

Uninvadable social behaviors and preferences
in group-structured populations

Ingela Alger, Laurent Lehmann and Jörgen W. Weibull



Uninvadable social behaviors and preferences in group-structured populations

Ingela Alger^{*}, Laurent Lehmann[†], and Jörgen W. Weibull[‡]

February 1, 2018[§]

Abstract

Humans have evolved in populations structured in groups that extended beyond the nuclear family. Individuals interacted with each other within these groups and there was limited migration and sometimes conflicts between these groups. Suppose that during this evolution, individuals transmitted their behaviors or preferences to their (genetic or cultural) offspring, and that material outcomes resulting from the interaction determined which parents were more successful than others in producing (genetic or cultural) offspring. Should one then expect pure material self-interest to prevail? Some degree of altruism, spite, inequity aversion or morality? By building on established models in population biology we analyze the role that different aspects of population structure—such as group size, migration rates, probability of group conflicts, cultural loyalty towards parents—play in shaping behaviors and preferences which, once established, cannot be displaced by any other preference. In particular, we establish that uninvadable preferences under limited migration between groups will consist of a materially self-interested, a moral, and an other-regarding component, and we show how the strength of each component depends on population structure.

Keywords: strategic interactions, preference evolution, evolution by natural selection, cultural transmission, pro-sociality, altruism, morality, spite.

JEL codes: A12, A13, B52, C73, D01, D63, D64, D91.

^{*}Toulouse School of Economics, CNRS, University of Toulouse Capitole, Toulouse, France, and Institute for Advanced Study in Toulouse. ingela.alger@tse-fr.eu

[†]Department of Ecology and Evolution, University of Lausanne, Switzerland. laurent.lehmann@unil.ch

[‡]Stockholm School of Economics, Institute for Advanced Study in Toulouse, and Toulouse School of Economics. jorgen.weibull@hhs.se

[§]All authors conceived the model, I.A. and L.L. derived the main results, and I.A. wrote the manuscript with input of all authors. Support by Knut and Alice Wallenberg Research Foundation and by ANR-Labex IAST is gratefully acknowledged. We also thank Agence Nationale de la Recherche for funding (Chaire d'Excellence ANR-12-CHEX-0012-01 for I. Alger, and Chaire IDEX ANR-11-IDEX-0002-02 for J. Weibull).

1 Introduction

The self-interested *Homo oeconomicus*—who only cares about own material payoff, and who once populated most economics models—is at odds with an overwhelming amount of data from laboratory experiments (see, e.g., Zelmer, 2003, Engel, 2011, Güth and Kocher, 2014) and from other sources (see, e.g., Bandiera, Barankay, and Rasul, 2005, Feld and Frey, 2007). Over the past few decades, behavioral economics has suggested richer motivations to account for observed behaviors, such as altruism (Becker, 1976), warm glow (Andreoni, 1990), a concern for fairness (Rabin, 1993), reciprocal altruism (Levine, 1998), inequity aversion (Fehr and Schmidt, 1999), identity concerns (Akerlof and Kranton, 2000), moral motivation (Brekke, Kverndokk, and Nyborg, 2003), self-image concerns (Bénabou and Tirole, 2006), and a concern for honesty (Alger and Renault, 2006).¹ While this literature has helped enhance predictive power in several interactions, further progress could be made if the underlying factors that shape human motivation were better understood. It has long been emphasized that in order to understand human behavior, the evolutionary foundations of economists’ assumption that behavior is driven by rational choice needs to be better understood and ascertained (e.g., Hirshleifer, 1977, Alexander, 1979, Lumsden and Wilson 1981, Bergstrom, 1996, Robson, 2001). This is precisely the task undertaken by the economics literature on preference evolution. The key question asked is: What preferences or moral values should humans be expected to have if preferences and moral values are transmitted across generations, and preferences and values that lead to higher individual fitness spread at the expense of other preferences and values?

Milton Friedman (1953) claimed that “unless the behavior of businessmen in some way or other approximated behavior consistent with the maximization of returns, it seems unlikely that they would remain in business for long”. However, neither Friedman’s claim, nor conjectures along similar lines in biological or cultural evolution, turn out to be generally validated by theoretical models. To see this, suppose that an individual’s material returns or payoffs in interactions with others enhance the individual’s reproductive success (survival and number of offspring). It may then appear plausible that biological evolution of genetically transmitted traits would favor individuals who strive to maximize their own material returns or payoffs. In formal models, it has been established that the following three conditions are necessary—except in knife-edge cases—and sufficient for this conjecture to be valid: *(i)* the population is very large and reproduction is asexual, *(ii)* interacting individuals do not know each other’s preferences or goal functions, and *(iii)* interactions are uniformly random in the

¹Throughout this paper we use concepts and terminology that are standard in economics, and model behavior as a choice of action (or stream of actions) from a set of feasible actions, where this choice is guided by a striving to maximize some goal (utility) function. The utility function together with the information and the constraints imposed by the environment are thus what biologists would call the proximate causes driving behavior. Furthermore, by contrast to the evolutionary biology literature where the terms “altruism” and “spite” are used to refer to the fitness consequences of a behavior on the actor and others, in economics they are used to describe the proximate causes behind behaviors. Thus, in economics, an individual who has a utility function which puts a positive weight on another individual’s material well-being is altruistic; and an individual who has a utility function which puts a negative weight on another individual’s material well-being is spiteful. For further discussion of the meaning of these terms in different academic disciplines, we refer to West et al. (2007) and Bshary and Bergmüller (2008).

sense that each encounter is just as likely, see Ok and Vega-Redondo (2001), Dekel, Ely, and Yilankaya (2007). However, these conditions rarely, if ever, apply, even as approximations. Indeed, all natural populations (human or otherwise) are structured into groups, connected to each other by limited migration. Limited migration causes limited genetic and/or cultural mixing among population members (Cavalli-Sforza and Bodmer, 1971, Hartl and Clark, 2007), which in turn has profound consequences for the evolution of traits in populations.²

Indeed, a key implication of spatial population structure is that it imposes constraints on the ways in which traits, including strategies, preferences and moral values, can spread. For example, consider a culturally transmitted trait and suppose that in an initially homogenous group or location suddenly a new trait appears in one individual. If children acquire their traits by copying their parents or an adult in their neighborhood, in the second generation some interactions may involve several carriers of the new trait. Hence, locally interacting individuals are more likely to share the same trait, even when rare in the population at large, than are individuals sampled at random from the whole population, and they will be culturally “related” through the common adult they copied, their “common cultural ancestor”. Such assortative matching—*relatedness* in the biology literature—between interacting individuals tends to favor behavior that promotes the survival and/or reproductive success of others, since such behavior is more likely to benefit individuals also carrying the trait determining such behavior than when matching is uniformly random (Hamilton, 1964, 1971, Grafen, 1985, Frank, 1998).

However, spatial population structure also means that individuals who share a common “ancestor” are more likely to be competing for the same local resources than are individuals drawn at random from the population at large. For example, two siblings who inherit their traits from their parents are likely to also compete for the resources provided by the parents. If interacting groups are finite, such *local competition* tends to favor behavior that hinders the survival and/or reproductive success of others, since achieving an edge over neighbors then translates into evolutionary success (Hamilton, 1971, Schaffer, 1988, Frank, 1998). As both assortative matching and local competition can, in general, not be separated, their joint effects need to be taken into account to understand how trait expression results in evolutionary success under limited dispersal, a question that has received much attention in the biology literature (e.g., Frank 1998 and Rousset, 2004, for general treatments). Clearly, then, this literature is relevant when it comes to understanding how spatial structure affects trait evolution in social interactions. Nonetheless the value of this previous work is limited for economists by the facts that (i) these analyses focus on the evolution of strategies, and not of preferences; (ii) the predictions are derived at the level of basic fitness components, such as reproduction and survival, but not at the level of the material payoffs generated by the interaction, and (iii) the underlying (genetic) concepts are exotic to most economists

²In industrial organization theory, Fershtman and Judd (1987) showed that if the managers of competing firms know each others’ contracts, and thus (ii) is violated, return-maximizing contracts are not viable. In abstract game theory, Heifetz, Shannon and Spiegel (2007a) likewise showed that if players know each other’s goal functions, then own material payoff is a goal function that is selected against. Alger and Weibull (2013) showed that if the random matching is assortative, and hence (iii) is violated, then again pure material self-interest is selected against.

with results scattered in a vast literature which often assumes low-dimensional strategy sets and focuses on specific games.

In earlier work, Lehmann, Alger, and Weibull (2015), we started to address these limitations by analyzing strategy evolution in group-structured populations under limited migration. We established conditions under which the emerging strategies can be interpreted as chosen by rational individuals endowed with certain preferences, and showed how features of the population structure translate into features of these preferences. Here we build on this work by proposing a framework more in line with those used by economists and analyze how the specifics of the transmission process—migration rate, group size, probability of war, cultural loyalty—determine evolutionarily viable strategies and preferences. Compared to the earlier economics literature on preference evolution, our contribution is to explicitly analyze the effects of spatial and socioeconomic group structure and limited migration between groups upon behavior and preferences; compared to the biology literature, our contribution is to deliver predictions on preferences at the level of material payoffs in the social interaction rather than at the fitness level (these literatures are discussed in greater detail in Section 6).

Specifically, our model allows for detailed analysis of the following thought experiment that takes place in a large population over an infinite sequence of discrete time periods. The population is structured into a large number of groups of equal size. Within each group, individuals engage in a strategic interaction in which all parties' strategies affect the material payoffs to all participants. The material payoffs so realized in turn determine the expected fitness of each individual in the population, where an individual's fitness is the number of individuals in the following time period who have acquired their trait from him or her. If transmission is genetic, an individual's fitness is the number of his surviving offspring and the individual himself if he survives. If transmission is cultural, an individual's fitness is the number of individuals in the next generation who acquire their trait by copying this individual. Offspring may migrate to another group or stay in the natal group. In the case of genetic transmission, groups may wage wars against each other, wars in which the winning group takes over the losing group's resources. In all these scenarios, genetic and cultural, the population is initially homogenous; all individuals are *ex ante* identical, carry the same *trait*. Suddenly, a different, mutant trait, spontaneously appears in exactly one individual. The original trait is *uninvadable* if for every possible mutant trait, the initial mutant does not produce enough descendants for its trait to be maintained in the population in the long run.

We begin with analysis of strategy evolution, that is, when the traits in question are strategies in the group interaction. We consider several standard transmission scenarios, which allow to build intuition about how their specifics affect trait evolution under limited migration. Once this foundation is laid out, we turn to analyzing preference evolution, that is, when the traits in question are utility functions that guide the carriers's choice of strategy in the group interaction.

Under preference evolution, we assume that an individual's preferences are his or her private information. We evaluate a utility function's fitness consequences for its carrier in terms of the material payoffs that result when groups play Bayesian Nash equilibria under incomplete information. Allowing for all continuous utility functions, we ask if there exist

utility functions that are uninvadable in the sense that any mutant utility function does worse, in terms of material payoffs, than the resident function in all equilibria. This analysis allows to establish a link between, on the one hand, the environment in which a population evolves—represented by the material game, the population structure, and the transmission scenario—and, on the other hand, preferences that motivate individuals in their choice of strategy. In particular, it allows us to establish whether such preferences simply correspond to maximization of own material payoff in the group interaction—in line with *Homo oeconomicus*—or whether these preferences also include other-regarding or moral considerations. In the latter case, what exactly are the other-regarding and moral components, and how do they depend on the specifics of the transmission process?

Precise answers to these questions will be provided in this paper, which is organized as follows. Section 2 provides the baseline model and defines three scenarios for how traits are transmitted across time and space. Section 3 defines and provides a characterization of uninvadability. It also presents a result that allows to characterize strategies that are uninvadable when selection is weak, that is, when the effects on vital rates are weak enough to permit first-order approximation. This allows us to define a utility function such that a strategy is uninvadable if and only if it is a symmetric Nash equilibrium among individuals endowed with it. Sections 4-5 establish the novel results. In Section 4, in the special case of a differentiable material payoff function we state a first-order condition that allows to identify the set of candidate uninvadable strategies under weak selection. A second part of Section 4 characterizes this set in the three scenarios. In Section 5 we provide a second model, in which preferences, rather than strategies, are the units of selection and show that the aforementioned utility function is uninvadable. Section 6 describes the related literatures, and Section 7 concludes. All the proofs are in the Appendix.

2 Model

We consider a population structured into a large number of distinct groups, locations, villages or islands. For simplicity, all the islands are assumed to have the same number n of adults living there in each generation. Our analysis will concern the limit case of infinitely many islands. Time is divided into demographic time periods, within each of which the following events unfold. At the beginning of a demographic time period the adults in each island engage in a social or economic interaction with each other, to be called the material game, the same on all islands and at all times. This is a non-cooperative normal-form game in which each player has the same set of (pure or mixed) strategies, X , and which results in material payoffs to all individuals in the island. The material payoff to an individual i who uses strategy $x_i \in X$ when the others on his or her island use strategies $\mathbf{x}_{-i} \in X^{n-1}$ is $\pi(x_i, \mathbf{x}_{-i})$, is assumed to be invariant under permutation of the components of the vector $\mathbf{x}_{-i} \in X^{n-1}$ of neighbors' strategies. An individual's trait determines her strategy choice. In the first part of the analysis, the trait will be the strategy to play, while in the second part it will be a utility (or goal) function that guides individuals' choice of strategy.

Upon play of the material game, but still within the demographic time period, certain

transmission events occur. These events determine, for each individual, the probabilities that her trait will be transmitted to one or more individuals in her island or in another island, at the beginning of the next demographic time period. An individual's *immediate descendants* consist of her surviving offspring, as well as herself if she survives, where "offspring" may be interpreted as cultural. The expected number of immediate descendants will be referred to as the individual's *fitness*. This fitness may depend on the individual's own material payoff, the material payoffs to the other members of her island, and on average material payoff in the population at large.

We next describe three canonical *transmission scenarios* whereby traits are transmitted from one demographic period to the next. We first define a standard scenario in evolutionary biology, then a "war scenario" in which there are occasional conflicts between islands, and finally a scenario of cultural transmission.

2.1 Scenario A: genetic transmission

This *base-line scenario* might occur if traits are determined genetically under asexual reproduction when each offspring inherits the trait of its parent.³ The scenario has four stages after play of the material game. First, each adult produces a Poisson-distributed number of offspring. Then each adult survives or dies, statistically independently of each other. Thereafter, each offspring either aspires to replace a deceased adult in her natal island or migrates to another, randomly chosen island, where she aspires to replace a deceased adult. Individual offsprings' migration decisions are statistically independent, and offspring who migrate disperse uniformly to the other islands, with statistical independence between their destinations.⁴ Finally, in each island the deceased adults are replaced by (uniformly) randomly drawn aspiring offspring, native and immigrant. The fortunate ones settle and become adults while the unfortunate ones die.⁵

Write $w(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*)$ for the fitness of individual i when the material payoffs earned in her island are $(\pi_1, \pi_2, \dots, \pi_n) \in \mathbb{R}^n$ and all others in the population earn the same material payoff $\bar{\pi}^* \in \mathbb{R}$. In the present scenario

$$w(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*) = s(\pi_i) + m \cdot [1 - s(\bar{\pi}^*)] n \cdot \frac{f(\pi_i)}{nf(\bar{\pi}^*)} \quad (1)$$

$$+ (1 - m) \cdot \left(n - \sum_{j=1}^n s(\pi_j) \right) \cdot \frac{f(\pi_i)}{(1 - m) \sum_{j=1}^n f(\pi_j) + nmf(\bar{\pi}^*)},$$

³We here consider asexual reproduction as is standard in essentially all the literature on preference evolution in economics (see Waldman, 1994, for an exception).

⁴"Uniform dispersal" refers to the limit of uniform dispersal when there are finitely many islands, and their number is taken to infinity.

⁵We assume that the total number of offspring is large enough for the probability of there being fewer aspiring offspring than there are deceased adults at an island to be negligible.

where $0 < m \leq 1$ is the probability for each newborn to migrate to another island, $0 \leq s(\pi_i) \leq 1$ is the probability that i survives to the next demographic time period, and $f(\pi_i) > 0$ is i 's expected number of offspring, i 's *fecundity*. The first term in (1) is the probability that individual i survives to the next period. The second term is the expected number of i 's offspring who migrate and manage to secure a breeding spot on another island. This term contains three factors: (a) the probability for migration, (b) the expected number of available spots in a randomly drawn island, where $s(\bar{\pi}^*)$ denotes the average survival probability for individuals in the population at large, and (c) the individual's expected number of offspring that compete for spots in a random island, relative to the total number of offspring aspiring for spots in such islands, with $f(\bar{\pi}^*)$ denoting the fecundity in the population at large.⁶ The third term, finally, is the expected number of i 's offspring who manage to secure a “breeding spot” on the natal island. Also this term is the product of three factors: (a') the probability for not migrating, (b') the expected number of available spots on the island, and, for each available spot, the competition for the spot, among native and migrating offspring from other islands.

This scenario contains two *vital rates*: survival and fecundity.⁷ We note that individual fitness, defined in (1), is scale-invariant with respect to the second rate; if all fecundities are multiplied by the same positive factor, individual fitnesses remain unchanged. This is not true for the first rate, however; in general individual fitnesses change if all survival rates are scaled up or down. We also note that individual fitness is invariant under permutation of the material payoffs to the individual's island neighbors (the components of the vector $\pi_{-i} \in \mathbb{R}^{n-1}$).

2.2 Scenario B: genetic transmission and wars

We now add potential wars between islands to the baseline scenario just described. Following play of the material game, but before reproduction and migration, islands are randomly engaged in pairwise wars. In each such war, one island wins and one loses. All individuals in the losing island thus die before they reproduce; the winning island takes over all reproductive resources of the other island and thus doubles its members' fecundity. Technically, the double-sized pool of offspring of the winning island will split in two halves, one for each of the two islands, that they will treat as their “home” island. After the initial (war) stage and subsequent reproduction stage, the stages in Scenario A follow for all islands that were either not involved in war or that won a war. For the sake of clarity and simplicity, we assume in this war scenario that no adult survives after reproduction; $s(\pi_i) = 0$ for all values of π_i .

⁶Since the total number of islands is infinite, the probability is zero for the event that more than one of i 's offspring happen to migrate to the same island. For a derivation of an equation like (1) in a similar population model, see Lehmann and Balloux (2007).

⁷This scenario nests both overlapping and non-overlapping generations models that are frequently used in evolutionary biology (see, e.g., Ewens, 2004, and Pen, 2000). Overlapping generations obtain when (at least some) survival rates are positive, while non-overlapping generations obtain when all survival rates are zero, which amounts to assuming that all adults die at the end of each demographic time period.

Let $0 \leq \rho \leq 1$ denote the probability that any given island is drawn into war, the *war risk*, and let $v(\boldsymbol{\pi}, \bar{\pi}^*)$ denote the probability that an island, in which material payoff profile $\boldsymbol{\pi} \in \mathbb{R}^n$ obtains, wins a war when the average payoff in the rest of the population is $\bar{\pi}^*$, where v is assumed to be increasing and permutation invariant with respect to the material payoffs earned by the inhabitants of the island in question (the components of the vector $\boldsymbol{\pi} \in \mathbb{R}$). In other words, v has the properties of standard welfare functions. In this *war scenario*, the fitness of individual i writes

$$w(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*) = [(1 - \rho) + 2\rho v(\boldsymbol{\pi}, \bar{\pi}^*)] \cdot \left[m \cdot \frac{f(\pi_i)}{f(\bar{\pi}^*)} + (1 - m)n \cdot \frac{f(\pi_i)}{(1 - m) \sum_{j=1}^n f(\pi_j) + nm f(\bar{\pi}^*)} \right]. \quad (2)$$

The first factor contains two terms. The first term, $1 - \rho$, is the probability that the individual's island will not go to war. The second term, $2\rho v(\boldsymbol{\pi}, \bar{\pi}^*)$, is twice the probability that the island will go to war and win (all winning islands double their fecundity and spread their offspring uniformly over the two islands they now possess). The second factor in (2) is the same as the right-hand side of (1), in Scenario A, when all grown-ups die after reproduction ($s(\pi_i) = s(\bar{\pi}^*) = 0$). To see why this is the case, note that migrants who arrive at any island, irrespective of whether this island has been involved in war or not, come with probability $1 - \rho$ from an island that was not in war, and (recalling that the average probability of winning a war is $1/2$) with probability $\rho/2$ from an island that won a war. Moreover, victorious islands send out twice as many migrants as islands that did not go to war. Hence, the expected number of migrants who compete for the breeding spots in any given island is $m(1 - \rho + 2\rho/2) \cdot f(\bar{\pi}^*) = mf(\bar{\pi}^*)$, the same as in Scenario A.

However, unlike scenario A, this war scenario includes an element of competition between groups; an individual's fitness depends in part on the material wealth of its island. In other words, behaviors that favor the material well-being of other island members are evolutionarily valuable in the present scenario. Like in scenario A, individual fitness is invariant under permutation of the material payoffs to the individual's island neighbors.

2.3 Scenario C: cultural transmission

We finally consider a scenario of cultural transmission, whereby the traits in each parent generation are carried over by cultural transmission to the next generation. In every demographic time period, each adult dies and is replaced by exactly one child, who searches a trait to emulate, from its deceased (single) parent, another adult in its island, or an adult in another island. With probability $s(\pi_i) \in [0, 1]$, the loyalty of i 's child, the (unique) child of individual i , emulates its parent's trait. With probability $1 - m$ a non-loyal child searches for a trait to emulate among the (now dead) grown-ups in its natal island (including its own parent). With the complementary probability, $m > 0$, such a child draws a sample of n grown-ups from the population at large, and emulates the trait of one of them. The probability that an adult on any island is chosen as role model, when compared to others in

her island (by a non-loyal child), depends on her trait’s attractiveness relative to the attractiveness of the other grown-ups’ traits in her island. Likewise, the probability that a child who searches outside its native island will pick a certain island, when looking for a “role model”, is assumed to be proportional to the island’s relative attractiveness in the world at large.

Let $w(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*)$ be the expected number of children who emulate their trait from an individual with material payoff π_i when the other island members earn the material payoff vector $\boldsymbol{\pi}_{-i}$, and individuals in all other islands earn material payoff $\bar{\pi}^*$. Then

$$\begin{aligned} w(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*) &= s(\pi_i) + \\ &+ (1 - m) \cdot \left(n - \sum_{j=1}^n s(\pi_j) \right) \cdot \frac{f(\pi_i)}{\sum_{j=1}^n f(\pi_j)} \\ &+ m \cdot [1 - s(\bar{\pi}^*)] \cdot \frac{f(\pi_i)}{f(\bar{\pi}^*)}, \end{aligned} \tag{3}$$

where, for any individual j in i ’s island, $f(\pi_j) > 0$ is the attractiveness of the trait used by j . The first term in (3) is the probability that i ’s child loyally emulates its parent’s trait, without comparison with other adults’ traits. The second term concerns the event that some or all the children in i ’s island emulate their trait from one among the parents on the island. The product of the first two factors in this term is the expected number of such children and the third factor is the probability, for each such child, that it will choose to imitate individual i . The third term concerns the event that children from other islands emulate their trait from one of the parents on i ’s island. Written more explicitly, this term can be spelled out as

$$mn [1 - s(\bar{\pi}^*)] \cdot \frac{\sum_{j=1}^n f(\pi_j)}{n \cdot f(\bar{\pi}^*)} \cdot \frac{f(\pi_i)}{\sum_{j=1}^n f(\pi_j)},$$

where the first factor is the expected number of children who search outside their native islands, the second factor is the probability for each such child to decide for i ’s island, and the third is the conditional probability that it will then choose i as role model.

We note that, comparing this scenario to Scenario A, loyalty plays a similar role to survival, and attractiveness to fecundity. Moreover, the *cultural import propensity* m plays a similar role to migration in Scenario A. (These observations motivated the notation.) Under cultural transmission, individual fitness is scale independent with respect to attractiveness, just as individual fitness is scale independent with respect to fecundity in scenarios A and B, and an individual’s fitness is invariant under permutation of the material payoffs to the individual’s island neighbors.

2.4 Fitness monotonicity

Each of the three scenarios imposes some structure on the fitness function w and relates it to material payoffs. The subsequent analysis will be premised on the following monotonicity condition on the fitness function w with respect to material payoffs:

[M] (i) *Fitness is strictly increasing in own material payoff.* (ii) *The marginal effect of a neighbor’s material payoff on fitness is not larger than the marginal effect of own material payoff.* (iii) *Fitness is decreasing in the average material payoff in the population at large.*

3 Uninvadable strategies

In this and the next section we take an individual’s trait to be the strategy to use in the material game and characterize uninvadable strategies. In Section 5 we build on this characterization and let the individual’s trait instead be her utility function.

What matters for the long-run evolutionary success of any heritable trait is whether individuals with that trait on average have more descendants in future generations than individuals with other traits. A trait is called *uninvadable* if any other (mutant) trait that initially appears in a single individual, is bound to disappear from the population. This is quite a stringent evolutionary robustness condition. In this section, we take the traits to be (pure or mixed) strategies in the strategic interaction under consideration. More precisely, consider any material game $\langle n, X, \pi \rangle$, any transmission scenario whereby strategies are transmitted from one demographic time period to the next according to some fitness function $w : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$ satisfying Assumption [M], any resident strategy $x \in X$, and any strategy $y \in X$ that initially appears in a single individual. The strategy x is *uninvadable* if the random number of demographic time periods during which any given mutant strategy $y \neq x$ remains in the population is finite with probability one.

To get a hold on what strategies, if any, are uninvadable in this sense, we use the notion of an individual’s *lineage*, defined as the collection of this individual’s all descendants: her immediate descendants, defined as her offspring and also herself if she survives, the immediate descendants of her immediate descendants, etc. *ad infinitum*. Suppose that initially everybody in the population uses strategy x , and that suddenly one individual on an island, a mutant, switches to another strategy, y . To be favored by selection, individuals from the mutant lineage must colonize (via migration) new islands that are occupied by the resident strategy, and may do so as singletons or possibly along with other mutants. The latter may, for instance, be the case in our scenario B, when individuals in the winning group reproduce on the island that lost the war. Irrespective of whether an island of residents is colonized by a single or several mutants, the random time of “first extinction” of the *local lineage* of mutants—defined as the collection of all descendants of mutants who initially jointly colonized that island—is finite with probability one, since the migration rate by assumption is positive and constant.⁸

The *lineage fitness* of the mutant strategy y is defined as the average fitness of a mutant’s local lineage members, the average being taken over all demographic time periods until the first period in which the local lineage is extinct and over all possible initial conditions of that

⁸Note that typically, while locally extinct, lineage members may still live on other islands, and some of them may even move to the mutant’s native island. However, this last event has probability zero in the limit when the number of islands is infinite.

local lineage (single or multiple simultaneous mutants). It turns out that lineage fitness, defined this way, can be used to characterize uninvadability. In order to make this precise, let x be the initial (incumbent or resident) strategy and let y be any mutant strategy. Let $\pi(y|k)$ and $\pi(x|k)$ denote the material payoffs to individuals using strategies y and x , respectively, in any island in which exactly k individuals use strategy y (and the others use x). Assuming henceforth that the transmission scenario is such that individual fitness is invariant under permutation of neighbors' material payoffs, the lineage fitness of strategy y can be formally defined as

$$W(y, x) = \sum_{k=0}^{n-1} p_k(y, x) \cdot w(\pi(y|k), \langle \pi(y|k), \pi(x|k) \rangle, \bar{\pi}(x)). \quad (4)$$

Here $p_k(y, x)$, for $k = 0, 1, \dots, n - 1$ is the probability for a randomly drawn member i from its local lineage to coexist with k other lineage members in its island. Likewise, $\langle \pi(y|k), \pi(x|k) \rangle \in \mathbb{R}^{n-1}$ denotes the vector of material payoffs to the other $n - 1$ members of i 's island—among whom k use the mutant strategy y and the others use the resident strategy x . Finally, $\bar{\pi}(x)$ is the material payoff earned by each individual in an island where all play the resident strategy x .

With lineage fitness so defined, uninvadability can be succinctly characterized as follows:⁹

Proposition 1 (Lehmann et al., 2016) *A strategy x is uninvadable if and only if*

$$W(y, x) \leq 1 \quad \forall y \in X. \quad (5)$$

The characterizing criterion (5) compares the lineage fitness of a single initial y -mutant, $W(y, x)$, to unity, the lineage fitness of each resident in this population. Given that the number of mutants is finite in an infinite population, such a resident has mutant neighbors with probability zero, and consequently has lineage fitness $W(x, x) = 1$. Equivalently, the criterion can be written

$$x \in \arg \max_{y \in X} W(y, x). \quad (6)$$

In other words, a strategy is uninvadable if and only if it is a best reply to itself in terms of lineage fitness.

In a sense, Proposition 1 provides a partial answer to our “as if”-question: any uninvadable strategy $x \in X$ is a best reply for a player in an n -player game in which the player obtains payoff $W(y, x)$ from playing strategy y if all others play strategy x . The relevance of this answer is, however, limited by (a) the fact that the weights $p_k(y, x)$ depend on the strategies x and y , a dependency that in general is non-trivial, and (b) the payoff function in this game is not fully defined, it is only specified in the special case when all others' strategies are identical. In order to get beyond this point, we henceforth focus on what biologists call weak selection.

⁹In Lehmann, Alger, and Weibull (2015) we proved this result for scenarios where new islands can be colonized only by singleton mutants (covering scenario A and C). Lehmann et al. (2016) extended that result to allow for scenarios in which multiple offspring from the same group can reproduce in the same non-natal island, such as in Scenario B.

3.1 Weak selection

In biology weak selection refers to situations in which the fitness effects of heritable traits, here strategies, are small and can be represented in terms of their first-order effects (see, e.g., Nagylaki, 1992, Rousset, 2004). Formally, we assume that an individual's material payoff affects his vital rates marginally, by multiplicative factors of the form $\exp(\delta\pi)$ for small $\delta > 0$ (for examples, see Section 4). In other words, all higher-order terms in a Taylor expansion of how fitness depends on strategies are neglected, with reproduction and survival being independent of payoff when $\delta = 0$. This amounts to a useful linearization, not unlike standard first-order-conditions in economics.

The key implication for evolutionary analysis of the effects of material payoffs on vital rates being weak, is that the matching probabilities $p_k(y, x)$, while still depending on the transmission events, do not depend on the strategies x and y at hand. In the analysis, these matching probabilities can then be evaluated under what biologists refer to as *neutral drift* or *the neutral process*, i.e., when every individual in the population is exchangeable, in the sense that they all face the exact same prospects of producing descendants (recall that fitness is the expected number of descendants, the realized number of descendants in the following generation being a random variable; in the neutral process all individuals face the same distribution for this random variable); see, e.g., Crow and Kimura (1970) or Rousset (2004). The probability for a randomly drawn descendant of an ancestor, be it a resident or mutant, to coexist in its island with k other descendants of the same ancestor is then solely determined (for each k) by the vital rates in a population in which everybody uses the same strategy, no matter which. Let p_k^0 denote the matching probabilities in the neutral process, that is, when $\delta = 0$, and hence all vital rates and matching probabilities are independent of what strategies are used in the population; p_k^0 is independent of mutant and resident strategies for all $k = 0, 1, \dots, n - 1$. Let $\mathbf{p}^0 = (p_0^0, p_1^0, \dots, p_{n-1}^0)$.

A key implication of weak selection is that it is possible to characterize the set of uninvadable strategies in terms of what we call a strategy's *lineage payoff-advantage*. Formally, we define the lineage payoff-advantage of a mutant strategy $y \in X$ in a population of residents using strategy $x \in X$ as

$$\Pi(y, x) = \sum_{k=0}^{n-1} p_k^0 \cdot \tilde{\pi}^{(k)}(y, x), \quad (7)$$

where $\tilde{\pi}^{(k)}(y, x)$ is the mutant's *payoff advantage* when there are k other mutants in her or his island, defined by

$$\tilde{\pi}^{(k)}(y, x) = \pi(y|k) - \lambda_0 \cdot \left[\frac{k}{n-1} \pi(y|k) + \frac{n-1-k}{n-1} \pi(x|k) \right]. \quad (8)$$

The first term in this expression is the payoff of a descendant of the initial mutant who finds herself in an island with k other such descendants. The term in square brackets is the average material payoff earned by the other members in the island. We refer to the parameter λ_0 as *local competitiveness*. It can be derived for each of the above transmission scenarios from the associated fitness equations (1), (2), and (3), respectively. The parameter can be interpreted as the relative effect on all neighbors' fitnesses, in comparison with the effect on own fitness,

from small changes in an individual’s material payoff. Formally, for any individual i and any neighbor j :

$$\lambda_0 = - \left. \frac{(n-1) w_n(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*)}{w_1(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*)} \right|_{\delta=0}, \quad (9)$$

where w_j denotes the partial derivative of w with respect to its j^{th} argument ($j = 1, \dots, n$). We will derive explicit expressions for this parameter in examples below. We note that the parameter λ_0 depends on the fitness function w , which in turn depends on population structure and the transmission scenario, but it does not depend on the material payoff function *per se*.

The lineage payoff-advantage is the average payoff advantage of the initial mutant’s descendants, in the island where the mutant initially appeared. It reflects two facts implied by population structure. First, it measures the ability to outcompete one’s neighbors, with λ_0 capturing the importance of doing so. Second, descendants of the initial mutant who stay in the natal island interact with each other. This is captured by the matching probabilities p_k^0 and can be quantified by the pairwise relatedness coefficient

$$r_0 = \sum_{k=0}^{n-1} \frac{k}{n-1} p_k^0, \quad (10)$$

which is the expected share of other lineage members among one’s neighbors, and is equivalent to the probability (in the neutral process) that two randomly sampled individuals in an island inherited their strategy from a common ancestor (Michod and Hamilton, 1980). In particular, $r_0 = 0$ is zero if and only if $p_0^0 = 1$, that is, when no mutant, when rare, ever meets another lineage member. The coefficient of relatedness will play a central role at various stages of our analysis.

Under our maintained payoff-monotonicity condition **[M]**, it can be shown that a strategy is uninvadable under weak selection if and only if there is no strategy that would have a lineage payoff-advantage if appearing as a rare mutant. Formally:¹⁰

Proposition 2 (Lehmann, Alger, and Weibull, 2015) *Under weak selection, a strategy $x \in X$ is uninvadable if and only if*

$$\Pi(y, x) \leq \Pi(x, x) \quad \forall y \in X. \quad (11)$$

Moreover, $n-1 \leq \lambda_0 \leq 1$.

The intuition behind this proposition is straightforward. By allowing for linearization of fitness, weak selection implies that average *fitness* of the initial mutant’s local lineage can be

¹⁰As was noted above, in Lehmann, Alger, and Weibull (2015) we did not allow for reproductive processes where offspring from the same group can reproduce in the same non-natal group, while we do allow for that here. However, in force of Proposition 1 and its proof in Lehmann, Alger, and Weibull (2015), we prove in the Appendix that Proposition 2 holds generally, and we there also determine a lower bound on λ_0 and identify its functional form, as given in (9).

operationally expressed in terms of average *material payoffs* in the local lineage. Condition (11) says that a resident strategy x is uninvadable if it also maximizes the lineage payoff-advantage that a mutant could hope to get in a population where x is the resident strategy. In doing so, such a strategy preempts entry by any mutant strategy.

3.2 Nash equilibrium

Thanks to this result, we are in a position to establish a general link between uninvadability under weak selection and individual maximizing behavior. In essence, a strategy is uninvadable under weak selection if and only if it is a Nash equilibrium strategy in a game with suitably chosen individual payoff functions. In other words, weak natural selection leads to behaviors that are identical with equilibrium behavior among rational individuals who freely choose their strategies in order to maximize their (expected) utility.

In order to express this result, we need some notation and a definition. Let Δ be the unit simplex of probability vectors $\boldsymbol{\mu} = (\mu_0, \mu_1, \dots, \mu_{n-1})$. (In particular, $\mathbf{p}^0 \in \Delta$.) For any $\boldsymbol{\mu} \in \Delta$, any player i , and any strategy profile $\mathbf{x} \in X^n$, let $\tilde{\mathbf{x}}$ be a *random strategy-profile* such that with probability μ_k (for $k = 0, 1, \dots, n-1$) exactly k of the $n-1$ components in \mathbf{x}_{-i} are replaced by x_i , with equal probability for each subset of k replaced components, while the remaining components in \mathbf{x}_{-i} keep their original value, and let the utility function $u_{\boldsymbol{\mu}, \lambda} : X^n \rightarrow \mathbb{R}$ be defined by

$$u_{\boldsymbol{\mu}, \lambda}(x_i, \mathbf{x}_{-i}) = \mathbb{E}_{\boldsymbol{\mu}} \left[\pi(x_i, \tilde{\mathbf{x}}_{-i}) - \lambda \cdot \sum_{j \neq i} \pi(\tilde{x}_j, \tilde{\mathbf{x}}_{-j}) \mid \mathbf{x} \right] \quad \forall \mathbf{x} \in X^n, \quad (12)$$

where $1 - n \leq \lambda \leq 1$.

In the special case when $\lambda = 0$ and $\mu_0 = 1$, this goal function is but that of the familiar *Homo oeconomicus*, individuals who evaluate any strategy profile $\mathbf{x} = (x_i, \mathbf{x}_{-i})$ solely in terms of its material consequences for the individual him- or herself; $u_{\boldsymbol{\mu}, \lambda}(x_i, \mathbf{x}_{-i}) = \pi(x_i, \mathbf{x}_{-i})$ for all $(x_i, \mathbf{x}_{-i}) \in X^n$.

In the more general case when $\lambda = 0$ but $\boldsymbol{\mu} \in \Delta$ is arbitrary, this goal function is identical with that of *Homo moralis* with morality profile $\boldsymbol{\mu}$ (Alger and Weibull, 2016), whereby an individual evaluates any strategy profile (x_i, \mathbf{x}_{-i}) by pondering her expected material payoff if all, some, or none of the others in her island would use the same strategy as herself (drawn randomly according to $\boldsymbol{\mu}$), instead of playing their strategies, given by \mathbf{x}_{-i} ; $u_{\boldsymbol{\mu}, \lambda}(x_i, \mathbf{x}_{-i}) = \mathbb{E}_{\boldsymbol{\mu}} [\pi(x_i, \tilde{\mathbf{x}}_{-i}) \mid \mathbf{x}]$.

When $\lambda \neq 0$ and $\boldsymbol{\mu} \in \Delta$ is arbitrary, the individual also considers the expected material payoff consequences, under the hypothetical play given by $\boldsymbol{\mu}$, for her island mates. She then evaluates any strategy profile (x_i, \mathbf{x}_{-i}) in terms of its expected *material payoff advantage*, as compared with the others in her island, $\pi(x_i, \tilde{\mathbf{x}}_{-i}) - \lambda \cdot \sum_{j \neq i} \pi(\tilde{x}_j, \tilde{\mathbf{x}}_{-j})$, rather than to her *material payoff*. A positive weight $\lambda > 0$ expresses a form of *envy* or *spite*. If instead $\lambda < 0$, then it is as if individuals care positively, or *altruistically*, about their neighbors. Because the utility function $u_{\boldsymbol{\mu}, \lambda}$ combines both morality and a comparison with others' material payoffs, we refer to an individual with such a utility function as a *competitive moralist*.

A competitive moralist of key importance is when $\lambda = \lambda_0$ and $\boldsymbol{\mu} = \boldsymbol{p}^0$, in which case we will write u^0 for $u_{\boldsymbol{\mu}, \lambda}$:

$$u^0(x_i, \boldsymbol{x}_{-i}) = \mathbb{E}_{\boldsymbol{p}^0} \left[\pi(x_i, \tilde{\boldsymbol{x}}_{-i}) - \lambda_0 \cdot \sum_{j \neq i} \pi(\tilde{x}_j, \tilde{\boldsymbol{x}}_{-j}) \mid \boldsymbol{x} \right] \quad \forall \boldsymbol{x} \in X^n \quad (13)$$

We note that for any strategy profile $\boldsymbol{x} = (x_i, \boldsymbol{x}_{-i}) \in X^n$ such that $x_i = y$ and $x_j = x$ for all $j \neq i$, $u^0(x_i, \boldsymbol{x}_{-i}) = \Pi(y, x)$, i.e., the competitive moralist's utility then coincides with the lineage payoff advantage of a mutant strategy y that appears in a population of residents playing strategy x (compare equations (12) and (7)). This is why u^0 is of particular importance. In fact, it provides a precise, operational, and general answer to our initial “as if”-question, for the case of weak selection:

Corollary 1 *A strategy \hat{x} is uninvadable under weak selection if and only if it is a symmetric Nash equilibrium strategy in the n -player game in which every player's strategy set is X and every player is a competitive moralist with payoff function u^0 .*

To see more concretely what this result means, consider first the case $n = 2$. Then uninvadability will induce behavior that appears to result from a game in which each individual freely chooses her strategy in order to maximize his utility, defined for player $i = 1, 2$ (and $j \neq i$) by

$$u^0(x_i, x_j) = p_0^0 \cdot [\pi(x_i, x_j) - \lambda_0 \cdot \pi(x_j, x_i)] + p_1^0 \cdot [\pi(x_i, x_i) - \lambda_0 \cdot \pi(x_i, x_i)]. \quad (14)$$

The results above show the evolutionary origin of these terms: the first square parenthesis corresponds to the individual's payoff advantage in the event that the opponent is not a member of the same lineage, and the second square parenthesis to the payoff advantage in the event that the opponent is a member of the same lineage. Each square parenthesis is given its probability weight, where $p_0^0 + p_1^0 = 1$. In the case $n = 2$, $p_1^0 = r_0$. Reorganizing the right-hand side in (14) reveals motivational factors familiar to behavioral economists. First, by writing

$$u^0(x_i, x_j) = (1 - \lambda_0)(1 - r_0) \cdot \pi(x_i, x_j) + \lambda_0(1 - r_0) \cdot [\pi(x_i, x_j) - \pi(x_j, x_i)] + (1 - \lambda_0)r_0 \cdot \pi(x_i, x_i), \quad (15)$$

the utility function can be interpreted as the sum of three terms, where the first represents “pure self-interest” (own material payoff), the second a “comparison with the Joneses” (the difference between own material payoff and that of the neighbor), and the third a “moral” concern (what is the “right thing to do if others in the population act like me”). Second, by instead bundling the terms as follows, the utility function can alternatively be interpreted as a combination of pure self-interest, “net altruism” (altruism minus spite), and morality:

$$u^0(x_i, x_j) = (1 - r_0) \cdot \pi(x_i, x_j) - \lambda_0(1 - r_0) \cdot \pi(x_j, x_i) + (1 - \lambda_0)r_0 \cdot \pi(x_i, x_i). \quad (16)$$

The two formulations show that evolution by natural selection lends support to behaviors that are compatible with motives that have been studied in economics before, in particular altruism (Becker, 1976), inequity aversion (Fehr and Schmidt, 1999), and morality (Alger and Weibull, 2013). Note also that if $r_0 = 0$ and $\lambda_0 = -1$, an individual with this utility function seeks to maximize total material welfare.

Irrespective of how one chooses to bundle the terms, if $r_0\lambda_0 \neq 0$ an individual with utility function u^0 takes into account not only his or her material self-interest, but also other motives of a social and/or moral nature.

As island size n increases from $n = 2$, both the number of terms in the utility function u^0 and the variety of matching probabilities increases. For example, for $n = 3$, and writing (x, y, z) for the strategy profile used in an island:

$$\begin{aligned} u^0(x, y, z) = & p_0^0 \cdot \left[\pi(x, y, z) - \lambda_0 \cdot \frac{\pi(y, z, x) + \pi(z, x, y)}{2} \right] \\ & + p_1^0 \cdot \left[\pi(x, x, z) - \lambda_0 \cdot \frac{\pi(x, z, x) + \pi(z, x, x)}{2} \right] \\ & + p_1^0 \cdot \left[\pi(x, y, x) - \lambda_0 \cdot \frac{\pi(y, x, x) + \pi(x, x, y)}{2} \right] \\ & + p_2^0 \cdot \left[\pi(x, x, x) - \lambda_0 \cdot \frac{\pi(x, x, x) + \pi(x, x, x)}{2} \right]. \end{aligned} \quad (17)$$

In spite of this complexity, as we show next there is a remarkably simple way to characterize uninvadable strategies when the material payoff function is differentiable.

4 First-order condition for uninvadability

We proceed to analyze strategy evolution under weak selection, focusing on strategic interactions in which the strategy space is one-dimensional, $X \subseteq \mathbb{R}$, and the material payoff function π is continuously differentiable.¹¹ We then write π_j for the partial derivative of π with respect to its j^{th} argument ($j = 1, \dots, n$). It is immediate that, for $n = 2$, Corollary 1 and any of the expressions (14)-(16) together imply that any interior uninvadable strategy \hat{x} must then satisfy the first-order condition:

$$\pi_1(y, \hat{x}) + \kappa_0 \cdot \pi_2(\hat{x}, y) = 0, \quad (18)$$

when evaluated at $y = \hat{x}$. The first term is the marginal effect upon own material payoff and the second term is the marginal effect upon the other individual's material payoff, which depends on the coefficient

$$\kappa_0 = \frac{r_0 - \lambda_0}{1 - \lambda_0 r_0}, \quad (19)$$

¹¹The generalization to higher-dimensional Euclidean spaces is straight-forward. If $X \subseteq \mathbb{R}^d$ for some integer $d > 1$, then the first-order conditions considered below will hold for each component of the strategy vector $y \in \text{int}(X)$ in question.

which summarizes the behavioral effects of assortativity and local competition brought about by population structure. In all scenarios $\lambda_0 r_0 < 1$, so the difference $r_0 - \lambda_0$ determines the sign of κ_0 , which indicates whether, on balance, an uninvadable strategy is materially pro- or anti-social, in the sense that if used by the two interacting individuals, it results in a higher or a lower average material payoff than a strategy that purely self-interested individuals would choose. Here, a strategy \hat{x} is *pro-social* if $\kappa_0 > 0$, *anti-social* if $\kappa_0 < 0$, and *individualistic* if $\kappa_0 = 0$.

The analytical power that differentiability provides becomes particularly striking for island of arbitrary size $n \geq 2$. As noted above (see (17)), if island size is $n = 3$, then u^0 has four terms, each receiving a different weight (p_0^0 , p_1^0 , or p_2^0). It turns out that when π is differentiable, an interior uninvadable strategy must still satisfy a differential equation with only two terms, like (18), and where the coefficient κ_0 can be expressed explicitly in terms of group size n , the coefficient of local competitiveness, λ_0 , and the coefficient of pairwise relatedness, r_0 (the probability that, in a homogeneous population, the neighbor of a randomly drawn member of a certain lineage is also a lineage member).

Proposition 3 *Let $n \geq 2$, and let $\pi : X^n \rightarrow \mathbb{R}$ be continuously differentiable with $X \subseteq \mathbb{R}$. If a strategy $\hat{x} \in \text{int}(X)$ is uninvadable under weak selection, then*

$$[\pi_1(y, \hat{\mathbf{x}}) + (n-1)\kappa_0 \cdot \pi_n(y, \mathbf{y})]_{y=\hat{x}} = 0, \quad (20)$$

where $\hat{\mathbf{x}}$ is the $(n-1)$ -dimensional vector whose components all are \hat{x} , and \mathbf{y} is the $(n-1)$ -dimensional vector whose components all are y , and

$$\kappa_0 = \frac{r_0 - \lambda_0 \hat{r}_0}{1 - \lambda_0 r_0}, \quad (21)$$

with $\hat{r}_0 = [1 + (n-2)r_0]/(n-1)$. Furthermore, $-1 \leq \kappa_0 \leq 1$; in particular, $\kappa_0 = -1$ if $\lambda_0 = 1$, $r_0 < 1$ and $n = 2$, and $\kappa_0 = 1$ if $\lambda_0 = 1 - n$.

The coefficient κ_0 has an intuitive interpretation as a marginal substitution rate. It gives the number of units of own material payoff that any given individual is willing to forgo to increase the material payoff of each neighbor by one unit. Absent local competition, i.e., when $\lambda_0 = 0$, we have $\kappa_0 = r_0$, meaning that the individual is willing to forgo material payoff at a rate given by relatedness, r_0 . This is intuitive, since (pairwise) relatedness can be thought of as a measure of the recipient's ability, relative to that of the donor, to transmit a given trait (here a strategy) to the next generation (see, e.g., Frank, 1998). When $\lambda_0 > 0$, a transfer of own payoff to neighbors has several consequences that need to be taken into account in the substitution rate. Let us first consider the case when there is but one neighbor, that is $n = 2$. A payoff transfer to this neighbor increases competition from the neighbor at rate λ_0 (since λ_0 measures the relative increase in competition in the neighborhood of an individual when its payoff is varied, see (9)). Hence, the fitness benefit to the donor from giving the transfer to the neighbor is reduced by λ_0 , so that the numerator in (21) becomes $r_0 - \lambda_0$. Moreover, a transfer of resources to the neighbor alleviates the competition that the

neighbor experiences, and the neighbor is related to the donor according to coefficient r_0 . Hence, the cost of the transfer is reduced by $\lambda_0 r_0$, which explains the denominator in (21).

Second, when there are multiple neighbors, $n > 2$, one also has to account for the fact that a transfer given to one neighbor affects competition not only for the donor but for every neighbor. The $n - 1$ neighbors of any one recipient include the donor (1 of the $n - 1$ neighbors) and each of the $n - 2$ other neighbors. The donor is related by coefficient 1 to itself, and the other $n - 2$ neighbors are related to the donor by r_0 , and this is precisely the average relatedness factor \hat{r}_0 . A transfer of resources to any neighbor thus enhances the competition the other neighbors experience. Therefore, the fitness benefit of the transfer to the donor is reduced by $\lambda_0 \hat{r}_0$, rather than by λ_0 (where we note that $\hat{r}_0 = 1$ when $n = 2$), which explains the numerator of κ_0 . In the denominator, the cost of the transfer is still reduced by $\lambda_0 r_0$, which is the expected alleviation of competition that the transfer induces for the individual's neighbors (recall that λ_0 accounts for all neighbors through the term $(n - 1)$).

Proposition 3 is useful not only because it proposes an easy way to find uninvadable strategies. It also allows the analyst to determine whether, on balance, uninvadable strategies are pro-social, individualistic, or anti-social for any island size $n \geq 2$. Indeed, (20) immediately implies that we obtain pro-sociality if $\kappa_0 > 0$, individualism if $\kappa_0 = 0$, and anti-sociality if $\kappa_0 < 0$. Straightforward calculations further show that $\kappa_0 > 0$ if and only if $\lambda_0 < r_0/\hat{r}_0$, and $\kappa_0 > r_0$ if and only if $\lambda_0 < 0$.

The remainder of this section is devoted to calculating r_0 and λ_0 , and thus κ_0 , in a few specifications of the three transmission scenarios, and to discussing how the extent of pro-sociality or anti-sociality depends on the vital rates.

4.1 Scenario A

Let each individual's fecundity in the fitness function for this scenario (equation (1)) be exponentially increasing in the individual's material payoff:

$$f(\pi_i) = f_0 \cdot \exp(\delta_f \cdot \pi_i), \quad (22)$$

where $f_0 > 0$ is baseline fecundity and $\delta_f \geq 0$ represents the sensitivity of fecundity with respect to material payoff; δ_f is called the *intensity of selection*. Furthermore, let the survival probability be fixed and the same for all individuals, $s(\pi_i) = s_0$, where $0 \leq s_0 < 1$.

Weak selection amounts to analyzing the effect of the material payoffs on fitness in the limit as δ_f tends to zero. To calculate pairwise relatedness r_0^A , one uses a recursion equation that details the events in which this happens. In the present Scenario A, one thus obtains (see Appendix):

$$r_0^A = \frac{(1 - m)^2 + (1 + m^2) s_0}{n - (n - 1)(1 - m)^2 + (1 - (n - 1)m^2) s_0}. \quad (23)$$

which is decreasing in migration rate m and group size n . In this scenario, one obtains for

local competitiveness that (see Appendix):

$$\lambda_0^A = \frac{(n-1)(1-m)^2}{n-(1-m)^2}. \quad (24)$$

Both r_0^A and λ_0^A are independent of the fecundity factor δ_f , the reason being that it scales all individuals' material payoffs equally. Moreover, λ_0^A is independent of the survival rate s_0 , since this is the same for all individuals and hence does not affect the composition of local competitors. Both r_0^A and λ_0^A depend on the migration propensity, m . In particular, if all offspring always migrate from their native island, that is, if $m = 1$, then $r_0^A = \lambda_0^A = 0$, and pure self-interest obtains: $u^0 = \pi$. This is because there are infinitely many islands and offspring migrate individually and statistically independently. Consequently, any two migrants from the same island have zero probability of arriving at the same island. By contrast, for intermediate migration probabilities, $m \in (0, 1)$, both r_0^A and λ_0^A lie strictly between zero and one. Moreover, both r_0^A and λ_0^A are decreasing in the migration rate m and r_0^A is increasing in the background survival rate s_0 . Applying (21), we obtain:

$$\kappa_0^A = \frac{2(1-m)s_0}{2(1-m)s_0 + n[2-m(1-s_0)]}. \quad (25)$$

To illustrate the role played by survival, consider first the special case when $s_0 = 0$. Then the relatedness and the competitiveness effects exactly cancel out, and $\kappa_0^A = 0$ for all migration rates. In other words, *Homo oeconomicus* then prevails regardless of migration rate m and the uninvadable strategy is independent of the level of migration. In biology, this result is known as Taylor's (1992a) cancellation result.¹² It is a pivotal result of strategy evolution in spatially structured populations, noticed initially in agent-based simulations by Wilson, Pollock, and Dugatkin (1992), proven formally by Taylor (1992a) for the island model, and then shown to hold for arbitrary migration patterns between groups (e.g., Taylor, 1992b, Rousset, 2004, and Ohtsuki, 2012). This result is a useful yardstick for understanding how changes in parameter values can tip the balance either in the direction of pro-sociality or anti-sociality. Such tipping effects will be investigated in what follows.

Turning now to the case where $s_0 > 0$, where grown-ups may survive, we see that since λ_0^A is independent of s_0 and r_0^A is increasing in s_0 , it must then be that $r_0^A > \lambda_0^A$, and thus $\kappa_0^A > 0$ for all migration rates. In this scenario, then, the positive survival probability induces pro-sociality. This is because survival enhances relatedness by making some individuals interact

¹²To see why this is the case, consider a focal individual expressing an action producing $B/(n-1)$ units of fecundity (by way of increasing material payoff) for each of her group neighbors. This implies that the gains in the number of fitness units to the trait underlying the action is Br_0^A , since each neighbour carries the type of the focal individual with probability r_0^A . Producing additional offspring, however, increases local competition if these offspring do not migrate to other groups. This implies that the loss in the number of fitness units to the type underlying the action is $B[1/n+r_0^A(n-1)/n](1-m)^2$, since each non-dispersing offspring [probability $(1-m)$] of the focal individual and of her related neighbours enters in competition with probability $(1-m)$ with the new created offspring, and where the focal individual's offspring is a fraction of $1/n$ of the group's offspring experiencing the increase in competition. Owing to the fact that $r_0^A = (1-m)^2 [1/n + r_0^A(n-1)/n]$ (see (78)), fitness gains and losses exactly cancel out ($Br_0^A = B[1/n + r_0^A(n-1)/n](1-m)^2$).

with their own descendants, which thus increases the likelihood of interaction with one's type.

Figure 1 shows how κ_0 depends on the migration rate m when $s_0 = 1/n$, for $n = 2$ (black solid) and $n = 10$ (black dashed), and when $s_0 = 0.8$ for $n = 2$ (blue) and $n = 10$ (blue dashed), as well as $s_0 = 0$ (pink). Pro-sociality declines with n , at any given migration rate $m < 1$, and it declines with m , for any given island size n . However, anti-sociality never obtains.

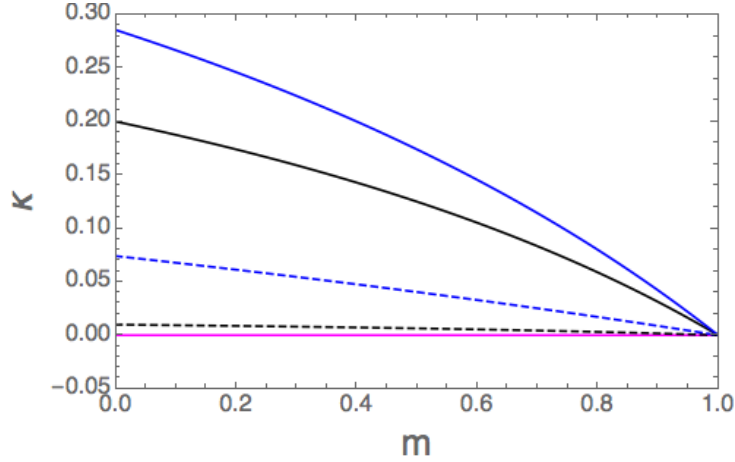


Figure 1: The value of κ_0^A as a function of the migration rate m .

4.2 Scenario B

Let the fecundity function in the fitness function for this scenario (equation (2)) be the same as in scenario A, (22), and assume that the probability of winning a war depends on the two islands' aggregate material payoffs according to

$$v_n(\boldsymbol{\pi}, \bar{\boldsymbol{\pi}}^*) = \frac{\exp(\delta_v \cdot V_n(\boldsymbol{\pi}))}{\exp(\delta_v \cdot V_n(\boldsymbol{\pi})) + \exp(\delta_v \cdot V_n(\boldsymbol{\pi}^*))}, \quad (26)$$

where $\boldsymbol{\pi}$ is the material payoff profile in the “home” island, $\boldsymbol{\pi}^*$ the material payoff profile in the “enemy” island, and $V_n : \mathbb{R}^n \rightarrow \mathbb{R}$ is a strictly increasing symmetric function (like any standard welfare function). The values $V_n(\boldsymbol{\pi})$ and $V_n(\boldsymbol{\pi}^*)$ of the function V_n represent the “strengths” of the two islands. Equation (26) is a logistic version of the Tullock contest function (Tullock, 1980), see Skaperdas (1996). It spans a continuum of cases, from all islands having the same chance to win any war, $\delta_v = 0$, to the case in which the materially wealthiest island is almost sure to win any war ($\delta_v \rightarrow +\infty$). Letting $\delta_f = \sigma_f \cdot \delta$ in equation (22) and $\delta_v = \sigma_v \cdot \delta$, for non-negative parameters $\sigma_f \geq 0$, $\sigma_v \geq 0$, and $\delta > 0$, we can let both sensitivity parameters tend to zero at proportional rates by focusing on the limit as $\delta \rightarrow 0$.

We show in the appendix that if an island's strength is proportional to its total material payoff ($V_n(\boldsymbol{\pi}) = n\bar{\pi}$ and $V_n(\boldsymbol{\pi}^*) = n\bar{\pi}^*$), then pairwise relatedness is

$$r_0^B = \frac{(1-m)^2}{n - (n-1)(1-m)^2}. \quad (27)$$

It is easily verified that this is exactly the same relatedness as in Scenario A (in the current special case when all grown-ups die). In other words, relatedness is not affected by the risk of war in this special case. Likewise, local competitiveness takes the following form:

$$\lambda_0^B = \frac{(n-1)(1-m)^2 - \rho(n-1)n\sigma_v/(2\sigma_f)}{n - (1-m)^2 + \rho n\sigma_v/(2\sigma_f)}. \quad (28)$$

We note that $\lambda_0^B = \lambda_0^A$ when $\rho = 0$, and that λ_0^B is reduced by the risk of war. Combining the expressions for r_0^B and λ_0^B , we obtain

$$\kappa_0^B = \frac{\rho\sigma_v}{\rho\sigma_v + 2m(2-m)\sigma_f}. \quad (29)$$

Like in Scenario A, both pairwise relatedness and local competitiveness are decreasing in the migration rate m , and so is κ_0^B . The threat of war ($\rho > 0$) nourishes pro-sociality: κ_0^B is increasing in ρ . We also note that local competitiveness changes sign when the risk of war is $\rho^* = 2\sigma_f(1-m)^2/(n\sigma_v)$. Local competitiveness is positive at lower risks of war and negative at higher risk levels for war. Hence, if the risk of war is above this critical value (which is lower the higher is the migration propensity), then the spiteful component of the utility function u^0 changes sign to altruism.¹³ This critical risk level is lower the more migration there is, and tends to zero as $m \rightarrow 1$.

Figure 2 shows κ_0^B as a function of the migration rate m , when $\sigma_f = \sigma_v$, for war risk $\rho = 0$ (the pink curve), $\rho = 0.4$ (the orange curve), and $\rho = 0.8$ (the blue curve).

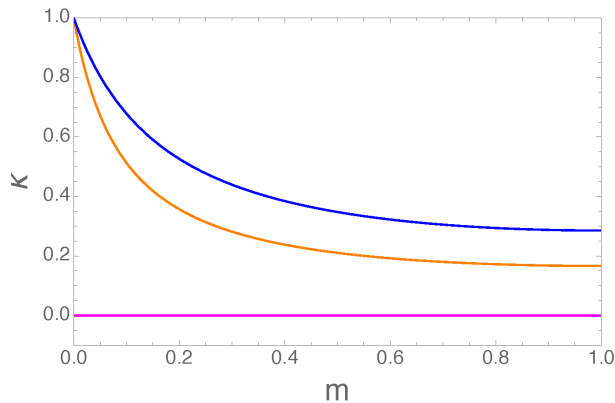


Figure 2: The value of κ_0^B as a function of the migration rate m .

Remark 1 *The above calculations are based on the assumption that the probability of winning a war depends on the group's total material payoff. Arguably, this is a natural benchmark case. However, sometimes the success or failure in conflicts depends on the strongest or weakest member of one's group.¹⁴ We here briefly consider the latter case, and focus on*

¹³A similar effect obtains if groups are exposed to the threat of environmental shocks, and if some aggregate measure of material wealth in the group mattered for its ability to withstand such shocks (see, e.g., Eshel, 1972, and Aoki, 1982, for models considering such cases under strategy evolution).

¹⁴A host of other hypotheses about group strength could be explored, see, e.g., Konrad (2014) and the references therein.

groups of size $n = 2$. Hence, now let $V(\boldsymbol{\pi}) = \min_j \pi_j$. By way of replacing $\pi_1 + \pi_2$ by $\min\{\pi_1, \pi_2\}$ the above general formula

$$u^0(x_i, x_j) = (1 - r_0) \cdot \pi(x_i, x_j) - \lambda_0(1 - r_0) \cdot \pi(x_j, x_i) + (1 - \lambda_0)r_0 \cdot \pi(x_i, x_i)$$

needs to let λ_0 depend on who has the highest material payoff. We obtain the following generalization:

$$\begin{aligned} u^0(x_i, x_j) &= (1 - r_0) \left(1 - \lambda_0^{Ba} \cdot 1_{\pi_i < \pi_j} - \lambda_0^{Bb} \cdot 1_{\pi_i \geq \pi_j} \right) \cdot \pi(x_i, x_j) \\ &\quad - \lambda_0^{Ba} (1 - r_0) \cdot [\pi(x_j, x_i) - \pi(x_i, x_j)]_+ \\ &\quad + \lambda_0^{Bb} (1 - r_0) \cdot [\pi(x_i, x_j) - \pi(x_j, x_i)]_+ \\ &\quad + (1 - \lambda_0)r_0 \cdot \pi(x_i, x_i). \end{aligned} \quad (30)$$

where

$$\lambda_0^{Ba} = \frac{(1 - m)^2}{2 - (1 - m)^2}$$

and

$$\lambda_0^{Bb} = \frac{(1 - m)^2 - \rho\sigma_v / (2\sigma_f)}{2 - (1 - m)^2 + \rho\sigma_v / (2\sigma_f)}.$$

We note the similarity with the well-known class of social preferences proposed in Fehr and Schmidt (1999). In agreement with their parameter assumption $\beta \leq \alpha$, our evolutionary scenario gives $\lambda_0^{Bb} \leq \lambda_0^{Ba}$. We also note that λ_0^{Ba} is independent of the war risk ρ , and that it is positive, just as their parameter α is assumed to be. By contrast, λ_0^{Bb} is decreasing in the war risk, and, unlike the parameter β in Fehr and Schmidt (1999), it may be negative. For sufficiently low war risks ρ , however, it is positive, in agreement with Fehr's and Schmidt's assumption about their parameter β . We finally note that our evolutionary model suggests that a morality term be added, a term absent in Fehr's and Schmidt's model.

4.3 Scenario C

Let the attractiveness of the strategy in the fitness function (equation (3)) used by an individual be either a constant or exponentially increasing with the strategy's payoff (in its island),

$$f(\pi_i) = f_0 \cdot \exp(\delta_f \cdot \pi_i) \quad (31)$$

for some $\delta_f \geq 0$, and let a child's loyalty be independent of its parent's strategy: $s(\pi_i) = s_0$, for some $0 < s_0 < 1$. Then

$$r_0^C = \frac{(1 - m) [2s_0 + (1 - m)(1 - s_0)]}{n(1 + s_0) - (1 - m)(n - 1) [2s_0 + (1 - m)(1 - s_0)]}, \quad (32)$$

and

$$\lambda_0^C = \frac{(n - 1)(1 - m)}{n - (1 - m)}, \quad (33)$$

so that

$$\kappa_0^C = -\frac{(1-m)(1-s_0)}{2n-1+s_0-m(n-1)(1-s_0)}. \quad (34)$$

Like in Scenario A, both pairwise relatedness, r_0^C , and local competitiveness, λ_0^C , are non-negative. Furthermore, r_0^C and λ_0^C are both decreasing in m . Comparing further genetic and cultural transmission, we note that pairwise relatedness is identical in these two scenarios when the background survival rate, s_0 , and background loyalty, s_0 , both are zero. We then have

$$r_0^C = r_0^A = \frac{(1-m)^2}{(1-m)^2 + nm(2-m)}. \quad (35)$$

Hence, in important special cases, the mathematical properties of genetic and cultural relatedness are identical. By contrast, however, local competitiveness is stronger under cultural evolution than under genetic evolution when $s_0 = 0$, since then, $\lambda_0^C > \lambda_0^A$ whenever $m > 0$. This difference can be seen in equations (1) and (3). For when $s_0 = 0$, the m -terms are identical in the two scenarios, while the $(1-m)$ -terms differ. Under cultural evolution, the latter term compares the individual's attractiveness to that of the average attractiveness in the island, while under genetic evolution, this term compares the individual's fecundity with a weighted average of the average fecundity in the island and in a representative island. Hence, fitness is more sensitive to neighbor's payoffs under cultural evolution. Comparing λ_0^A to λ_0^C , this effect appears through the absence of the factor $(1-m)$ in λ_0^C . It turns out that in the present scenario, local competitiveness is always strong enough to outweigh the effect of relatedness.

To illustrate this, Figure 3 shows that κ_0^C is strictly negative for all $m < 1$, for different loyalty rates and different island sizes: for $s_0 = 0$ and $n = 2$ (the pink curve), $s_0 = 0.4$ and $n = 2$ (the orange curve), $s_0 = 0.8$ and $n = 2$ (the blue curve), $s_0 = 0$ and $n = 10$ (the pink dashed curve), $s_0 = 0.4$ and $n = 10$ (the orange dashed curve), $s_0 = 0.8$ and $n = 10$ (the blue dashed curve).

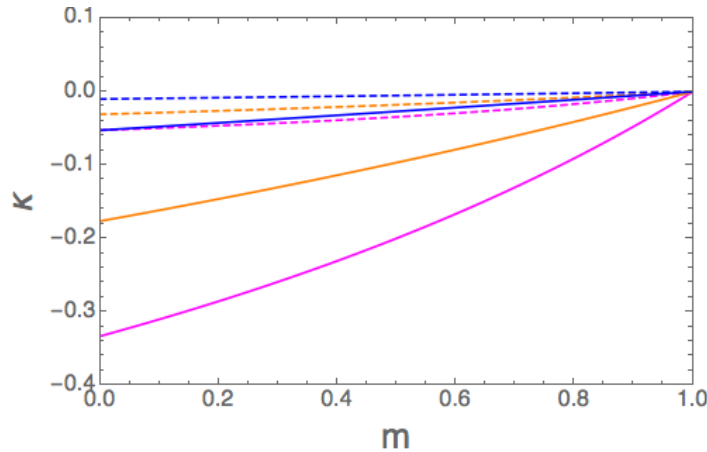


Figure 3: The value of κ_0^C as a function of the cultural import rate m , for different degrees of background loyalty of offspring towards parents.

In this example cultural transmission thus leads to anti-sociality, and anti-sociality is worse

for low values of m ; this is because a low cultural import rate enhances local competitiveness. Note that although genetic and cultural transmission here lead to opposite predictions regarding sociality, one qualitative similarity that appears is that survival under genetic transmission has a qualitatively similar effect as loyalty under cultural transmission: they both have a positive effect on κ_0 .

4.4 A public goods example

The above results apply to many strategic interactions studied in economics. We illustrate this with a canonical public goods game. Consider a population in which individuals in each island make contributions $x_i \geq 0$ to the provision of some public good, where

$$\pi(x_i, \mathbf{x}_{-i}) = F\left(\sum_{j=1}^n x_j\right) - \psi(x_i). \quad (36)$$

For twice differentiable benefit and cost functions F and ψ with $F', \psi' > 0$, Proposition 3 implies that an uninvadable strategy $\hat{x} > 0$ must satisfy the first-order condition

$$\frac{\psi'(\hat{x})}{F'(n\hat{x})} = 1 + (n-1)\kappa_0.$$

In the special case when F and ψ are power functions, with F having power $1/2$ and ψ having power 2 , this equation is also sufficient for uninvadability, and it gives

$$\hat{x} = \left(\frac{[1 + \kappa_0(n-1)]}{4\sqrt{n}}\right)^{2/3}. \quad (37)$$

We note that the uninvadable contribution is, not surprisingly, increasing in κ_0 . However, its dependence on group size, n , is complex. For any given value of $\kappa_0 \in (0, 1)$, \hat{x} is decreasing in n for $n < (1 - \kappa_0)/\kappa_0$ and increasing in n for $n > (1 - \kappa_0)/\kappa_0$. In particular, \hat{x} is monotonically increasing in n (for all $n \geq 2$), if and only if $\kappa_0 \geq 1/3$. For such values of κ_0 , individuals act sufficiently pro-socially to increase their contributions when more people benefit from the public good.

5 Uninvadable preferences under weak selection

In the above analysis, the transmitted traits are strategies, that is, rules of behavior for a given strategic interaction. This assumption is in line with standard evolutionary game theory, and, moreover, it allows to distill how population structure affects anti- and pro-sociality. However, it is at variance with classical game theory, which assumes that players are fully rational, know the set of strategies available to them, and have well-defined goal functions, which we henceforth will call *utility functions* (to distinguish them from material payoff functions). The standard assumption is that each player strives to maximize the mathematical expectation of his or her utility function, and the benchmark for predicting

outcomes is Nash equilibrium. The approach to let utility functions (or preferences), rather than strategies, be subject to evolutionary selection is usually referred to as the ‘indirect evolutionary approach’, pioneered by Güth and Yaari (1992), see also Heifetz, Shannon, and Spiegel (2007 a,b), Dekel, Ely, and Yilankaya (2007), and Alger and Weibull (2013, 2016), and references in these papers. Each participant in a strategic interaction is then assumed to act in his or her subjective self-interest, that is, to choose a strategy that maximizes his or her expected utility when interacting with the others. In this section we analyze preference evolution in our model.

Specifically, let Θ be the set of all continuous aggregative functions $u : X^n \rightarrow \mathbb{R}$.¹⁵ To define uninvadability of utility functions in this class, suppose that initially everybody in the population has the same utility function $u \in \Theta$, and that suddenly one individual, a *mutant*, switches to another utility function $v \in \Theta$. We focus on the case when each individual’s utility function, or *type*, is his or her private information. Then an individual’s behavior cannot be conditioned on the types of those with whom (s)he has been matched. However, individual behavior may be adapted to the population state at hand—that is, the current type distribution in the population. What matters for an individual’s choice of strategy is his or her matching profile, that is, the probability distribution of residents and mutants in her own island. Let $\mathbf{q} = (q_0, q_1, \dots, q_{n-1}) \in \Delta$ be a resident’s matching profile, where q_k is the probability that there are (precisely) k mutants in any given resident’s island. Likewise, let $\mathbf{p} = (p_0, p_1, \dots, p_{n-1}) \in \Delta$ be the mutants’ matching profile, where p_k is the probability that there are (precisely) k *other* mutants in any given mutant’s island. In the following definition, $\hat{\mathbf{y}}^{(j)}$ is a j -dimensional vector whose components all equal \hat{y} , and $\hat{\mathbf{x}}^{(j)}$ a j -dimensional vector whose components all equal \hat{x} (for $j = 1, \dots, n - 1$).

Given a pair of utility functions $(u, v) \in \Theta^2$, a (type-homogenous Bayesian) *Nash equilibrium* under the matching profile $\langle \mathbf{q}, \mathbf{p} \rangle$ is a pair of strategies, one for each preference type, such that each strategy is a best reply for any player of that type, given the matching probabilities $\langle \mathbf{q}, \mathbf{p} \rangle$:

$$\begin{cases} \hat{x} \in \arg \max_{x \in X} \sum_{k=0}^{n-1} q_k \cdot u(x, \hat{\mathbf{y}}^{(k)}, \hat{\mathbf{x}}^{(n-k-1)}) \\ \hat{y} \in \arg \max_{y \in X} \sum_{k=0}^{n-1} p_k \cdot v(y, \hat{\mathbf{y}}^{(k)}, \hat{\mathbf{x}}^{(n-k-1)}) \end{cases} \quad (38)$$

In particular, if $q_k = 0$ for all $k > 0$, then the first equation in (38) defines symmetric Nash equilibrium play among n individuals of the resident type u (and is hence independent of the mutant type).

We proceed to formally define uninvadability as a property of utility functions. Informally, a utility function will be called *uninvadable* if any other (mutant) utility function, which initially appears in a single individual, is bound to disappear from the population. Since the population is infinite while any mutant’s local lineage population is finite, it is sufficient to consider Nash equilibria under the matching profile $\langle \mathbf{q}^0, \mathbf{p}^0 \rangle$, where \mathbf{q}^0 is the unit vector $(1, 0, 0, \dots, 0)$ and \mathbf{p}^0 is the vector of mutants’ matching probabilities in the neutral process. For any given pair of utility functions $(u, v) \in \Theta^2$, let $B_{\text{NE}}^0(u, v) \subseteq X^2$ denote the set of

¹⁵By *aggregative* we mean that for any $x_i \in X$ and $\mathbf{x}_{-i} \in X^{n-1}$, and any bijection $h : \{2, 3, \dots, n\} \rightarrow \{2, 3, \dots, n\}$: $u(x_i, x_{h(2)}, x_{h(3)}, \dots, x_{h(n)}) = u(x_i, \mathbf{x}_{-i})$.

(type-homogenous Bayesian) Nash equilibria under matching profile $\langle \mathbf{q}^0, \mathbf{p}^0 \rangle$.¹⁶ In force of Proposition 1, according to which a strategy x is uninvadable by a strategy y if and only if $W(y, x) \leq 1$, we call a utility function $u \in \Theta$ *uninvadable* if $W(\hat{y}, \hat{x}) \leq 1$ for all utility functions $v \in \Theta$ and all $(\hat{x}, \hat{y}) \in B_{\text{NE}}^0(u, v)$. Any utility function that is not uninvadable will be called *invadable*.

The above definition does not presume that mutants know the preferences of residents. To a mutant, residents' preferences matter only in so far as they influence what strategies residents use. Our definition uses Nash equilibrium as a testing ground. It requires that if the residents were playing some Nash equilibrium among themselves (thus having mutually adjusted their individual behaviors to best replies, given their preferences), then there will not exist any mutant who could enter the population and earn a higher expected material payoff than the residents by acting optimally according to his or her own preferences.

The following result is an immediate consequence of our material-payoff characterization of uninvadability of strategies under weak selection, Proposition 2:

Corollary 2 *A utility function $u \in \Theta$ is uninvadable under weak selection if and only if $\Pi(\hat{y}, \hat{x}) \leq \Pi(\hat{x}, \hat{x})$ for all utility functions $v \in \Theta$ and all $(\hat{x}, \hat{y}) \in B_{\text{NE}}^0(u, v)$.*

For any given type $u \in \Theta$, let X^u be its set of symmetric Nash equilibrium strategies,

$$X(u) = \left\{ \hat{x} \in X : \hat{x} \in \arg \max_{x \in X} u(x, \hat{\mathbf{x}}^{(n-1)}) \right\}.$$

We are now in a position to state and prove our main result for preference evolution:

Theorem 1 *The utility function u^0 is uninvadable under weak selection. A utility function $u \in \Theta$ is invadable under weak selection if there exists a $\hat{x} \in X(u)$ such that $\hat{x} \notin X(u^0)$.*

In other words, selection favors competitive moralists with the utility function u^0 , defined in (13) and specified in scenarios A-C. Moreover, all other utility functions, unless they have an identical best reply to some resident equilibrium, are selected against.¹⁷

¹⁶By Berge's maximum theorem and Kakutani's fixed-point theorem, this set is nonempty if (i) the strategy set X is non-empty, finite-dimensional, compact and convex, (ii) the function u is continuous, (iii) the first maximand in (38) is quasi-concave in $x \in X$ and the second maximand in (38) is quasi-concave in $y \in X$. For generalizations to infinite-dimensional spaces, see e.g. Aliprantis and Border (2006).

¹⁷Theorem 1 enables us to recover the characterization in Proposition 2 of uninvadable strategies by way of a suitable restriction of the type space Θ . To see this, let $\Theta^* \subset \Theta$ be the subset of continuous functions of the special form $u_{\hat{x}}(x, \mathbf{y}) \equiv -\|x - \hat{x}\|$, for any given $\hat{x} \in X$. Individuals endowed with any such utility function will always play his dominant strategy, \hat{x} , irrespective of what strategies other individuals are using. It follows from the above that, under weak selection, a utility function $u_{\hat{x}} \in \Theta^*$ is uninvadable by utility functions in Θ^* if and only if \hat{x} is an uninvadable strategy. We note that this connection between preference evolution and strategy evolution relies on the assumption that interactions take place under incomplete information. If interacting individuals would know each others' types, or observe signals about each others' types, then types would also become commitment devices, which may give "bully" types extra help to survive over "mild" types. An important but big topic that falls outside the scope of this study.

6 Related literature

Several strands of literature have proposed answers to the question: in strategic interactions which strategies should evolution by natural selection give rise to? Here we summarize these strands of literature, and relate them to our model.

6.1 Strategy evolution in group-structured populations

Our model is based on the *island model*, originally due to Wright (1931). In evolutionary biology this has become a workhorse model to understand the conditions favoring pro- and anti-sociality at the level of fitness in spatially structured populations; this literature often refers to pro-sociality as “helping”, and to anti-sociality as “harming”. The early literature considered traits affecting environmentally induced group extinction (Eshel, 1972, Aoki, 1982); this is reminiscent of our Scenario B, but with the probability of surviving a shock depending only on the behaviors in the focal group.¹⁸ This research produced the insight that limited migration tends to favor pro-sociality. Later work showed that when traits impact individual fecundity, population structure tends to strengthen local competition, ultimately leading to a situation where local competition cancels the effect of relatedness (Wilson, Pollock, and Dugatkin, 1992, Taylor 1992a). These results spurred an extensive theoretical literature seeking to delineate how the assumptions pertaining to demography, life-history, the environment, and the modes of transmission, tip the balance in favor of pro- or anti-sociality at the survival or fecundity level, and which cover our three scenarios (e.g., Taylor and Irwin, 2000, Gardner and West, 2006, Johnstone and Cant, 2008, Lehmann, Foster, and Feldman, 2008, Lion and Gandon, 2010, Bao and Wild, 2012, and Alberto et al. 2017, for a few relevant examples). In such models with differentiable fitness functions, under any strength of selection one generally obtains a first-order condition at the vital rate level which is similar to the equation that we derive at the level of material payoffs under weak selection (our equation (20)), but with material payoff being replaced by survival or fecundity. In this first-order condition a *scaled relatedness* coefficient aggregates all the consequences of spatial structure. This work is reviewed by Lehmann and Rousset (2010); see also Van Cleve (2015) and Dos-Santos and Peña (2017).

Since we assumed that vital rates are affine functions of material payoffs, in our setup scaled relatedness is represented by the marginal rate of substitution, κ_0 .¹⁹ By contrast to the previous literature, our derivation of the first-order condition at the level of material

¹⁸The analytical models of Bowles (2006, 2009) for the evolution of “parochial altruism” are also close to our scenario B; in particular, the expected number of groups $[1 - \rho + 2\rho\nu(\pi, \bar{\pi}^*)]$ to which a focal group has access for reproduction after warfare also appears in Bowles’s formalization. However, since in his model there are no explicit assumptions that allow to close the lifecycle, it is impossible to derive the explicit values of λ_0 , r_0 , and κ_0 for his model.

¹⁹Some earlier results concerning scaled relatedness, which were derived in the special case of a public goods game, are confirmed in our setting. In particular, for Scenario A, we recover the pro-social scaled relatedness found by Taylor and Irwin (2000, equation A-10), and for Scenario C with $l_0 = 0$ we recover the anti-social scaled relatedness found by Lehmann, Foster, and Feldman (2008, equation (26)).

payoffs (equation (19)) allows us to have a single substitution rate, even when strategies affect multiple vital rates. A common feature of the biological literature is also that scaled relatedness is expressed directly in terms of specifics of the transmission process (e.g., migration rate, group size, probability of war, cultural loyalty). By contrast, we decompose scaled relatedness for constant demography, by showing exactly how it depends on pairwise relatedness on the one hand, and local competitiveness on the other hand. This contribution is valuable for two reasons. First, it allows to disentangle the pathways that affect pro- and anti-sociality. Second, in the uninvadable goal function this decomposition shows how pairwise relatedness and local competitiveness contribute to the weights attached to self-interest (own material well-being), altruism or spite (others' material well-being), and moral concerns (the material well-being that would arise should others choose the same strategy).

This study extends ends our previous work, Lehmann, Alger, and Weibull (2015), in which we used the island model to study strategy evolution in social interactions. We derived necessary and sufficient conditions for uninvadable strategies to constitute symmetric Nash equilibria under two utility functions that are weighted averages of the fitnesses, as in (4). We also studied weak selection and arrived at a result similar to the present Corollary 1. We allowed neither for wars between islands, nor for cultural transmission. Furthermore, we did not study the differentiable case and thus did not analyze whether the resulting behavior is, on balance, pro- or anti-social, and we did not derive expressions for the coefficient of competitiveness, λ_0 . Finally, we did not study preference evolution.

In sum, our current approach is consistent with, and extends previous formalizations of strategy evolution in the island model. Furthermore, with our three scenarios we capture qualitatively and quantitatively canonical outcomes on the pro- and anti-sociality spectrum, from strong pro-sociality (when competition is mainly between groups) to strong anti-sociality (when competition is mainly within groups).

6.2 Preference evolution

Since economists view individuals as choosing strategies based on their preferences, it is not surprising that behavioral economists interested in how evolution may have shaped human behavior sought to model the evolution of preferences rather than of strategies. The literature on preference evolution in social interactions dates back to the ground-breaking study by Güth and Yaari (1992), who extended the concept of evolutionary stability from strategies to parametrized classes of *utility functions*;²⁰ this approach is also referred to as “indirect evolution.”²¹ While in many such models interactions are assumed to take place under complete information,²² our model falls into the part of the literature which studies

²⁰See also Frank (1987).

²¹There is also an extensive literature on preference evolution in decision problems. See, e.g., Robson (1996), Rayo and Becker (2007), and Robson and Samuelson (2011a,b).

²²See Güth and Yaari (1992), Ockenfels (1993), Huck and Oechssler (1996), Ellingsen (1997), Bester and Güth (1998), Fershtman and Judd (1987), Fershtman and Weiss (1998), Koçkesen, Ok and Sethi (2000a,b),

interactions under incomplete information.²³ In this literature, the work most closely related to ours is that by Alger and Weibull (2013, 2016), who allow for assortative matching, i.e., the possibility that the probability of being matched with a mutant differs for residents and for mutants. They do not explicitly model the population structure or the population dynamics, however. Instead, assortativity is modeled as an abstract function that maps the distribution of traits in the population to probabilities governing the matching of interacting individuals.²⁴

Our model makes two contributions to this literature. First, we explicitly model the population structure that gives rise to assortativity. Population structure is shown to imply that relatedness—which in our formalization is given by the probability that individuals who share a common ancestor get to interact more—must go hand in hand with local competitiveness, a force which does not appear in Alger and Weibull (2013, 2016). It further implies that relatedness and local competitiveness can be formally traced back to vital rates such as migration rates, survival probabilities, war probabilities, loyalty to parental values, etc.

Second, by explicitly taking into account the dynamics, the transmission process, our model allows to establish a clear distinction between fitness and material payoff. In particular, under weak selection our model thus delivers predictions regarding the shape of unavoidable utility functions, showing that it consists of a combination of self-interest, morality and other-regard. The relative importance of these motives depend on the specifics of the transmission process. For complete migration and no local competition ($r_0 = 0$ and $\lambda_0 = 0$), we recover the result that evolution favors self-interested individuals (e.g., Dekel et al., 2007). By contrast, under no migration and no local competition ($r_0 = 1$ and $\lambda_0 = 0$, which obtains when no individual migrates single-handedly, irrespective of if there are wars or not, since in the latter case it is as if whole groups migrate, but do not have local competitors), the unavoidable goal function is full morality, which corresponds to the purely Kantian *Homo moralis* (Alger and Weibull, 2013, 2016).²⁵

Furthermore, when vital rates are differentiable, our Proposition 3 shows that the index of assortativity used in the literature on the evolutionary stability of strategies and preferences (Bergstrom, 2003, Alger and Weibull, 2013, 2016) can be interpreted as the parameter κ_0 of

Bolle (2000), Possajennikov (2000), Sethi and Somanathan (2001), Heifetz, Shannon and Spiegel (2007a,b), Alger (2010), Alger and Weibull (2010, 2012), and Wu (2017). Biologists have also begun studying preference evolution under complete information and parameterized utilities; e.g., see notably Akçay and van Cleve (2013), which applies to the island model.

²³See Ok and Vega-Redondo (2001), and Dekel, Ely and Yilankaya (2007).

²⁴This formalization of assortativity, which goes back to Hamilton (1971), was initially used in economics by Bergstrom (1995, 2003), who focused on strategy evolution; see also Bowles and Gintis (1998).

²⁵Interestingly, the same result obtains under complete information under uniform random matching (see Dekel, 2007, Proposition 2). This shows that a Pareto efficient strategy profile (in terms of material payoffs) obtains either under perfect assortative matching, a special case of our model, or under preference evolution under complete information. However, the latter case is likely to be evolutionarily unstable in a wide range of strategic interactions as it is prone to mimickry, that is, dishonest signaling of one's type. This is a huge and important but complex topic that warrants further investigation.

the present model, and is hence driven both by relatedness and by local competition. Indeed, in that case, the first-order condition in Proposition 3 can be written

$$\frac{d}{dy} [(1 - \kappa_0) \cdot \pi(y, (\hat{x}, \hat{x}, \dots, \hat{x})) + \kappa_0 \cdot \pi(y, (y, y, \dots, y))]_{y=\hat{x}} = 0. \quad (39)$$

In words, any interior uninvadable strategy \hat{x} satisfies the same first-order condition as in any symmetric Nash equilibrium of a game in which every player's utility (technically, payoff) function is a certain convex combination of own material payoff and the material payoff that the player would obtain should all the others always use the same strategy as the individual herself. The weight given to the second, moral goal is $\kappa_0 \in [-1, 1]$, a parameter determined by population structure and the transmission scenario. In particular, it is independent of the material payoff function in the strategic interaction in question.

6.3 Economics models of cultural transmission

In the economics literature on cultural transmission of traits, a commonly used model is that of Bisin and Verdier (2001).²⁶ Here we compare our model to theirs.

Like in our model, in Bisin and Verdier (2001) each grown-up has exactly one child, and each child inherits its parent's trait with some probability, and otherwise it inherits the trait of another grown-up in the population. By contrast to our model, the population is not structured into islands, and there is no strategic interaction between individuals. Furthermore, in their model a parent cares about whether her child has the same trait as her, but not about whether the child inherited this trait from the parent or from someone else. For comparison with our model we thus need to derive the probability that a child will adopt the same trait, θ , as its parent, irrespective of whether this trait is emulated from the parent or another grown-up.

Let $q_{\theta g}$ be the share of θ -individuals on island g , and let q_θ be the share of θ -individuals in the population at large. Let $a_{\theta g}$ be the attractiveness of individuals with trait θ on island g , and let \bar{a}_θ be the average attractiveness of individuals with trait θ on all islands. Likewise, for any trait θ let $l_{\theta g}$ and \bar{l}_θ be the loyalty of individuals on island g and on average, respectively. Then the probability that the child of a randomly drawn adult of trait θ will adopt the same trait as its parent (whether from the parent or from someone else) is:

$$\begin{aligned} P_{\theta\theta} &= (1 - m) \cdot \bar{l}_\theta + m \cdot \frac{\bar{a}_\theta}{\bar{a}} \cdot q_\theta + (1 - m) \cdot (1 - \bar{l}) \cdot \frac{\bar{a}_\theta}{\bar{a}} \cdot q_\theta \\ &= (1 - m) \bar{l}_\theta + [m + (1 - m)(1 - \bar{l})] \frac{\bar{a}_\theta}{\bar{a}} \cdot q_\theta \end{aligned} \quad (40)$$

The first term represents the average probability that the child adopts his trait directly from his parent. The second term is the probability, on average, that the child adopts trait θ upon sampling individuals from the previous generation outside of his natal island. These

²⁶Other models include Hauk and Saez-Marti (2002) and Lindbeck and Nyberg (2006). However, for the sake of space, we do not provide a detailed comparison to these models here.

individuals being drawn randomly from the population at large, on average the share of individuals with trait θ in any such sample is q_θ , and each such individual has relative attractiveness \bar{a}_θ/\bar{a} . The last term is the average probability that a child who does not directly emulate the trait θ of his parent, but who does emulate his trait from someone in his natal island (including his parent as a potential target), ends up with the same trait as his parent. Since this probability may vary between islands, it is necessary to take the average over islands. The corresponding formula in Bisin and Verdier (2001) is given in their equation (1) as

$$P_{\theta\theta} = l_\theta(q_\theta) + [1 - l_\theta(q_\theta)] \cdot q_\theta, \quad (41)$$

where $l_\theta(q_\theta)$ is the probability for direct vertical transmission of a parent's trait to its (unique) child.²⁷ If we, in our model, abstract away the island structure, so that the whole population is treated like one large island, then there is no emulation from outside the island ($m = 0$) and island averages are population averages, in which case (40) boils down to

$$P_{\theta\theta} = l_\theta + (1 - l_\theta) \frac{\bar{a}_\theta}{\bar{a}} \cdot q_\theta, \quad (42)$$

which shows the similarity and the difference between the two models in terms of the probability that a trait is transmitted from parent to child. Crucially, however, the key difference between our model and that of Bisin and Verdier (2001) is that we derive the consequences of the said transmission probabilities for the evolution of preferences in social interactions, where the preferences can be any continuous function, while they analyze the dynamics of two given traits in a population in which parents actively try to influence their children's traits.

6.4 Evolutionarily stable strategies

To clarify the difference between an uninvadable strategy and an evolutionarily stable strategy (Maynard Smith and Price, 1973; for a textbook treatment, see Weibull, 1995), consider the special case of our three transmission scenarios when $n = 2$ and $m = 1$. Then $p_0(y, x) = 1$, that is, it is as if matching is uniformly random. Proposition 1 consequently implies that $x \in X$ is uninvadable if and only if $w(x, x) \geq w(y, x)$. We recall that a strategy is evolutionarily stable against another strategy if users of the former obtain a higher average material payoff than users of the latter, for all sufficiently small positive population shares using the latter strategy. Formally, a (resident) strategy $x \in X$ is evolutionarily stable if for each mutant strategy $y \neq x$ there exists an $\bar{\varepsilon}_y > 0$ such that for all population shares $\varepsilon \in (0, \bar{\varepsilon}_y)$ of users of the mutant strategy y , the following inequality holds:

$$(1 - \varepsilon) \cdot w(x, x) + \varepsilon \cdot w(x, y) > (1 - \varepsilon) \cdot w(y, x) + \varepsilon \cdot w(y, y). \quad (43)$$

By continuity of both sides of this inequality, a sufficient condition for x to be ESS is $w(x, x) > w(y, x)$. If $w(x, x) = w(y, x)$, it is also necessary that $w(x, y) > w(y, y)$. The fact that for uninvadability the weak inequality is sufficient stems from the fact that with

²⁷Instead of traits θ they write traits i , instead of q_θ they write q^i , and instead of $t_\theta(q_\theta)$ they write $d^i(q^i)$.

a finite number of mutants in an infinitely large population, when $m = 1$ the probability that a mutant is matched with another mutant is nil. By contrast, an evolutionarily stable strategy must resist the invasion by a positive mass of mutants, and hence, for any $\varepsilon > 0$ mutants face a positive probability of being matched together.²⁸

7 Conclusions

Going back to the advent of life on Earth, we all have a huge number of ancestors, each of whom managed to survive until (s)he successfully reproduced. Our recent ancestors most likely lived in small groups of hunter-gatherers (probably ranging from 5 to 150 grown-ups), extending beyond the nuclear family, for more than two millions of years (Grueter, Chapais, and Zinner, 2012, Malone, Fuentes, and White, 2012, van Schaik, 2016, Layton et al., 2012). This is part of the environment of evolutionary adaptation of the human lineage (e.g., van Schaik, 2016). Heritable behavioral traits should thus be a reflection of the ability to successfully survive and reproduce in such a group-structured environment, in which individuals' behavior certainly had effects on other group members. Our model represents the first attempt in economics to formally analyze how the migration patterns between groups, and the competition for resources within and between groups, may impact the qualitative nature of the preferences that evolutionary forces should select for. By applying a classic model developed by evolutionary biologists, we derive such predictions with the purpose of generating valuable insights for economists.

Our analysis shows that if material payoffs impact individual fitness only marginally (weak selection), evolution favors individuals who care not only about own material payoff. More specifically, evolution favors *competitive moralists*, who display a combination of material self-interest, a Kantian moral concern, and a (positive or negative) concern for others' payoffs (i.e., altruism or spite). The moral concern is driven by kin selection, that is, the fact that in group-structured populations interactions occur between individuals who are more likely to share a common ancestor (be it cultural or genetic) who lived in their group than individuals sampled randomly from the population. The concern for neighbors' material payoffs is driven by two opposing forces. On the one hand, a negative concern due to local competition, that is, the reproductive benefit from outcompeting neighbors, whose offspring compete locally with own offspring. On the other hand, a positive concern due to global competition, that is, the reproductive benefit from one's group's ability to win conflicts with other groups (or resist exogenous collective shocks, such as natural disaster).

When groups consist of two individuals ($n = 2$), or when $n \geq 2$, fitness functions are differentiable, and selection is weak, only two parameters are needed for uninvadable utility functions to be identified: the *coefficient of pairwise relatedness* and the *coefficient of local*

²⁸This makes the implicit assumption that evolution can operate in each demographic time period to increase the relative frequency of traits with increased fitness, a feature that, owing to recombination, is not guaranteed to hold under realistic genetic underpinning of traits, unless selection is weak (e.g., Eshel, Feldman, and Bergman 1998 for a useful discussion). Hence, the concept of an evolutionarily stable strategy has been supplanted by that of an uninvadable strategy in population biology.

competitiveness. Because both relatedness and local competition arises as soon as individuals have any propensity to stay in their natal group, our analysis implies that the selfish *Homo oeconomicus* survives evolutionary pressure only in the highly unrealistic scenario in which *all* individuals always leave their natal group. Thus, under any realistic transmission scenario and population structure, and when individuals do not know each others' preferences, evolution leads to social preferences that include (a) own material payoff, (b) a form of Kantian morality similar to the one expressed by *Homo moralis* preferences (Alger and Weibull, 2013, 2016), and (c) a comparison between own and other group members' material payoffs, akin to altruism/spite (Becker, 1976) or inequity aversion (Fehr and Schmidt, 1999).

Our analysis further reveals whether, on balance, a competitive moralist behaves pro-socially or anti-socially at the level of material payoffs, in the sense that equilibrium strategies enhance or reduce others' material payoffs. We show that, on balance, pro-sociality obtains if the coefficient of pairwise relatedness exceeds the coefficient of local competitiveness, and anti-socially if the reverse is true. Whether the former or the latter is true depends on the specifics of the transmission process, as illustrated by three scenarios. Under the vertical transmission scenario (**A**), the coefficient of local competitiveness is always positive; an individual's survival probability and fecundity benefit from a reduction in the neighbors' material payoffs. On balance, however, anti-sociality never obtains, and pro-sociality obtains as soon as individuals may survive to get more than one batch of offspring; longevity enhances relatedness. The second scenario (**B**) adds inter-group conflicts to the baseline scenario. We show that if a group's probability of winning a war depends positively on both group members' material payoffs, then this adds an altruistic component to the uninventable utility function. In a war with another group, individual fitness increases both with one's own material payoff and with that of one's neighbor. Hence, it may be fitness-enhancing to exhibit altruism towards one's neighbor, on top of Kantian morality. If the risk of war is high enough, this pro-social effect may dominate the counter-social effect induced by local competition, and the net result may be altruism. In such transmission scenarios, a combination of self-interest, altruism, and morality obtains. Finally, in the cultural transmission scenario (**C**) we consider, behavior is, on balance, always anti-social. In a nutshell, under cultural transmission competition is fiercer than under genetic transmission, and competitiveness then always outweighs relatedness.

We have explored these three transmission scenarios in detail, but a host of other scenarios could be considered. For instance, it would be interesting to model explicit age-structure within groups, and intergenerational transfer of resources in order to understand the evolution of social and time preferences in a cooperative breeding context. This could be done by letting the expected number of offspring that survive to compete for an adult's spot depend on the material payoffs of several adults in the group (Weibull and Salomonsson, 2006), like in our war scenario. Furthermore, it would also be interesting to lift the assumptions of asexual reproduction, equal and fixed group size, the inability of individuals to learn about neighbors' types, and of fixed migration to determine whether this will qualitatively affect the components of uninventable preferences. These and other scenarios may then allow to understand the impact of population structures throughout human history on the evolution of behaviors and preferences.

8 Appendix

8.1 Proof of Proposition 2

We start by evaluating personal fitness under weak selection. We write the partial derivatives of the fitness function $w : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$, which maps payoff profiles $(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*)$ to i 's fitness, by indexing with respect to the argument in question. By using a first-order Taylor expansion of the fitness of an arbitrary individual i in an arbitrary island with respect to δ , evaluated at $\delta = 0$, we obtain

$$w(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*) = 1 + \delta \cdot \left[\beta \cdot (\pi_i - \bar{\pi}^*) - \frac{\gamma}{n-1} \sum_{j \neq i} (\pi_j - \bar{\pi}^*) \right] + \mathcal{O}(\delta^2), \quad (44)$$

where $\beta = w_1(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*)|_{\delta=0}$ and

$$\gamma = - \sum_{k=2}^n w_k(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*)|_{\delta=0} = (n-1) \cdot w_n(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*)|_{\delta=0} \quad (45)$$

are coefficients that depend on structural demographic parameters.

For example, in Scenario B, where

$$w(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*) = \left[1 - \rho + 2\rho \frac{1 + \delta \sum_{j=1}^n \pi_j/n}{2 + \delta \left(\sum_{j=1}^n \pi_j/n + \bar{\pi}^* \right)} \right] \cdot \left[\frac{n(1-m)(1 + \delta\pi_i)}{(1-m) \sum_{j=1}^n (1 + \delta\pi_j) + mn(1 + \delta\bar{\pi}^*)} + \frac{m(1 + \delta\pi_i)}{(1 + \delta\bar{\pi}^*)} \right], \quad (46)$$

straight-forward calculations (see below) yield

$$\beta = 1 + \frac{\rho}{2n} - \frac{(1-m)^2}{n} \quad (47)$$

and

$$\gamma = \frac{n-1}{n} \left[(1-m)^2 - \frac{\rho}{2} \right]. \quad (48)$$

We return to the general case, and note that the expression in (44) follows from four facts about $w(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*)$: (i) to the first order in δ , fitness is an affine function of the individual's payoff; (ii) all individuals $j \neq i$ on i 's island have the same effect on the fitness of individual i (by the assumed permutation invariance of payoffs); (iii) all individuals on other islands have the same effect on the fitness of individual i (they all play the same strategy, \hat{x}), and (iv) the total selective effects must sum to zero in a monomorphic population as the expected change in type number or frequency is necessarily nil; formally,

$$w_1(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*) + (n-1) w_n(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*) + w_{n+1}(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*) = 0, \quad (49)$$

where all derivatives are evaluated at $\delta = 0$.

Letting

$$\lambda_0 = \gamma/\beta \quad (50)$$

and factoring out $\beta > 0$ from (44), we obtain

$$w(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*) = 1 + \delta\beta \left[\pi_i - \lambda_0 \sum_{j \neq i} \frac{\pi_j}{n-1} - (1 - \lambda_0)\bar{\pi}^* \right] + \mathcal{O}(\delta^2). \quad (51)$$

This shows that λ_0 quantifies local competitiveness among patch members (Lehmann, Alger, and Weibull, 2015; see also Frank, 1998, and Gardner and West, 2004, for a description, but without a formal derivation, of λ_0).

We proceed to determining the implications of Assumption [M] for the bounds of $\lambda_0 = -(n-1) \cdot w_n(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*)/w_1(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*)$ (evaluated at $\delta = 0$); we focus on the non-trivial case $w_n(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*) \neq 0$. First, part (ii) of Assumption [M] implies $-(n-1) \leq \lambda_0$. Second, recalling (49) we obtain $\lambda_0 \leq 1$, with a strict inequality if either $w_{n+1}(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*) < 0$, or $w_{n+1}(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*) = 0$ and $n > 2$.

Fitness being a function of the material payoffs, which in turn are functions of the strategy profiles, we can write fitness directly as a function of the strategy profiles. Using \tilde{w} to denote this function, and letting \hat{x} denote the resident strategy:

$$\tilde{w}(x_i, \mathbf{x}_{-i}, \hat{x}) = 1 + \delta\beta \left[\pi_i(x_i, \mathbf{x}_{-i}, \hat{x}) - \lambda_0 \sum_{j \neq i} \frac{\pi(x_j, \mathbf{x}_{-j}, \hat{x})}{n-1} - (1 - \lambda_0)\bar{\pi}(\hat{x}) \right] + \mathcal{O}(\delta^2). \quad (52)$$

We now note that since strategies only affect material payoffs, vital rates (such as fecundity and survival) are the same for all strategies when $\delta = 0$. Hence, also fitness is strategy independent and thus equal to 1 when $\delta = 0$ (since the population by assumption is constant in size). The resulting evolutionary process is neutral (Crow and Kimura, 1970, Ewens, 2004, Gillespie, 2004, and, for an explicit example, Rousset, 2004), and hence we can write

$$p_k(y, \hat{x}) = p_k^0 + \mathcal{O}(\delta) \quad \forall k, \quad (53)$$

where $\mathcal{O}(\delta)$ is the deviation (relative to the neutral process) of the strategy-profile distribution induced by selection (i.e., $\delta > 0$) that is at most of order δ , where p_k^0 is strategy independent.

From equations (52) and (53), we obtain:

$$p_k(y, \hat{x}) \cdot \tilde{w}(y, (\mathbf{y}^{(k)}, \hat{\mathbf{x}})) = p_k(y, \hat{x}) + \delta\beta p_k^0 \cdot [\tilde{\pi}^{(k)}(y, x) - (1 - \lambda_0)\pi(\hat{\mathbf{x}})] + \mathcal{O}(\delta^2), \quad (54)$$

where $(\mathbf{y}^{(k)}, \hat{\mathbf{x}})$ is the $(n-1)$ -dimensional vector with k components equal to y and the remaining components equal to \hat{x} , and

$$\tilde{\pi}^{(k)}(y, \hat{x}) = \pi(y|k) - \lambda_0 \left[\frac{k}{n-1} \pi(y|k) + \frac{n-1-k}{n-1} \pi(\hat{x}|k) \right] \quad (55)$$

(see equation (8)).

Substituting this into lineage fitness (4) produces

$$\begin{aligned} W(y, \hat{x}) &= \sum_{k=0}^{n-1} p_k(y, \hat{x}) \cdot \tilde{w}(y, (\mathbf{y}^{(k)}, \hat{\mathbf{x}})) \\ &= 1 + \delta\beta \cdot \sum_{k=0}^{n-1} p_k^0 \cdot [\tilde{\pi}^{(k)}(y, \hat{x}) - (1 - \lambda_0)\pi(\hat{\mathbf{x}})] + \mathcal{O}(\delta^2). \end{aligned} \quad (56)$$

Expressed in words: to the first order in selection intensity δ , lineage fitness is the expectation of fitness taken over the neutral experienced-weighted lineage-size distribution, which is a standard result of evolutionary dynamics in population genetics [a machinery that applies both to finite and infinite populations, see Roze and Rousset (2003), Rousset (2004), Lehmann and Rousset (2009), and Lessard (2009)].

Recalling the definition of lineage payoff-advantage (see (7)),

$$\Pi(y, \hat{x}) = \sum_{k=0}^{n-1} p_k^0 \cdot \tilde{\pi}^{(k)}(y, \hat{x}), \quad (57)$$

and combining it with (54) we can write lineage fitness as

$$W(y, \hat{x}) = 1 + \delta\beta \cdot [\Pi(y, \hat{x}) - (1 - \lambda_0)\pi(\hat{\mathbf{x}})] + \mathcal{O}(\delta^2). \quad (58)$$

Neglecting higher order terms in δ in this equation allows us to write the condition for uninvadability [$W(y, \hat{x}) \leq W(\hat{x}, \hat{x})$ for all $y \in X$] for weak selection as $\Pi(y, \hat{x}) \leq \Pi(\hat{x}, \hat{x})$ for all $y \in X$, which implies that \hat{x} is uninvadable if and only if

$$\hat{x} \in \arg \max_{y \in X} \Pi(y, \hat{x}). \quad (59)$$

8.2 Proof of Proposition 3

We write the proof for the special case when $y \in X \subseteq \mathbb{R}$, but the proof holds for each component y_t of vector strategies $y \in X \subseteq \mathbb{R}^d$ for any $d \in \mathbb{N}$ and $t \in \{1, \dots, d\}$. Letting $(\mathbf{y}^{(k)}, \hat{\mathbf{x}})$ denote the $(n-1)$ -dimensional vector with $k=0, \dots, n-1$ components equal to y and the remaining components equal to \hat{x} , (20) can be written

$$[\pi_1(y, (\mathbf{y}^{(0)}, \hat{\mathbf{x}})) + \kappa_0(n-1) \cdot \pi_n(y, (\mathbf{y}^{(0)}, \hat{\mathbf{x}}))]_{|y=\hat{x}} = 0. \quad (60)$$

We proceed to prove that this first-order condition must hold for any interior uninvadable strategy \hat{x} . To begin, write

$$\begin{aligned} \Pi(y, \hat{x}) &= \sum_{k=0}^{n-1} p_k^0 \cdot \pi(y, (\mathbf{y}^{(k)}, \hat{\mathbf{x}})) \\ &\quad - \lambda_0 \sum_{k=0}^{n-1} p_k^0 \cdot \frac{k}{n-1} \cdot \pi(y, (\mathbf{y}^{(k)}, \hat{\mathbf{x}})) \\ &\quad - \lambda_0 \sum_{k=0}^{n-1} p_k^0 \cdot \frac{n-1-k}{n-1} \cdot \pi(\hat{x}, (\mathbf{y}^{(k+1)}, \hat{\mathbf{x}})). \end{aligned} \quad (61)$$

By virtue of Proposition 2, a necessary condition for a strategy \hat{x} to be uninvadable when π is differentiable is that $\partial\Pi(y, x)/\partial y$ vanishes at $y = \hat{x}$, or

$$\begin{aligned}
0 &= \sum_{k=0}^{n-1} \left[p_k^0 \sum_{j=1}^{k+1} \pi_j (y, (\mathbf{y}^{(k)}, \hat{\mathbf{x}})) \right]_{y=\hat{x}} \\
&\quad - \lambda_0 \cdot \sum_{k=0}^{n-1} \left[p_k^0 \cdot \frac{k}{n-1} \cdot \sum_{j=1}^{k+1} \pi_j (y, (\mathbf{y}^{(k)}, \hat{\mathbf{x}})) \right]_{y=\hat{x}} \\
&\quad - \lambda_0 \cdot \sum_{k=0}^{n-1} \left[p_k^0 \cdot \frac{n-1-k}{n-1} \cdot \sum_{j=2}^{k+2} \pi_j (\hat{x}, (\mathbf{y}^{(k+1)}, \hat{\mathbf{x}})) \right]_{y=\hat{x}}.
\end{aligned} \tag{62}$$

Thanks to permutation invariance, this can be written

$$\begin{aligned}
0 &= \sum_{k=0}^{n-1} p_k^0 \cdot \pi_1 (\hat{x}, (\mathbf{y}^{(0)}, \hat{\mathbf{x}})) + \sum_{k=0}^{n-1} [p_k^0 \cdot k \cdot \pi_n (\hat{x}, (\mathbf{y}^{(0)}, \hat{\mathbf{x}}))] \\
&\quad - \lambda_0 \cdot \sum_{k=0}^{n-1} \left[p_k^0 \cdot \frac{k}{n-1} \cdot \pi_1 (\hat{x}, (\mathbf{y}^{(0)}, \hat{\mathbf{x}})) \right] \\
&\quad - \lambda_0 \cdot \sum_{k=0}^{n-1} \left[p_k^0 \cdot \frac{k}{n-1} \cdot k \cdot \pi_n (\hat{x}, (\mathbf{y}^{(0)}, \hat{\mathbf{x}})) \right] \\
&\quad - \lambda_0 \cdot \sum_{k=0}^{n-1} \left[p_k^0 \cdot \frac{n-1-k}{n-1} \cdot (k+1) \cdot \pi_n (\hat{x}, (\mathbf{y}^{(0)}, \hat{\mathbf{x}})) \right].
\end{aligned} \tag{63}$$

Collecting terms, this becomes

$$\begin{aligned}
0 &= \pi_1 (\hat{x}, (\mathbf{y}^{(0)}, \hat{\mathbf{x}})) \cdot \sum_{k=0}^{n-1} p_k^0 \cdot \left(1 - \lambda_0 \cdot \frac{k}{n-1} \right) \\
&\quad + \pi_n (\hat{x}, (\mathbf{y}^{(0)}, \hat{\mathbf{x}})) \cdot \sum_{k=0}^{n-1} (p_k^0 \cdot k) \\
&\quad - \lambda_0 \cdot \pi_n (\hat{x}, (\mathbf{y}^{(0)}, \hat{\mathbf{x}})) \cdot \sum_{k=0}^{n-1} \left(p_k^0 \cdot \frac{k}{n-1} \cdot k \right) \\
&\quad - \lambda_0 \cdot \pi_n (\hat{x}, (\mathbf{y}^{(0)}, \hat{\mathbf{x}})) \cdot \sum_{k=0}^{n-1} \left[p_k^0 \cdot \frac{n-1-k}{n-1} \cdot (k+1) \right],
\end{aligned} \tag{64}$$

or

$$\begin{aligned}
0 &= \pi_1(\hat{x}, (\mathbf{y}^{(0)}, \hat{\mathbf{x}})) \cdot \sum_{k=0}^{n-1} p_k^0 \cdot \left(1 - \lambda_0 \cdot \frac{k}{n-1}\right) \\
&\quad + \pi_n(\hat{x}, (\mathbf{y}^{(0)}, \hat{\mathbf{x}})) \cdot \sum_{k=0}^{n-1} (p_k^0 \cdot k) \\
&\quad - \lambda_0 \cdot \pi_n(\hat{x}, (\mathbf{y}^{(0)}, \hat{\mathbf{x}})) \cdot \sum_{k=0}^{n-1} \left[p_k^0 \cdot \frac{(n-2)k + (n-1)}{n-1} \right].
\end{aligned} \tag{65}$$

Recalling that $\sum_{k=0}^{n-1} p_k^0 = 1$, and noting that from (10) we have

$$r_0 = \sum_{k=0}^{n-1} \frac{k}{n-1} \cdot p_k^0, \tag{66}$$

(65) becomes

$$\begin{aligned}
0 &= (1 - \lambda_0 r_0) \cdot \pi_1(\hat{x}, (\mathbf{y}^{(0)}, \hat{\mathbf{x}})) \\
&\quad + r_0 (n-1) \cdot \pi_n(\hat{x}, (\mathbf{y}^{(0)}, \hat{\mathbf{x}})) \\
&\quad - \lambda_0 r_0 (n-2) \cdot \pi_n(\hat{x}, (\mathbf{y}^{(0)}, \hat{\mathbf{x}})) \\
&\quad - \lambda_0 \cdot \pi_n(\hat{x}, (\mathbf{y}^{(0)}, \hat{\mathbf{x}})),
\end{aligned} \tag{67}$$

or

$$\begin{aligned}
0 &= \pi_1(\hat{x}, (\mathbf{y}^{(0)}, \hat{\mathbf{x}})) \\
&\quad + (n-1) \cdot \frac{r_0 - \lambda_0 [1/(n-1) + (n-2)r_0/(n-1)]}{1 - \lambda_0 r_0} \cdot \pi_n(\hat{x}, (\mathbf{y}^{(0)}, \hat{\mathbf{x}})),
\end{aligned} \tag{68}$$

or

$$0 = \pi_1(\hat{x}, (\mathbf{y}^{(0)}, \hat{\mathbf{x}})) + (n-1) \cdot \kappa_0 \cdot \pi_n(\hat{x}, (\mathbf{y}^{(0)}, \hat{\mathbf{x}})), \tag{69}$$

which is identical to (60).

Finally, we determine the bounds for κ_0 . First,

$$\begin{aligned}
\kappa_0 &\leq 1 \Leftrightarrow \frac{r_0 - \lambda_0 \left[\frac{1}{n-1} + r_0 \frac{n-2}{n-1} \right]}{1 - \lambda_0 r_0} \leq 1 \\
&\Leftrightarrow \frac{\lambda_0}{n-1} (r_0 - 1) \leq 1 - r_0 \\
&\Leftrightarrow \lambda_0 \geq -(n-1),
\end{aligned} \tag{70}$$

an inequality which was proved in the proof of Propostion 2. Second,

$$\kappa_0 \geq -1 \Leftrightarrow \lambda_0 \leq \frac{(n-1)(1+r_0)}{1+r_0(2n-3)}. \tag{71}$$

Now, $n \geq 2$ and $r_0 \leq 1$ together imply $\frac{(n-1)(1+r_0)}{1+r_0(2n-3)} \geq 1$. Hence, $\lambda_0 < 1$ implies that the second inequality in (71) is verified strictly, and thus $\kappa_0 > -1$. Finally, if $\lambda_0 = 1$, then $\kappa_0 = -1$ iff $r_0 < 1$ and

$$\begin{aligned} & \frac{r_0 - \left[\frac{1}{n-1} + r_0 \frac{n-2}{n-1} \right]}{1 - r_0} = -1 \\ \Leftrightarrow & \frac{(n-1)r_0 - [1 + r_0(n-2)]}{(n-1)(1-r_0)} = -1 \\ \Leftrightarrow & \frac{r_0 - 1}{(n-1)(1-r_0)} = -1 \\ \Leftrightarrow & n = 2. \end{aligned}$$

8.3 Calculating local competitiveness and pairwise relatedness

We here derive expressions for local competitiveness and pairwise relatedness for pairwise interactions in Scenarios A-C. To calculate λ_0 we rely on the method outlined in the proof of Proposition 2, while to calculate r_0 we use a recursion equation, which is a standard technique for calculating probabilities of identity by descent (see Nagylaki, 1992, and Rousset, 2004, for background).

8.3.1 Scenario A

In Scenario A the individual fitness writes:

$$\begin{aligned} w^A(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*) &= s_0 + m(1-s_0) \cdot \exp[\delta_f(\pi_i - \bar{\pi}^*)] \\ &+ (1-m)(1-s_0) \cdot \frac{\exp[\delta_f(\pi_i - \bar{\pi}^*)]}{m + (1-m)n^{-1} \sum_{j=1}^n \exp[\delta_f(\pi_j - \bar{\pi}^*)]}. \end{aligned} \quad (72)$$

Use a first-order Taylor expansion with respect to δ_f , evaluated at $\delta_f = 0$, to write

$$\begin{aligned} w^A(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*) &= 1 + \delta_f m(1-s_0)(\pi_i - \bar{\pi}^*) e^{\delta_f(\pi_i - \bar{\pi}^*)} \\ &+ \frac{\delta_f(1-m)(1-s_0)(\pi_i - \bar{\pi}^*) e^{\delta_f(\pi_i - \bar{\pi}^*)} \left(m + \frac{1-m}{n} \sum_{j=1}^n e^{\delta_f(\pi_j - \bar{\pi}^*)} \right)}{\left(m + \left(\frac{1-m}{n} \right) \sum_{j=1}^n e^{\delta_f(\pi_j - \bar{\pi}^*)} \right)^2} \\ &- \frac{\delta_f \frac{(1-m)^2}{n} (1-s_0) e^{\delta_f(\pi_i - \bar{\pi}^*)} \cdot \sum_{j=1}^n (\pi_j - \bar{\pi}^*) e^{\delta_f(\pi_j - \bar{\pi}^*)}}{\left(m + \left(\frac{1-m}{n} \right) \sum_{j=1}^n e^{\delta_f(\pi_j - \bar{\pi}^*)} \right)^2} \\ &+ \mathcal{O}(\delta^2) \\ &= 1 + \delta_f m(1-s_0) \cdot (\pi_i - \bar{\pi}^*) \\ &+ \delta_f(1-m)(1-s_0) \left[(\pi_i - \bar{\pi}^*) - \frac{1-m}{n} \sum_{j=1}^n (\pi_j - \bar{\pi}^*) \right] + \mathcal{O}(\delta^2). \end{aligned} \quad (73)$$

Neglecting higher-order terms, and rearranging terms,

$$\begin{aligned}
w^A(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*) &= 1 + \delta_f \left[(1-m)(1-s_0) \cdot \left[(\pi_i - \bar{\pi}^*) - (1-m) \frac{\sum_{j=1}^n (\pi_j - \bar{\pi}^*)}{n} \right] \right. \\
&\quad \left. + m(1-s_0) \cdot (\pi_i - \bar{\pi}^*) \right] \\
&= 1 + \delta_f \left[\left[(1-m)(1-s_0) \left(1 - \frac{1-m}{n} \right) + m(1-s_0) \right] (\pi_i - \bar{\pi}^*) \right. \\
&\quad \left. - (1-m)^2(1-s_0) \frac{\sum_{j \neq i} (\pi_j - \bar{\pi}^*)}{n} \right] \\
&= 1 + \delta_f \left[\left(1 - s_0 - \frac{(1-m)^2(1-s_0)}{n} \right) (\pi_i - \bar{\pi}^*) \right. \\
&\quad \left. - \frac{(n-1)(1-m)^2(1-s_0)}{n} \cdot \frac{\sum_{j \neq i} (\pi_j - \bar{\pi}^*)}{n-1} \right].
\end{aligned} \tag{74}$$

Recalling (44), we obtain

$$\beta = \frac{n(1-s_0) - (1-m)^2(1-s_0)}{n} \tag{75}$$

and

$$\gamma = \frac{(n-1)(1-m)^2(1-s_0)}{n}, \tag{76}$$

so that, using (50), we get

$$\lambda_0 = \frac{(n-1)(1-m)^2(1-s_0)}{n(1-s_0) - (1-m)^2(1-s_0)}. \tag{77}$$

To calculate r_0 , one uses a recursion equation. In Scenario A this equation writes

$$\begin{aligned}
r_0 &= s_0^2 r_0 + 2s_0(1-s_0)(1-m) \left(\frac{1}{n} + \frac{n-1}{n} r_0 \right) \\
&\quad + (1-s_0)^2 (1-m)^2 \left(\frac{1}{n} + \frac{n-1}{n} r_0 \right).
\end{aligned} \tag{78}$$

The left-hand side is the average probability that, in a monomorphic population, the neighbor of a randomly drawn member of a certain lineage, is also a member of this lineage (i.e., the probability that two randomly sampled individuals without replacement in the same group are identical by descent). The terms on the right-hand side details the events in which this happens. The first term on the right hand side corresponds to the event that both the mutant at hand and the randomly drawn neighbor survived from the previous period. The second term on the right hand side corresponds to the two events in which either the mutant at hand or the randomly drawn neighbor survived from the previous period while the other didn't, and the one who didn't survive from the previous period did not migrate in from

another island. In this case, there is a probability $1/n$ that one is the offspring of the other, in which case they are both mutants with certainty; with the complementary probability, they are not parent and child, in which case the probability that they are both mutants equals r_0 . The third term on the right hand side corresponds to the event in which neither the mutant at hand nor the randomly drawn neighbor survived from the previous period, and neither of them migrated in from another island. In this case, there is a probability $1/n$ that they have the same parent, in which case they are both mutants with certainty; with the complementary probability, they have different parents, in which case the probability that they are both mutants equals r_0 . Solving (78) for r_0 yields

$$r_0 = \frac{(1-m)[2s_0 + (1-s_0)(1-m)]}{n(1+s_0) - (n-1)(1-m)[2s_0 + (1-s_0)(1-m)]}. \quad (79)$$

8.3.2 Scenario B

In Scenario B the individual fitness of a focal individual i in a focal island writes:

$$w^B(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*) = \left[1 - \rho + \frac{2\rho \exp(\delta_v \bar{\pi})}{\exp(\delta_v \bar{\pi}) + \exp(\delta_v \bar{\pi}^*)} \right] \cdot \left[\frac{(1-m) \exp[\delta_f(\pi_i - \bar{\pi}^*)]}{m + (1-m)n^{-1} \sum_{j=1}^n \exp[\delta_f(\pi_j - \bar{\pi}^*)]} + m \exp[\delta_f(\pi_i - \bar{\pi}^*)] \right]. \quad (80)$$

Let $\delta_v = \delta \cdot \sigma_v$ and $\delta_f = \delta \cdot \sigma_f$, and use a first-order Taylor expansion with respect to δ , evaluated at $\delta = 0$, to write

$$w^B(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*) = 1 + \frac{\delta 2\rho}{[\exp(\delta_v \bar{\pi}) + \exp(\delta_v \bar{\pi}^*)]^2} \cdot [\bar{\pi} \sigma_v \exp(\delta_v \bar{\pi}) [\exp(\delta_v \bar{\pi}) + \exp(\delta_v \bar{\pi}^*)] - \exp(\delta_v \bar{\pi}) [\bar{\pi} \sigma_v \exp(\delta_v \bar{\pi}) + \bar{\pi}^* \sigma_v \exp(\delta_v \bar{\pi}^*)]] \cdot \left[\frac{(1-m) \exp[\delta_f(\pi_i - \bar{\pi}^*)]}{m + (1-m)n^{-1} \sum_{j=1}^n \exp[\delta_f(\pi_j - \bar{\pi}^*)]} + m \exp[\delta_f(\pi_i - \bar{\pi}^*)] \right] + \delta \left[1 - \rho + \frac{2\rho \exp(\delta_v \bar{\pi})}{\exp(\delta_v \bar{\pi}) + \exp(\delta_v \bar{\pi}^*)} \right] \cdot \left[\frac{(1-m) \sigma_f(\pi_i - \bar{\pi}^*) \left[m + \frac{1-m}{n} \sum_{j=1}^n \exp[\delta_f(\pi_j - \bar{\pi}^*)] \right] \exp[\delta_f(\pi_i - \bar{\pi}^*)]}{\left[m + (1-m)n^{-1} \sum_{j=1}^n \exp[\delta_f(\pi_j - \bar{\pi}^*)] \right]^2} - \frac{\frac{(1-m)^2}{n} \exp[\delta_f(\pi_i - \bar{\pi}^*)] \sum_{j=1}^n \sigma_f(\pi_j - \bar{\pi}^*) \exp[\delta_f(\pi_j - \bar{\pi}^*)]}{\left[m + (1-m)n^{-1} \sum_{j=1}^n \exp[\delta_f(\pi_j - \bar{\pi}^*)] \right]^2} + m \sigma_f(\pi_i - \bar{\pi}^*) \cdot \exp[\delta_f(\pi_i - \bar{\pi}^*)] + \mathcal{O}(\delta^2) \right]. \quad (81)$$

Neglecting higher-order terms and simplifying,

$$\begin{aligned}
w^B(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*) &= 1 + \delta\rho \cdot \frac{\sigma_v(\bar{\pi} - \bar{\pi}^*)}{2} \\
&+ \delta \left[(1-m)\sigma_f \left[(\pi_i - \bar{\pi}^*) - \frac{1-m}{n} \sum_{j=1}^n (\pi_j - \bar{\pi}^*) \right] \right. \\
&\left. + m\sigma_f \cdot (\pi_i - \bar{\pi}^*) \right],
\end{aligned} \tag{82}$$

or

$$\begin{aligned}
w^B(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*) &= 1 + \delta\rho\sigma_v \cdot \frac{\sum_{j=1}^n (\pi_j - \bar{\pi}^*)}{2n} \\
&+ \delta\sigma_f \left[(\pi_i - \bar{\pi}^*) - (1-m)^2 n^{-1} \sum_{j=1}^n (\pi_j - \bar{\pi}^*) \right] \\
&= 1 + \delta \left[\left[\frac{\rho\sigma_v}{2n} + \sigma_f \left(1 - \frac{(1-m)^2}{n} \right) \right] \cdot (\pi_i - \bar{\pi}^*) \right. \\
&\left. - \left[\sigma_f(1-m)^2 - \frac{\rho\sigma_v}{2} \right] \frac{n-1}{n} \cdot \frac{\sum_{j \neq i} (\pi_j - \bar{\pi}^*)}{n-1} \right].
\end{aligned} \tag{83}$$

Recalling (44), we obtain

$$\beta = \frac{\rho\sigma_v}{2n} + \sigma_f \left(\frac{n - (1-m)^2}{n} \right) \tag{84}$$

and

$$\gamma = \left[\sigma_f(1-m)^2 - \frac{\rho\sigma_v}{2} \right] \frac{n-1}{n}, \tag{85}$$

so that, using (50), we get

$$\lambda_0 = \frac{(n-1) [2\sigma_f(1-m)^2 - \rho\sigma_v]}{\rho\sigma_v + 2\sigma_f [n - (1-m)^2]}. \tag{86}$$

For r_0 , the recursion equation writes

$$r_0 = (1-m)^2 \left(\frac{1}{n} + \frac{n-1}{n} r_0 \right). \tag{87}$$

In this scenario, the only event in which a randomly drawn mutant can be related to a randomly drawn neighbor, is when both stayed in their natal island. In this case, there is a probability $1/n$ that they have the same parent, in which case they are both mutants with certainty; with the complementary probability, they have different parents, in which case the probability that the neighbor is a mutant is r_0 . Solving for r_0 yields

$$r_0 = \frac{(1-m)^2}{n - (n-1)(1-m)^2}. \tag{88}$$

8.3.3 Scenario C

In Scenario C the individual fitness of a focal individual i in a focal island writes:

$$w^C(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*) = s_0 + (1-m)(1-s_0) \cdot \frac{\exp(\delta_f \cdot \pi_i)}{n^{-1} \sum_{j=1}^n \exp(\delta_f \cdot \pi_j)} + m(1-s_0) \cdot \exp[\delta_f \cdot (\pi_i - \bar{\pi}^*)]. \quad (89)$$

Use a first-order Taylor expansion with respect to δ_f , evaluated at $\delta_f = 0$, to write

$$w^C(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*) = 1 + \delta(1-m)(1-s_0) \frac{\pi_i \exp(\delta_f \pi_i) n^{-1} \sum_{j=1}^n \exp(\delta_f \pi_j)}{\left[n^{-1} \sum_{j=1}^n \exp(\delta_f \pi_j) \right]^2} - \delta(1-m)(1-s_0) \frac{\exp(\delta_f \pi_i) n^{-1} \sum_{j=1}^n \pi_j \exp(\delta_f \pi_j)}{\left[n^{-1} \sum_{j=1}^n \exp(\delta_f \pi_j) \right]^2} + \delta m(1-s_0)(\pi_i - \bar{\pi}^*) \cdot \exp[\delta_f \cdot (\pi_i - \bar{\pi}^*)] + \mathcal{O}(\delta^2). \quad (90)$$

Neglecting higher-order terms, simplifying, and rearranging,

$$\begin{aligned} w^C(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*) &= 1 + \delta \left[(1-m)(1-s_0) \left(\pi_i - \frac{1}{n} \sum_{j=1}^n \pi_j \right) + m(1-s_0)(\pi_i - \bar{\pi}^*) \right] \\ &= 1 + \delta \left[\frac{mn(1-s_0) + (n-1)(1-m)(1-s_0)}{n} \cdot (\pi_i - \bar{\pi}^*) - \frac{(n-1)(1-m)(1-s_0)}{n} \cdot \frac{\sum_{j \neq i} (\pi_j - \bar{\pi}^*)}{n-1} \right]. \end{aligned} \quad (91)$$

Recalling (44), we obtain

$$\beta = \frac{nm(1-s_0) + (n-1)(1-m)(1-s_0)}{n} \quad (92)$$

and

$$\gamma = \frac{(n-1)(1-m)(1-s_0)}{n}, \quad (93)$$

so that, using (50), we get

$$\lambda_0 = \frac{(n-1)(1-m)}{nm + (n-1)(1-m)}. \quad (94)$$

For r_0 the recursion equation writes

$$\begin{aligned} r_0 &= s_0^2 r_0 + 2(1-m)s_0(1-s_0) \left(\frac{1}{n} + \frac{n-1}{n} r_0 \right) \\ &\quad + (1-m)^2 (1-s_0)^2 \cdot \left(\frac{1}{n} + \frac{n-1}{n} r_0 \right). \end{aligned} \quad (95)$$

The first term on the right-hand side corresponds to the event that both the mutant at hand and the randomly drawn neighbor have been loyal to their parents, where the neighbor's parent belongs to the mutant's lineage with probability r_0 . The second term on the right hand side corresponds to the event that either the mutant was loyal to its parent but the randomly drawn neighbor was not loyal to its parent, or the other way around. In this case, there is a probability $1/n$ that the non-loyal child acquired its trait from the loyal child's parent, in which case they are both mutants with certainty, while with the complementary probability this did not happen, in which case the probability that the randomly neighbor is a mutant is r_0 . The third term on the right hand side corresponds to the event that neither the mutant at hand nor the randomly drawn neighbor were loyal to their parents but both of them acquired their trait from someone in the island. In this case, there is a probability $1/n$ that they acquired their type from the same adult, in which case they are both mutants with certainty; with the complementary probability they have different cultural parents, in which case the probability that the randomly drawn neighbor is a mutant is r_0 . We note that the equation simplifies to

$$r_0 = s_0^2 r_0 + 2s_0(1-s_0)(1-m) \left(\frac{1}{n} + \frac{n-1}{n} r_0 \right) + (1-s_0)^2(1-m)^2 \left(\frac{1}{n} + \frac{n-1}{n} r_0 \right),$$

which has the same form as (78). Hence,

$$r_0 = \frac{(1-m)[2s_0 + (1-m)(1-s_0)]}{n(1+s_0) - (1-m)(n-1)[2s_0 + (1-m)(1-s_0)]}. \quad (96)$$

8.4 Proof of Theorem 1

To establish the first claim, let $v \in \Theta$ and consider any $(\hat{x}, \hat{y}) \in B_{\text{NE}}^0(u^0, v)$. Then, $u^0(\hat{x}, \hat{\mathbf{x}}) \geq u^0(\hat{y}, \hat{\mathbf{x}})$. By definition of u^0 , this inequality is equivalent to $\Pi(\hat{x}, \hat{\mathbf{x}}) \geq \Pi(\hat{y}, \hat{\mathbf{x}})$. Hence, u^0 is uninvadable under weak selection. For the second claim, let $u \neq u^0$. Assume further that $v = u^0$. By hypothesis in the claim, there exists some $(\hat{x}, \hat{y}) \in B_{\text{NE}}^0(u, u^0)$ for which $\hat{x} \notin X(u^0)$. Hence, there exists $y \in X$ such that $u^0(y, \hat{\mathbf{x}}^{(n-1)}) > u^0(\hat{x}, \hat{\mathbf{x}}^{(n-1)})$, or, equivalently, $\Pi(y, \hat{\mathbf{x}}) > \Pi(\hat{x}, \hat{\mathbf{x}})$.

References

- [1] Akçay, E., and J. van Cleve (2013): "Behavioral responses in structured populations pave the way to group optimality," *American Naturalist*, 179, 257-269.
- [2] Akerlof, G. and R. Kranton (2000): "Economics and Identity," *Quarterly Journal of Economics*, 115, 715-753.
- [3] Alberto, J. C. Micheletti, D. Ruxton and A. Gardner (2017): "Intrafamily and intragenomic conflicts in human warfare", *Proceedings of the Royal Society of London Series B (Biological Sciences)* 284, issue 1849.

- [4] Alexander, R.A. (1979): *Darwinism and Human Affairs*. University of Washington Press, Seattle.
- [5] Alger, I. (2010): “Public Goods Games, Altruism, and Evolution,” *Journal of Public Economic Theory*, 12, 789-813.
- [6] Alger, I. and R. Renault (2006): “Screening Ethics when Honest Agents Care about Fairness,” *International Economic Review*, 47, 59-85.
- [7] Alger, I. and J.W. Weibull (2010): “Kinship, Incentives and Evolution,” *American Economic Review*, 100, 1725-1758.
- [8] Alger, I. and J.W. Weibull (2012): “A Generalization of Hamilton’s Rule—Love Others How Much?” *Journal of Theoretical Biology*, 299, 42-54.
- [9] Alger, I., and J.W. Weibull (2013): “Homo moralis—Preference evolution under incomplete information and assortative matching,” *Econometrica* 81: 2269-2302.
- [10] Alger, I., and J.W. Weibull (2016): “Evolution and Kantian morality,” *Games and Economic Behavior* 98, 56-67.
- [11] Andreoni, J. (1990): “Impure altruism and donations to public goods: A theory of warm-glow giving,” *Economic Journal*, 100, 464-477.
- [12] Aoki, K. (1982): “A condition for group selection to prevail over counteracting individual selection,” *Evolution*, 36, 832-842.
- [13] Bandiera, O., I. Barankay, and I. Rasul (2005): “Social preferences and the response to incentives: Evidence from personnel data,” *Quarterly Journal of Economics*, 120, 917-962.
- [14] Bao, M., and G. Wild (2012): “Reproductive skew can provide a net advantage in both conditional and unconditional social interactions,” *Theoretical Population Biology*, 82, 200-208.
- [15] Becker, G. (1976): “Altruism, Egoism, and Genetic Fitness: Economics and Sociobiology,” *Journal of Economic Literature*, 14, 817-826.
- [16] Bénabou, R. and J. Tirole (2006): “Incentives and Prosocial Behavior,” *American Economic Review*, 96, 1652-1678.
- [17] Bergstrom, T. (1995): “On the Evolution of Altruistic Ethical Rules for Siblings,” *American Economic Review*, 85, 58-81.
- [18] Bergstrom, T. (1996): “Economics in a Family Way,” *Journal of Economic Literature*, 34, 1903-1934.
- [19] Bergstrom, T. (2003): “The Algebra of Assortative Encounters and the Evolution of Cooperation,” *International Game Theory Review*, 5, 211-228.

- [20] Bester, H. and W. Güth (1998): “Is Altruism Evolutionarily Stable?” *Journal of Economic Behavior and Organization*, 34, 193–209.
- [21] Bisin, A., and T. Verdier (2001): “The economics of cultural transmission and the dynamics of preferences”, *Journal of Economic Theory* 97: 298-319.
- [22] Bolle, F. (2000): “Is Altruism Evolutionarily Stable? And Envy and Malevolence? Remarks on Bester and Güth” *Journal of Economic Behavior and Organization*, 42, 131-133.
- [23] Bowles, S. (2006): “Group competition, reproductive leveling, and the evolution of human altruism,” *Science* 314, 1569-1572.
- [24] Bowles S. (2009): “Did warfare among ancestral hunter-gatherers affect the evolution of human social behaviors?” *Science* 324, 1293-8.
- [25] Bowles, S., and H. Gintis (1998): “The Moral Economy of Communities: Structured Populations and the Evolution of Pro-Social Norms,” *Evolution and Human Behavior*, 19, 3-25.
- [26] Brekke, K.A., S. Kverndokk, and K. Nyborg (2003): “An Economic Model of Moral Motivation,” *Journal of Public Economics*, 87, 1967–1983.
- [27] Bshary, R., and R. Bergmüller (2008): “Distinguishing four fundamental approaches to the evolution of helping,” *Journal of Evolutionary Biology*, 21, 405-20
- [28] Cavalli-Sforza, L.L., and W.F. Bodmer (1971): *The genetics of human populations*, W.H. Freeman, San Francisco.
- [29] Crow, J. F. and M. Kimura (1970): *An Introduction to Population Genetics Theory*. Harper and Row, New York.
- [30] Dekel, E., J.C. Ely, and O. Yilankaya (2007): “Evolution of preferences,” *Review of Economic Studies*, 74, 685-704.
- [31] dos Santos, M., and J. Peña (2017): “Antisocial rewarding in structured populations,” *Scientific Reports*, 7, 6212.
- [32] Ellingsen, T. (1997): “The Evolution of Bargaining Behavior,” *Quarterly Journal of Economics*, 112, 581-602.
- [33] Engel, C. (2011): “Dictator games: a meta study,” *Experimental Economics*, 14, 583–610.
- [34] Eshel, I. (1972): “On the neighbor effect and the evolution of altruistic traits,” *Theoretical Population Biology*, 11, 258–277.
- [35] Eshel, I., M.W. Feldman, and A. Bergman (1998): “Long-term evolution, short-term evolution, and population genetic theory,” *Journal of Theoretical Biology*, 191, 391-396.

- [36] Ewens, W. J. (2004): *Mathematical Population Genetics*. Springer Verlag, New York.
- [37] Fehr, E., and K. Schmidt (1999): “A Theory of Fairness, Competition, and Cooperation,” *Quarterly Journal of Economics*, 114, 817-868.
- [38] Feld, L.P. and B.S. Frey (2007): “Tax compliance as the result of a psychological tax contract: The role of incentives and responsive regulation,” *Law & Policy*, 29, 102–120.
- [39] Fershtman, C. and K. Judd (1987): “Equilibrium Incentives in Oligopoly,” *American Economic Review*, 77, 927–940.
- [40] Fershtman, C., and Y. Weiss (1998): “Social Rewards, Externalities and Stable Preferences,” *Journal of Public Economics*, 70, 53-73.
- [41] Frank, R.H. (1987): “If Homo Economicus Could Choose His Own Utility Function, Would He Want One with a Conscience?” *American Economic Review*, 77, 593-604.
- [42] Frank, S.A. (1998): *Foundations of Social Evolution*. Princeton University Press, Princeton, NJ.
- [43] Friedman, M. (1953): *Essays in Positive Economics*. Chicago: University of Chicago Press.
- [44] Gardner, A., and S.A. West (2004), “Spite and the scale of competition,” *Journal of Evolutionary Biology*, 17, 1195-1203.
- [45] Gardner, A., and S.A. West (2006): “Demography, altruism, and the benefits of budding,” *Journal of Evolutionary Biology*, 19, 1707-1716.
- [46] Gillespie, J. H. (2004): *Population Genetics: a Concise Guide*. Johns Hopkins, Baltimore & London.
- [47] Grafen, A. (1985): “A geometric view of relatedness,” in Dawkins, R. and M. Ridley, eds., *Oxford Surveys in Evolutionary Biology*, Oxford University Press, Oxford.
- [48] Grueter, C.C., B. Chapais, and D. Zinner (2012): “Evolution of multilevel social systems in nonhuman primates and humans,” *International Journal of Primatology*, 33, 1002–1037.
- [49] Güth, W., and M.G. Kocher (2014): “More than thirty years of ultimatum bargaining experiments: Motives, variations, and a survey of the recent literature,” *Journal of Economic Behavior & Organization*, 108, 396–409.
- [50] Güth, W., and M. Yaari (1992): “An evolutionary approach to explain reciprocal behavior in a simple strategic game,” in U. Witt (ed.), *Explaining Process and Change – Approaches to Evolutionary Economics*. Ann Arbor: University of Michigan Press.
- [51] Hamilton, W.D. (1964): “The genetical evolution of social behaviour,” *Journal of Theoretical Biology*, 7, 1-52.

- [52] Hamilton, W.D. (1971): “Selection of selfish and altruistic behavior in some extreme models,” in J.F. Eisenberg and W. S.Dillon, eds. *Man and beast: comparative social behavior*. Smithsonian Institution Press, Washington, D.C.
- [53] Hartl, D.L., and A.G. Clark (2007): *Principles of Population Genetics* (4th edition). Sinauer and Associates, Sunderland, MA.
- [54] Hauk, E., and M. Saez-Martí (2002): “On the cultural transmission of corruption.” *Journal of Economic Theory*, 107, 311-335.
- [55] Heifetz, A., C. Shannon, and Y. Spiegel (2007a): “What to Maximize if You Must,” *Journal of Economic Theory*, 133, 31-57.
- [56] Heifetz, A., C. Shannon, and Y. Spiegel (2007b): “The Dynamic Evolution of Preferences,” *Economic Theory*, 32, 251-286.
- [57] Hirshleifer, J. (1977): “Economics from a biological viewpoint,” *Journal of Law and Economics*, Vol. 20, 1-52.
- [58] Huck, S., and J. Oechssler (1999): “The Indirect Evolutionary Approach to Explaining Fair Allocations,” *Games and Economic Behavior*, 28, 13–24.
- [59] Johnstone, R.A., and M.A. Cant (2008) “Sex differences in dispersal and the evolution of helping and harming,” *American Naturalist*, 172, 318-30.
- [60] Kant, I. (1785): *Grundlegung zur Metaphysik der Sitten*. [In English: Groundwork of the Metaphysics of Morals. 1964. New York: Harper Torch books.]
- [61] Koçkesen, L., E.A. Ok, and R. Sethi (2000a): “The Strategic Advantage of Negatively Interdependent Preferences,” *Journal of Economic Theory*, 92, 274-299.
- [62] Koçkesen, L., E.A. Ok, and R. Sethi (2000b): “Evolution of Interdependent Preferences in Aggregative Games,” *Games and Economic Behavior* 31, 303-310.
- [63] Konrad, K.A. (2014) “Strategic aspects of fighting in alliances,” in K. Wärneryd, ed. *The Economics of Conflict: Theory and Empirical Evidence*. MIT Press, Cambridge, MA.
- [64] Layton, R., S. O’Hara, and A. Bilsborough (2012): “Antiquity and social functions of multilevel social organization among human hunter-gatherers,” *International Journal of Primatology*, 33, 1215–1245.
- [65] Lehmann, L., I. Alger, and J.W. Weibull (2015): “Does evolution lead to maximizing behavior?” *Evolution* 69-7, 1858–1873.
- [66] Lehmann, L., and Balloux (2007): “Natural selection on fecundity variance in subdivided populations: kin selection meets bet hedging”, *Genetics* 176, 361-377.
- [67] Lehmann, L., K. Foster, and F. Feldman (2008): “Cultural transmission can inhibit the evolution of altruistic helping,” *American Naturalist*, 172, 12-24

- [68] Lehmann, L., C. Mullon, E. Akçay, and J. Van Cleve (2016): “Invasion fitness, inclusive fitness, and reproductive numbers in heterogeneous populations,” *Evolution*, 70, 1689–1702.
- [69] Lehmann, L. and F. Rousset (2009): “Perturbation expansions of multilocus fixation probabilities for frequency-dependent selection with applications to the Hill-Robertson effect and to the joint evolution of helping and punishment”, *Theoretical Population Biology* 76: 35–51.
- [70] Lehmann, L., and F. Rousset (2010): “How life history and demography promote or inhibit the evolution of helping behaviours,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365,. 2599-2617.
- [71] Lessard, S. (2009): “Diffusion approximations for one-locus multi-allele kin selection, mutation and random drift in group-structured populations: a unifying approach to selection models in population genetics”, *Journal of Mathematical Biology* 59: 659–696.
- [72] Levine, D. (1998): “Modelling Altruism and Spite in Experiments,” *Review of Economic Dynamics*, 1, 593-622.
- [73] Lindbeck, A., and S. Nyberg (2006): “Raising children to work hard: altruism, work norms and social insurance,” *Quarterly Journal of Economics*, 121, 1473-1503.
- [74] Lion, S., and S. Gandon (2010): “Life history, habitat saturation and the evolution of fecundity and survival altruism,” *Evolution*, 64, 1594-606.
- [75] Lumsden, C.J., and E.O. Wilson (1981): *Genes, Mind, and Culture*. Harvard University Press, Cambridge, MA.
- [76] Malone, N., A. Fuentes, and F.J. White (2012): “Variation in the social systems of extant hominoids: Comparative insight into the social behavior of early hominins,” *International Journal of Primatology*, 33, 1251–1277.
- [77] Maynard Smith, J., and G.R. Price (1973): “The logic of animal conflict,” *Nature* 246: 15-18.
- [78] Michod, R.E., and W.D. Hamilton (1980): “Coefficients of relatedness in sociobiology,” *Nature* 288, 694–697.
- [79] Nagylaki, T. (1992): *Introduction to Population Genetics*. Springer, Berlin.
- [80] Ockenfels, P. (1993): “Cooperation in Prisoners’ Dilemma—An Evolutionary Approach”, *European Journal of Political Economy*, 9, 567-579.
- [81] Ohtsuki, H. (2012): “Does synergy rescue the evolution of cooperation? An analysis for homogeneous populations with non-overlapping generations,” *Journal of Theoretical Biology*, 307, 20-28.

- [82] Ok, E.A., and F. Vega-Redondo (2001): “On the Evolution of Individualistic Preferences: An Incomplete Information Scenario,” *Journal of Economic Theory*, 97, 231-254.
- [83] Pen, I. (2000): “Reproductive effort in viscous populations,” *Evolution*, 54, 293-297.
- [84] Possajennikov, A. (2000): “On the Evolutionary Stability of Altruistic and Spiteful Preferences,” *Journal of Economic Behavior and Organization*, 42, 125-129.
- [85] Rabin, M. (1993): “Incorporating Fairness into Game Theory and Economics,” *American Economic Review*, 83, 1281-1302.
- [86] Rayo, L., and G.S. Becker (2007): “Evolutionary efficiency and happiness,” *Journal of Political Economy*, 115, 302-337.
- [87] Robson, A. (1996): “A biological basis for expected and non-expected utility,” *Journal of Economic Theory*, 68, 397-424.
- [88] Robson, A. (2001): “The Biological Basis of Economic Behavior,” *Journal of Economic Literature*, 39, 11-33.
- [89] Robson, A. and Samuelson, L. (2011a): “The evolution of decision and experienced utilities,” *Theoretical Economics*, 6: 311–339.
- [90] Robson, A. and Samuelson, L. (2011b): “The evolutionary foundations of preferences,” in Alberto Bisin and Matt Jackson. eds., *Handbook of Social Economics* North-Holland.
- [91] Rousset, F. (2004): *Genetic Structure and Selection in Subdivided Populations*. Princeton University Press, Princeton, NJ.
- [92] Roze, D. and F. Rousset (2003): “Selection and drift in subdivided populations: A straightforward method for deriving diffusion approximations and applications involving dominance, selfing and local extinctions,” *Genetics* 165: 2153–2166.
- [93] Schaffer, M.E. (1988): “Evolutionarily stable strategies for finite populations and variable contest size,” *Journal of Theoretical Biology*, 132, 467-478.
- [94] Sethi, R., and E. Somanathan (2001): “Preference evolution and reciprocity,” *Journal of Economic Theory*, 97, 273-297.
- [95] Skaperdas, S. (1996): “Contest success functions,” *Economic Theory* 7, 283-290.
- [96] Taylor, P.D. (1992a) “Altruism in viscous populations - an inclusive fitness model,” *Evolutionary Ecology*, 6, 352-356.
- [97] Taylor, P.D. (1992b) “Inclusive fitness in a homogeneous environment,” *Proceedings of the Royal Society B*, 249, 299-302.
- [98] Taylor, P.D., and A.J. Irwin (2000) “Overlapping generations can promote altruistic behavior,” *Evolution*, 54, 1135-41.

- [99] Tullock, G. (1980): “Efficient rent seeking,” in Buchanan, J., Tollison, R. and Tullock, G., Eds., *Toward a Theory of Rent Seeking Society*, Texas A and M University Press, College Station, 97-112.
- [100] Van Cleve, J. (2015): “Social evolution and genetic interactions in the short and long term,” *Theoretical Population Biology*, 2013, 2-26.
- [101] Van Schaik, C. P. (2016) *The Primate Origin of Human Behavior*. Wiley-Blackwell, Hoboken, NJ.
- [102] Waldman, M. (1994): “Systematic errors and the theory of natural selection,” *American Economic Review*, 84, 482-497.
- [103] Weibull, J.W. (1995): *Evolutionary Game Theory*. MIT Press, Cambridge, MA.
- [104] Weibull, J.W. and M. Salomonsson (2006): “Natural selection and social preferences,” *Journal of Theoretical Biology*, 239, 79-92.
- [105] West, S.A., A. Griffin, and A. Gardner (2007): “Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection,” *Journal of Evolutionary Biology*, 20, 415-32.
- [106] Wilson, D.S., G.B. Pollock, and L.A. Dugatkin (1992): “Can altruism evolve in purely viscous populations?” *Evolutionary Ecology*, 6, 331-341.
- [107] Wright, S. (1931): “Evolution in Mendelian populations,” *Genetics*, 16, 97–159.
- [108] Wu, J. (2017): “A note on altruism in asymmetric games: An indirect evolutionary approach,” *Biological Theory*, 12, 181-188.
- [109] Zelmer, J. (2003): “Linear public goods experiments: A meta-analysis,” *Experimental Economics*, 6, 299-310.