utility with high probability, and only with probability ϵ chooses an inferior type.

Note that ϵ determines both the probabilities of catastrophes and arrivals: that is to say, in the limit I will study both probabilities will be small. However, no assumption has been made about the relative rates at which these probabilities go to zero: these depend on the magnitude of the resistance to catastrophes, about which no assumption has been made.

Stochastic stability

I call the state in which $n_t^i = \overline{N} - M + m^i$, that is, the population is at a maximum, and the variable population consists entirely of type i, the *i-state*. If type i is conformist, the i state is a *conformist state*. I say that a type i is a *resilience providing type,* or *resilient type* for short, if it is conformist and for any other conformist j we have $w^i \geq w^j$. That is, a resilient type maximizes resilience among conformist types. Since there is a type (the nasty type) that is conformist, a resilient type exists. If i is a resilient type, then the i-state is a *resilient state*. The goal of this paper is to characterize stochastically stable states. To minimize notation and maximize readability I will abbreviate "there exists an N such that for $\overline{N} > N$" as "if $\overline{N}$ is sufficiently large."

Proposition 2.1. *For given $\underline{N}$ and M, if $\overline{N}$ is sufficiently large, then for all $\epsilon > 0$ the Markov process defined by the transition probabilities in equation 2.1 has a unique ergodic distribution μ_ϵ that converges to a unique limit μ_0 as $\epsilon \to 0$. The limit distribution μ_0 is an ergodic distribution of the Markov process with $\epsilon = 0$ and the ergodic classes of that process are conformist i-states.*

Following standard practice, those conformist i-states that have positive probability in the limit distribution μ_0 are called *stochastically stable*. The implication of stochastic stability is that, when ϵ is small, "most of the time" we will observe one of these stochastically stable states. In 5, I illustrate this by Monte Carlo simulation.

Proof. For $\epsilon > 0$ the Markov process defined by the transition probabilities in equation 2.1 has a positive probability path from any state to any other state, and because the resistance to no arrival is zero there are no deterministic cycles. From Young (1993) this implies μ_ϵ is unique and has a unique limit μ_0 and that limit is an ergodic distribution for $\epsilon = 0$. Lemma 4.2 below shows that when $\epsilon = 0$ states that are not conformist i-states are transient, and Corollary 4.4 shows that the conformist i-states are absorbing. Hence the ergodic classes for $\epsilon = 0$ are individual conformist i-states and not, for example, best response cycles. □

3. Conformism and Resilience

Recall that $c^i = \alpha w^i + \psi \max\{0, P^i - \underline{P}\}$ where ψ measures the cost of punishment and that v is the importance of resilience in reducing catastrophes. The following are the main results of the paper:

Theorem 3.1. *For given* $\underline{N}$*:*

(i) If $\overline{N}$ *is sufficiently large, only* ***conformist states*** *are stochastically stable*

(ii) There exists a $\hat{v}$ *such that for any* $v > \hat{v}$ *if* $\overline{N}$ *is sufficiently large, only* ***resilient states*** *are stochastically stable.*

All of these results assume that $\overline{N}$ is large relative to $\underline{N}$. The first result (already implied by Proposition 2.1) is general, and says that only conformist states are stochastically stable. But which conformist states? The second result says that if public goods are highly effective in reducing catastrophes (v large), then only resilient states are stochastically stable.

Below in Proposition 6.1 is a third result showing that, when the baseline resistance to catastrophes is large, then only nasty states are stochastically stable. As explained there, this result requires additional assumptions.

Efficient production and punishment size

Can it be the case that very large punishments, by imposing a much higher cost on opponents than on the type, might be "more stable" than smaller punishments? That is, is it true that evolution favors large punishments? The next result shows that this is not the case for resilient types. This is a duality result: under conformity, to maximize resilience, punishment must be minimized for a given cost.

Proposition 3.2. *Suppose that $P^i > \underline{P}$ and that there is a conformist j with $P^j < P^i$ and $c^j \geq c^i$. Then i is not a resilient type.*

Proof. If $c^i \geq P^i$, then i is not a conformist type, so not a resilient type either. So assume $c^i < P^i$. Since $P^i > \underline{P}$ it follows that $\max\{0, P^i - \underline{P}\} = P^i - \underline{P}$. Since $w^\gamma = c^\gamma/\alpha - (\psi/\alpha)\max\{0, P^\gamma - \underline{P}\}$ for any type γ, it follows that $w^j > w^i$, implying that i is not a resilient type. □

4. Proof of the Main Result

I first give an overview of the proof. As indicated, for $\epsilon = 0$ conformist states, by discouraging deviations, are absorbing, while all other states are transient. Hence, ergodic distributions place weight only on conformist states, so only these can be stochastically stable. These facts are in Lemmas 4.1 and 4.2. Lemma 4.3 is a technical result showing that conformity constraints can be perturbed by a fixed amount while remaining valid.

If v is sufficiently large, then the public good is important for preventing catastrophes. This is Lemma 4.6. If $\overline{N}$ is sufficiently large, then, in addition, it is difficult to move between conformist states when $N_t = \overline{N}$, because it requires a great deal of arrival to unfit types. This is Corollary 4.4. The key idea is that once the population has fallen, only few arrivals to unfit types are enough to tilt the system to a new state. This is Corollary 4.5. Since it is relatively easy to move between conformist states once the population has fallen (Lemma 4.7), stochastic stability requires a high level of resilience (Lemma 4.8).

To analyze stochastic stability using standard methods, it is necessary to assess the resistance in moving from one state to another. The *resistance of the transition* from n_t to n_{t+1} is the maximum of the resistances in equation 2.1. For any sequence of transitions the *resistance of the path* is the sum of the resistances of the transitions. The *resistance from one state to another* is the least resistance over any path starting in the first state and ending in the other, and such a path is called a *least resistance path.*

Lemma 4.1. *If type i has least cost C^i, then there is a zero resistance transition to a state with an arrival to type i that still has least cost. In particular there is a zero resistance path to the i-state.*

Proof. The least cost type always has zero resistance to arrival. Hence what must be shown is that after such an arrival, type i still has least cost. The cost difference between i and $j \neq i$ is

$$C^i(\phi_t) - C^j(\phi_t) = c^i - c^j + \sum_{\omega \neq i} \phi_t^\omega P^\omega - \sum_{\omega \neq j} \phi_t^\omega P^\omega$$

$$= c^i - c^j + \phi_t^j P^j - \phi_t^i P^i. \tag{4.1}$$

An arrival to i weakly increases ϕ^i and weakly decreases ϕ^j, regardless of whether or not some type $\gamma \neq i$ is removed from the population, and whether or not $\gamma = j$. It follows that for all $j \neq i$ the cost difference $C^i(\phi_t) - C^j(\phi_t)$ is weakly decreased, so i remains least cost. □

Lemma 4.2. *For a given M, if $\overline{N}$ is sufficiently large, and if i is not conformist, then, in the i-state, there is zero resistance to reaching a conformist state.*

Proof. By Lemma 4.1 it suffices to show that if i is not conformist, and $\overline{N}$ is sufficiently large, then there is a conformist γ with least cost in the i-state.

For type i, cost is $C^i(\phi_t) = c^i + \sum_{j \neq i}(m^j/N_t)P^j$. As i is not conformist $c^i \geq P^i$. Let γ be chosen as a minimizer of $\sum_{j \neq \gamma} m^j P^j$ over $\gamma \neq i$ subject to $c^\gamma \equiv 0$, and notice that it must be that $P^\gamma > 0$ so that γ is conformist.

For any $\omega \neq i$ we have

$$C^{\omega}(\phi_t) = c^{\omega} + (1/\overline{N}) \sum_{j \neq \omega} m^j P^j + (1 - M/\overline{N}) P^i$$

$$= c^{\omega} + (1 - M/\overline{N}) P^i + (1/\overline{N}) \sum_{j \neq \omega} m^j P^j.$$

If $c^{\omega} = 0$, then by construction $C^{\gamma}(\phi_t) \leq C^{\omega}(\phi_t)$. If $c^{\omega} > 0$ and since $c^{\gamma} = 0$, we have $c^{\gamma} + P^i < c^{\omega} + P^i$. Hence, for $\overline{N}$ sufficiently large, $C^{\gamma}(\phi_t) < C^{\omega}(\phi_t)$.

Finally, compare $C^{\gamma}(\phi_t)$ with $C^i(\phi_t)$. Since i is not conformist $c^i \geq P^i$, so

$$C^{\gamma}(\phi_t) - C^i(\phi_t) = -c^i + (1 - M/\overline{N} + m^i/\overline{N}) P^i - (m^{\gamma}/\overline{N}) P^{\gamma} < 0.$$

□

Lemma 4.3. *There exists $\lambda > 0$ and $\theta < 1$ such that if i is a conformist type, then for $\phi_t^i > \theta$ and $j \neq i$, we have $C^i(\phi_t) + \lambda < C^j(\phi_t)$.*

Proof. Since there are finitely many types, it suffices to find a $\lambda > 0$ and $\theta < 1$ for each pair consisting of a conformist i and $j \neq i$. By the definition of conformity $c^i < c^j + P^i$. In other words, for $\phi_t^i = 1$ it is the case that $c^i - c^j + (1 - \phi_t^i) P^j - \phi_t^i P^i < 0$. By continuity, there is a $\lambda > 0, \theta < 1$ pair such that if $\phi_t^i > \theta$, then $c^i - c^j + (1 - \phi_t^i) P^j - \phi_t^i P^i + \lambda < 0$. Using equation 4.1, the result then follows from

$$C^i(\phi_t) - C^j(\phi_t) + \lambda = c^i - c^j + \phi_t^j P^j - \phi_t^i P^i + \lambda$$

$$\leq c^i - c^j + (1 - \phi_t^i) P^j - \phi_t^i P^i + \lambda.$$

□

Corollary 4.4. *For given $\underline{N}$, if $\overline{N}$ is sufficiently large, there is an $a > 0$ so that in a conformist i-state to reach another conformist state without a catastrophe has a resistance at least $a\overline{N}$. This means that for $\epsilon = 0$, conformist states are absorbing.*

By Lemma 4.2, all other states have zero resistance paths to one of these absorbing states so they are transient. Hence, with $\epsilon = 0$, only conformist states have positive weight in any ergodic distribution, so only they can be stochastically stable. This is part (i) of the Theorem.

Proof. Let i be conformist, and choose λ and θ by Lemma 4.3, so that for $\phi_t^i > \theta$ and $j \neq i$ we have $C^i(\phi_t) + \lambda < C^j(\phi_t)$. Let $\overline{C}^i = \max_{\varphi:\varphi^i \geq \theta} C^i(\varphi)$ and $\overline{C} = \max c^j + \max P^j$. Recall that the average cost of all types is $\overline{z} = \sum_j \phi^j C^j$. Then, for $\phi_t^i > \theta$, the average cost $\overline{z}$ is at most $(1-\phi_t^i)\overline{C} + \phi_t^i \overline{C}^i$, while

$$C^j(\phi_t) > \overline{C}^i + \lambda \geq \frac{\overline{z} - (1-\phi_t^i)\overline{C}}{\phi_t^i} + \lambda$$

$$= \overline{z} - \frac{1-\phi_t^i}{\phi_t^i}\left(\overline{C} - \overline{z}\right) + \lambda.$$

Hence, for

$$\phi_t^i \geq \frac{\overline{C} - \overline{z}}{\overline{C} - \overline{z} + \lambda/2} \equiv \overline{\theta},$$

$C^j(\phi_t) > \overline{z} + \lambda/2$, that is $j \neq i$ has above average cost. Taking $\underline{\theta} = \frac{M - m^i}{\overline{N}}$, we may choose $a = 1 - \max\{\theta + \underline{\theta}, \overline{\theta} + \underline{\theta}\}$, so that at least $a\overline{N}$ arrivals of cost at least one are needed to escape from an i-state. □

Define $\overline{r} = \max_{i \in I} \max_n r_p^i(n)$.

Corollary 4.5. *For any conformist type i, after a catastrophe, the resistance to reach the conformist i-state is at most $\hat{R} \equiv \overline{r}(1 + \underline{N}/(1-\theta))$.*

Proof. To see this, add $K \geq \underline{N}/(1-\theta)$ of conformist type i to the post-catastrophe population of $\underline{N}$, which raises the proportion of type i in the group to $\phi_t^i > \theta$. By definition, each addition has resistance no greater than $\overline{r}$. By Lemma 4.3, in this new state, i has cost at least λ less than any other type. Hence, the resistance to reach the i-state after a catastrophe is at most $\hat{R}$. □

Let $\overline{w}$ denote the resilience provided by a resilient type, and by $\overline{W}$ the average resilience in a resilient state.

Lemma 4.6. *There is a $\hat{v}$ large enough that, for $v > \hat{v}$, the resistance of a resilient state to a catastrophe is at least $\hat{R}$ greater than a non-resilient state.*

Proof. Since r_κ is strictly increasing and has slope bounded uniformly above zero in the first argument, it follows that $\overline{W} > w^i$ for all $w^i < \overline{w}$, thus there exists a $\overline{v}$ large enough so that for $v > \overline{v}$, we have $r_\kappa(v\overline{W}, N_t/\overline{N}) - r_\kappa(vw^i, N_t/\overline{N}) > \hat{R}$. □

Lemma 4.7. *For fixed v, there is an $\overline{N}$ large enough such that the least resistance route from one conformist i-state to another is by having an immediate catastrophe.*

Proof. The greatest resistance if there is an immediate catastrophe is $r_\kappa(v\overline{w}, 1) + \hat{R}$, while, if there is no catastrophe at all, it is at least $a\overline{N}$. Hence for $\overline{N} > (r_\kappa(v\overline{w}, 1) + \hat{R})/a$ the least resistance paths between conformist states must have a catastrophe.

Finally, I show that if $\overline{N}$ is large enough, the catastrophe must be immediate. That is, it is not possible to reduce resistance below $r_\kappa(vw^i, N_t/\overline{N})$ by increasing a type with low resilience and then having a catastrophe. As I just showed that we cannot leave the basin of the conformist state i without greater resistance than an immediate catastrophe, it suffices to show that an above average cost arrival with cost at least one reduces the resistance to a catastrophe by less than one. Consider then a given W and an above average cost of arrival leading to $\hat{W}$. Continue to let $\overline{w} = \max_j w^j$. As resilience is an average, $|W - \hat{W}| \leq \overline{w}/\overline{N}$. Recall that r_κ has slope bounded away from zero in the first argument, say by $b > 0$. Hence

$$|r_\kappa(vW, 1) - r_\kappa(v\hat{W}, 1)| \leq vb\overline{w}/\overline{N}$$

and we see that, if $\overline{N}$ is sufficiently large, this is less than one. □

Lemma 4.8. *For given $\underline{N}$ there exists a $\hat{v}$ and for any $v > \hat{v}$, if $\overline{N}$ is sufficiently large, only resilient states are stochastically stable.*

This is part (ii) of the Theorem.

Proof. Fix $\underline{N}$ and choose $\hat{v}$ from Lemma 4.6. Choose $v > \hat{v}$ and define sufficiently large $\overline{N}$ by Lemma 4.7.

Young (1993) characterizes stochastically stable states by analyzing trees having as nodes the conformist states. Each branch on a tree of all conformist states has a resistance equal to the least resistance path from the node on the branch furthest from the root to the node closest to the root, and the resistance of the tree is equal to the sum of resistances of all branches. Stochastically stable states then correspond to the root nodes of least resistance trees.

Suppose i is at the root of a least resistance tree and is not a resilient state. Find some resilient state j in the tree and cut it out from the state γ to which it was connected. By Lemma 4.7 this saves a resistance of at least $r_\kappa(v\overline{W}, 1)$. Attach the previous root i to j making j the root. Also by Lemma 4.7 and Corollary 4.5 this adds a resistance of at most $r_\kappa(vw^i, 1) + \hat{R}$. Hence, resistance is decreased by

$$r_\kappa(v\overline{W}, 1) - r_\kappa(vw^i, 1) - \hat{R},$$

which, by Lemma 4.6, is strictly positive. Consequently no state that is not a resilient state can be at the root of a least cost tree, so is not stochastically stable. □

5. Simulations

The reliance of the theoretical results on small ϵ and large $\overline{N}$ relative to $\underline{N}$ may raise concerns about the range of parameter values for which the dynamics described by Theorem 3.1 are verified. In fact, the evolutionary processes are numerically quite robust: this is supported by a recent theoretical literature, including Kreindler and Young (2014) and Ellison, Fudenberg and Imhof (2016). I illustrate this through a Monte Carlo simulation that also highlights the main finding.

Recall that $c^i = \alpha w^i + \psi \max\{0, P^i - \underline{P}\}$ where ψ measures the cost of punishment and that v is the importance of resilience in reducing catastrophes. I take

α	$\underline{P}$	ψ
0.10	1.00	0.33

I will work with four benchmark types: selfish types, nasty types, altruistic types and resilient types given by

i	w^i	P^i	c^i
selfish	0	0	0
nasty	0	$1 = \underline{P}$	0
altruistic	10	0	$1 = \alpha w^i$
resilient	10	4	$2 = \alpha w^i + \psi(P^i - \underline{P})$

The residual population has one of each type, the maximum population is $\overline{N} = 40$, the minimum population is $\underline{N} = 6$ and $\epsilon = 0.5$. Notice that the population decrease from the largest sustainable population in case of a catastrophe is 85%, which is large but not extreme, and $\epsilon = 0.5$, which is hardly negligible. The numbers bear a reasonable relationship to human history. From Bowles and Choi (2013) we know that most human evolution took place in relatively small groups, on the order of $\overline{N} = 40$. With typical ages on the same order, the replacement rate is about one per year, so that the relevant length of a period is one year, and $40,000$ periods is roughly half the history of behaviorally modern homo sapiens.

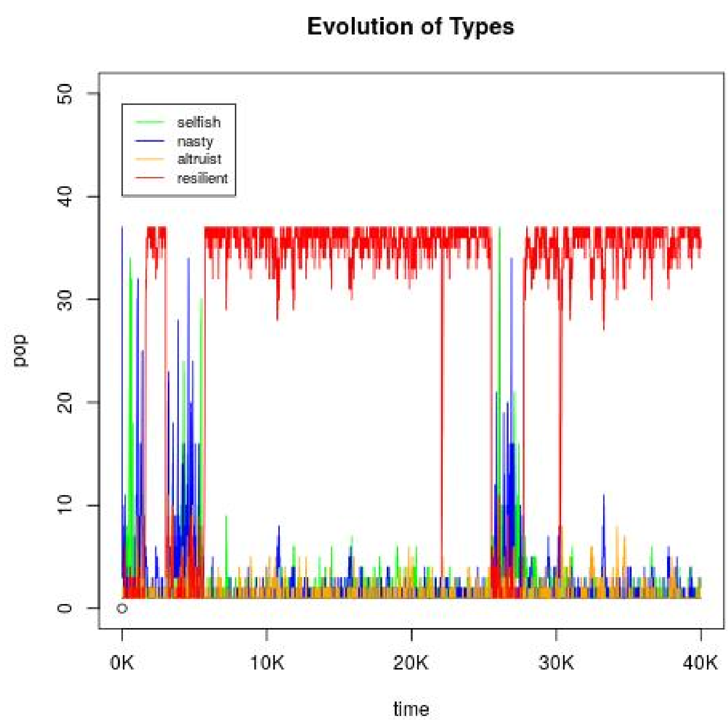
The probability of catastrophe is given by

$$h_\kappa(vW, N_t/\overline{N}) = \epsilon^{3.2+W}$$

and in particular $h_\kappa(0, N_t/N) \approx 1/10$ and $h_\kappa(10, N_t/N) \approx 1/1000$, which is to say, for the types that produce no resilience, catastrophes are roughly once every ten years, while, for the types that do produce resilience, it is about once every thousand years.

The probability that arrival to some type takes place in the absence of a catastrophe is $0.9N_t/\overline{N}$. The Darwinian dynamic is given by constant resistances of 1 for unfit types, and 0 for those with cost weakly below average. The multipliers are $H_p^i = \phi^i \cdot 0.9N_t/\overline{N}$, that is, proportional to the fraction of the type and the arrival probability. The initial population $N_1 = 40$ and the initial variable population consists entirely of nasty types. The Monte Carlo was conducted in R for 40,000 periods.

The results of the Monte Carlo are shown in the graph below.



As can be seen, the types that are not conformist, the selfish and the altruistic types, play little role. Despite the fact that the simulations are slanted in favor of the nasty types by starting in the nasty state, nevertheless the resilient type predominates. In 82% of the periods they constitute more than 75% of the population. To verify this, I took the seed used in the figure and repeated the simulation 100 times, incrementing the seed by one each time. The average number of periods during which the resilient type constituted more than 75% of the population over these 100 simulations was slightly above 50%. However, this understates the importance of the resilient type. When there are many resilient types, the population tends to be large, as they have few catastrophes: this can be seen in the figure. A better measure of evolutionary success is the fraction of resilient types in the cumulative population. This is considerably higher, at 69%.

6. Survival of the Nasty

When catastrophes are not as important, there are conditions under which nasty types dominate.

Proposition 6.1. *For given* $\underline{N}$ *if* $\psi > 1/2$*, then in the best response dynamic if* $\overline{N}$ *is sufficiently large, there is a* $\overline{R} > 0$ *such that for* $\inf_{\varphi} r_{\kappa}(0, \varphi) > \overline{R}$*, only* ***nasty states*** *are stochastically stable.*

To see why an assumption on ψ is needed, suppose that ψ is very small and that there is a conformist super-punishing type that punishes vastly more than any other type. As this super punishment has very little cost, once a decent fraction of super punishing types enter the population, only this super punishing type has below average cost, and, if catastrophes do not play an important role, only the super punishing state is stochastically stable. As it is not clear why one particular type should punish vastly more than any other type, in Theorem 3.1 this possibility is ruled out by assuming that $\psi > 1/2$. This means that types that punish very heavily also bear very high costs, so there can be no super punishing types.

However, even ruling out super punishing types and reducing the chances of a catastrophe is not enough to lead to stochastic dominance by the nasty type. Without catastrophes, we need to deal carefully with the size of the basins of the absorbing states, and this is complicated by the possibility of a mixture of different types arriving. This is an ubiquitous problem in the literature on evolution in repeated games, as can be seen from the analysis of Johnson, Levine and Pesendorfer (2001), and is often avoided by limiting the number of types, or introducing special assumptions such as supposing that evolution largely proceeds through imitation, as in Levine and Pesendorer (2007).

To illustrate the problem, suppose some fairly high cost type enters in a decent fraction. This can raise the average cost so that some third type's cost is now below average, and thus might not have resistance to arrival. However, this cannot occur in the best-response dynamic in which only least cost types have zero resistance to arrival. With this assumption, we can now prove Proposition 6.1.

Proof. The condition that the punishment cost is high $\psi > 1/2$ guarantees that if a nasty type is 50% or more of the population, then it does strictly better than any other type. So the same remains true for some fraction $\theta < 1/2$. By making $\overline{N}$ large enough, we ensure this remains true after accounting for the residual types, and that the actual variable population of nasty types needed is strictly less than half the population. If we then take $\overline{R}$ large enough that $\overline{r} < \overline{R}$, then least resistance paths cannot include a catastrophe, so we are down to the standard case of analyzing least resistance paths for a fixed population of $\overline{N}$. However, the fact that it takes more than half the population to have arrivals with strictly positive resistance to escape a nasty state, while it takes a nasty type to have strictly fewer arrivals with resistance strictly greater than zero to go from any other conformist i-state to the nasty state, leads to the standard result, for example, in Morris, Rob and Shin (1995) or Ellison (2000), that only the nasty state is stochastically stable. This is easily proven by taking any tree in which the nasty state is not the root, cutting it and attaching the root to it, and observing that this strictly reduces the resistance. □

Simulation

To illustrate Proposition 6.1, I modify the first Monte Carlo example by reducing the importance of catastrophes, using

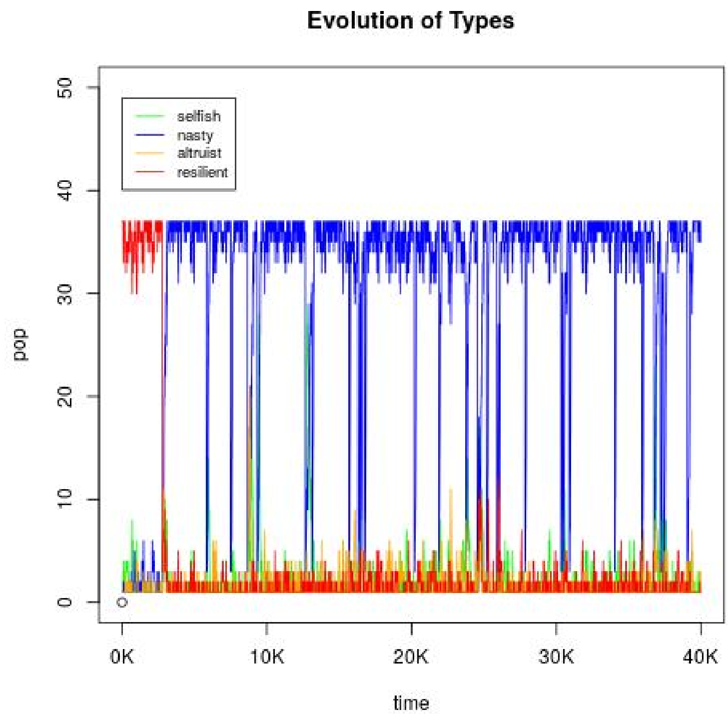
$$h_\kappa(vW, N_t/\overline{N}) = \epsilon^{8.1+W}$$

in place of

$$h_\kappa(vW, N_t/\overline{N}) = \epsilon^{3.2+W}.$$

I also switch to an initial population of resilient types.

The result is plotted in the graph below, in which we see that the nasty type dominates.



7. Group Selection and the Identification of Types

Crucial to the theory is the assumption that types are observable, although possibly with error. As indicated in the literature review, this has been a fraught subject in the evolutionary literature on group selection. Robson (1990)'s discussion of the secret handshake, Levine and Pesendorer (2007)'s assumption that lying is difficult, Levine and Szentes (2006) discussion of the feasibility of identifying those using the same rules, and the difficulty in observing off path play in the repeated game setting, are all examples of obstacles to identifying types.

In the present setting, if costs, public good production, and the act of punishment are observable, and members can communicate, the issue is not problematic. Individuals can be required to identify themselves as a type by making a public announcement, that is, stating a type i. It is the duty of each individual of a particular type to incur the prescribed cost and produce the prescribed amount of public good, and to punish anyone who announces a different type or fails to carry out the duties of their own type. If costs, public good production, the act of punishment, and the announcements are perfectly observed, then so are types.

In practice observability need not be so complete, so it is worth emphasizing that the model is consistent with the imperfect observation of types. Specifically, suppose that there is a chance of "accidentally" punishing their own type, as in Levine and Modica (2016b) or Levine and Mattozzi (2020). I maintain the assumption that small punishments have no cost, and assume in addition that they do not hurt the population that issues them. For example, they might be an insult that is offensive to other types but not to members who are type i. I now assume that while P^i represents the expected punishment to a member not of type i, there is also an expected accidental punishment of πP^i to type i members as well.

Type i then issues $(1-\phi^i)+\phi^i\pi$ punishments costing $\psi\max\{0, P^i - \underline{P}\}$. It also receives $\phi^i\pi$ of $\max\{0, P^i - \underline{P}\}$ "by accident" due to inaccurate signals. Hence the cost function

$$c(w^i, P^i)[\phi^i] = \alpha w + \phi^i\pi(1+\psi)\max\{0, P^i - \underline{P}\}) + (1-\phi^i)\psi\max\{0, P^i - \underline{P}\}. \tag{7.1}$$

now depends upon ϕ^i as well as w^i and P^i. Note, however, that if $\pi = \psi/(1+\psi)$, this becomes $c(w^i, P^i)[\phi^i] = \alpha w + \psi\max\{0, P^i - \underline{P}\}$ as assumed previously. Generalizing to allow cost to depend on ϕ^i does not change the main results in Theorem 3.1, as shown in the earlier working paper version Levine (2024).

8. Conclusion

The model in this paper shows how resilient types that provide incentives for the provision of public goods will emerge from a process of social evolution, provided that those public goods provide resilience by protecting against catastrophes. The model also suggests that non-excludable public goods are generally produced only insofar as they contribute to resilience. Some public goods - most notably defense and environmental spending - clearly contribute to resilience. However, modern economies spend far more on other programs, primarily social insurance, than they do on either defense or environment. It is less clear that these other expenditures are non-excludable, but to the extent that they are, the model does not provide a complete explanation of the public goods we see.

References

Akerlof, G. and R. E. Kranton (2005): "Identity and the Economics of Organizations," *Journal of Economic Perspectives.*

Alger, I., and Weibull, J. W. (2013): "Homo moralis—preference evolution under incomplete information and assortative matching," *Econometrica* 81: 2269-2302.

Axelrod, R. and W. D. Hamilton (1981): "The evolution of cooperation," *Science.*

Bailey, Mark (2014): *The decline of serfdom in late medieval England: from bondage to freedom*, Boydell & Brewer Ltd4.

Benassy, J. P. (1998): "Conformism and multiple sycophantic equilibria," in P. Howitt and A. Leijonhufvud (eds), *Money, Markets and Method*, Edward Elgar.

Bilancini, Ennio, Leonardo Boncinelli, and Pablo Marcos-Prieto (2022): "Conflict Initiation Function Shapes the Evolution of Persistent Outcomes in Group Conflict" mimeo IMT Lucca

Binmore, Kenneth G., and Larry Samuelson (1992): "Evolutionary stability in repeated games played by finite automata," *Journal of Economic Theory* 57: 278-305.

Bowles, Samuel and Jung-Kyoo Choi (2013): "Coevolution of farming and private property during the early Holocene" *Proceedings of the National Academy of Science*, doi:10.1073/pnas.1212149110.

Boyd, Robert and Peter J. Richerson (1982): "Cultural transmission and the evolution of cooperative behavior," *Human Ecology*: 325–351.

Boyd, Robert and Peter J. Richerson (1988): *Culture and the evolutionary process*, University of Chicago press.

Cui, Shiwei and Jian Zhai (2010): "Escape dynamics and equilibria selection by iterative cycle decomposition," *Journal of Mathematical Economics*, 46:1015 – 1029.

Dal Bó and Pujals (2015): "The Evolutionary Robustness of Forgiveness and Cooperation," mimeo.

Cho, In-Koo and Kenneth Kasa (2014): "Learning and model validation," *Review of Economic Studies* 82 :45–82.

Choi, Jung-Kyoo and Samuel Bowles (2007): "The coevolution of parochial altruism and war," *Science* 318: 636-40.

Coase, R. H. (1960): "The Problem of Social Cost," *Journal of Law and Economics* 3: 1-44.

Diamond, Jared (2005): *Collapse: How Societies Choose to Fail or Succeed*, Viking Press.

Dutta, R., D. K. Levine and S. Modica (2022): "The Whip and the Bible: Punishment Versus Internalization," *Journal of Public Economic Theory*, 23: 858-894

Ellison, Glenn (2000): "Basins of Attraction, Long Run Stochastic Stability and the Speed of Step-by-step Evolution," *Review of Economic Studies* 67: 17-45.

Ellison, G., D. Fudenberg and L. A. Imhof (2016): "Fast convergence in evolutionary models: A Lyapunov approach," *Journal of Economic Theory* 161: 1-36.

Ely, Jeffrey C. (2002): "Local conventions," *Advances in Theoretical Economics* 2.

Gardner, Andy and Stuart A. West (2010): "Greenbeards," *Evolution* 64:25–38.

Juang, W-T. and Sabourian, H. (2021): "Rules and Mutation - A Theory of How Efficiency and Rawlsian Egalitarianism/Symmetry May Emerge," mimeo Cambridge.

Jablonski, David (2001): "Lessons from the past: Evolutionary impacts of mass extinctions," *Proceedings of the National Academy of Sciences* 98: 5393-5398.

Johnson, P., D. K. Levine and W. Pesendorfer (2001): "Evolution and Information in a Gift Giving Game," *Journal of Economic Theory* 100: 1-22.

Hamilton, William D. (1964a): "The genetical evolution of social behaviour. I," *Journal of Theoretical Biology* 7: 1–16.

Hamilton, William D. (1964b): "The genetical evolution of social behaviour. II," *Journal of Theoretical Biology* 7:17–52.

Heller, Yuval and Ilan Nehama (2023): "Evolutionary foundation for heterogeneity in risk aversion," *Journal of Economic Theory* 208: 105617, 2023.

Jansen, Vincent A. A. and Minus Van Baalen (2006): "Altruism through beard chromodynamics," *Nature* 440:663–666.

Kandori, Michihiro, George Mailath, and Rafael Rob (1993): "Learning, Mutation, and Long Run Equilibria in Games," *Econometrica* 61: 29-56.

Kreindler, G. E. and H. P. Young (2014): "Rapid innovation diffusion in social networks," *Proceedings of the National Academy of Sciences* 111: 10881-10888.

Twine, K. (1992): "The city in decline: Rome in late antiquity," *Middle States Geographer* 25: 134-138.

Levine, David K. (2024): "Phoenix From the Ashes: Resilience and the Evolution of Mechanism Designers," working paper, www.dklevine.com.

Levine, David and Salvatore Modica (2012): "Conflict and the Evolution of Societies," working paper, www.dklevine.com

Levine, David and Salvatore Modica (2013): "Anti-Malthus: Conflict and the Evolution of Societies," *Research in Economics* 67: 289-306

Levine, David and Salvatore Modica (2016a): "Dynamics in stochastic evolutionary models", *Theoretical Economics* 11: 89-131

Levine, David and Salvatore Modica (2016b): "Peer Discipline and Incentives within Groups," *Journal of Economic Behavior and Organization* 123: 19-30

Levine, D. K. and S. Modica (2022): "Survival of the Weakest: Why the West Rules," *Journal of Behavior and Economic Organization*, forthcoming.

Levine, David K., and Wolfgang Pesendorfer (2007): "The evolution of cooperation through imitation," *Games and Economic Behavior* 58: 293-315.

Levine, D. K. and B. Szentes (2006): "Can A Turing Player Identify Itself?," *Economics Bulletin* 1: 1-6.

Levine, David K. and Andrea Mattozzi (2020): "Voter Turnout with Peer Punishment," *American Economic Review* 110: 3298–3314.

Marean, C.W., M. Bar-Matthews, J. Bernatchez, E. Fisher, P. Goldberg, A. I. Herries, Z. Jacobs, A. Jerardino, P. Karkanas and P. J. Nilssen (2007), "Early human use of marine resources and pigment in South Africa during the Middle Pleistocene," *Nature* 449: 905-908.

Molleman, Lucas, Andrés E. Quiñones, and Franz J. Weissing (2013): "Cultural evolution of cooperation: the interplay between forms of social learning and group selection," *Evolution and Human Behavior* 34: 342–349.

Morris, S., R. Rob and S. Shin (1995): "p-Dominance and belief potential," *Econometrica*: 145-157.

Newton, J. (2021): "Conventions under heterogeneous behavioural rules," *Review of Economic Studies* 88: 2094-2118.

Ostrom, Elinor (1990): *Governing the commons: The evolution of institutions for collective action*, Cambridge university press.

Peski, M. (2010): "Generalized risk-dominance and asymmetric dynamics," *Journal of Economic Theory*, 145: 216 – 248.

Peteet, D. (1995): "Global younger dryas?" *Quaternary International* 28: 93-104.

Robson, A. J. (1990): "Efficiency in evolutionary games: Darwin, Nash and the secret handshake.," *Journal of theoretical Biology*, 144: 379-396.

Robson, A. J. (1992): "Status, the distribution of wealth, private and social attitudes to risk," *Econometrica*: 837-857.

Ross, David (2002): *Ireland: History of a Nation*, New Lanark: Geddes & Grosset.

Rusch, Hannes (2014): "The evolutionary interplay of intergroup conflict and altruism in humans: a review of parochial altruism theory and prospects for its extension," *Proceedings of the Royal Society B: Biological Sciences* 281.

Serrano, Roberto and Oscar Volij (2008): "Mistakes in cooperation: the stochastic stability of Edgeworth's recontracting," *Economic Journal* 118: 1719–1741.

Skarbek, D. (2014): "The social order of the underworld: How prison gangs govern the American penal system," O*xford University Press.*

Szekely, Aron, Francesca Lipari, Alberto Antonioni, Mario Paolucci, Angel Sánchez, Luca Tummolini, and Giulia Andrighetto (2021): "Evidence from a long-term experiment that collective risks change social norms and promote cooperation," *Nature Communications* 12: 5452.

Townsend, R. M. (1994): "Risk and insurance in village India," *Econometrica*, 539-591.

Winter, S. G. (1971): "Satisficing, selection, and the innovating remnant," *Quarterly Journal of Economics* 85: 237-261.

Young, P. (1993): "The Evolution of Conventions," *Econometrica* 61: 57-83