




# The wisdom of crowds versus the madness of mobs: An evolutionary model of bias, polarization, and other challenges to collective intelligence

Collective Intelligence  
August-September 2022: 1–22  
© The Author(s) 2022  
Article reuse guidelines:  
[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)  
DOI: 10.1177/26339137221104785  
[journals.sagepub.com/home/col](https://journals.sagepub.com/home/col)  


## Andrew W Lo

MIT Laboratory for Financial Engineering, Cambridge, MA, USA; MIT Sloan School of Management, Cambridge, MA, USA; MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA, USA; Santa Fe Institute, Santa Fe, NM, USA

## Ruixun Zhang

School of Mathematical Sciences, Peking University, Beijing, China; Center for Statistical Science, Peking University, Beijing, China; National Engineering Laboratory for Big Data Analysis and Applications, Peking University, Beijing, China

## Abstract

Despite its success in financial markets and other domains, collective intelligence seems to fall short in many critical contexts, including infrequent but repeated financial crises, political polarization and deadlock, and various forms of bias and discrimination. We propose an evolutionary framework that provides fundamental insights into the role of heterogeneity and feedback loops in contributing to failures of collective intelligence. The framework is based on a binary choice model of behavior that affects fitness; hence, behavior is shaped by evolutionary dynamics and stochastic changes in environmental conditions. We derive collective intelligence as an emergent property of evolution in this framework, and also specify conditions under which it fails. We find that political polarization emerges in stochastic environments with reproductive risks that are correlated across individuals. Bias and discrimination emerge when individuals incorrectly attribute random adverse events to observable features that may have nothing to do with those events. In addition, path dependence and negative feedback in evolution may lead to even stronger biases and levels of discrimination, which are locally evolutionarily stable strategies. These results suggest potential policy interventions to prevent such failures by nudging the “madness of mobs” towards the “wisdom of crowds” through targeted shifts in the environment.

## Keywords

Collective intelligence, political polarization, bias, discrimination, evolutionarily stable strategy, group selection

---

## Corresponding author:

Ruixun Zhang, School of Mathematical Sciences, Peking University, 5 Yiheyuan Road, Beijing, 100871, China.

Email: [zhangruixun@pku.edu.cn](mailto:zhangruixun@pku.edu.cn)



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

## Significance statement

Collective intelligence refers to the group knowledge and wisdom that emerges from the collaboration and competition among many individuals. Despite its ubiquity and significance in financial markets and other domains, collective intelligence is not easy to achieve and can also fail dramatically under certain conditions. Examples include infrequent but repeated financial crises, political polarization and deadlock, and various forms of bias and discrimination. We propose an evolutionary framework that provides fundamental insights into the failure of collective intelligence by answering the following questions: In what environments are polarization and discrimination likely emerge? What are the drivers behind these phenomena? And more importantly, how can we avoid “collective ignorance” and promote collective intelligence instead? We derive collective intelligence as an emergent property of evolution and specify conditions under which it fails. Political polarization emerges in stochastic environments with reproductive risks that are correlated across individuals. Bias and discrimination emerge when individuals incorrectly attribute random adverse events to observable features that may have nothing to do with those events. Moreover, path dependence and negative feedback in evolution may lead to even stronger levels of discrimination. These results suggest potential policy interventions to prevent such failures by nudging the “madness of mobs” towards the “wisdom of crowds” through targeted shifts in the environment, which is likely to be more effective than attempting to outlaw undesirable behaviors. As long as the environmental factors giving rise to these behaviors are still in force, the banned behaviors will re-emerge in one form or another.

## Introduction

Collective intelligence—a term for shared or group knowledge and wisdom that emerges from the collaboration and competition of many individuals—has been studied across decades in many disciplines ranging from the cognitive neurosciences to evolutionary biology to economics and sociology to engineering and computer science. However, despite its ubiquity and importance, collective intelligence is not easy to achieve and can also fail, sometimes repeatedly. One such example is the prevalence of bubbles and crashes in financial markets (Lo, 2013), such as the dot-com bubble in 1990s, the financial crisis of 2007–2008, and most recently, the financial turmoil during the first few months of the COVID-19 pandemic. No matter how different the latest financial frenzy or crisis appears to be, there are usually similarities to past experience (Reinhart and Rogoff, 2009).

Two of the most hotly debated issues today—political polarization and discrimination—are also examples of the failure of collective intelligence. Since the 2010s, we have witnessed the rise of populism and nationalism as part of a reaction against the global policies of the last 30 years in Western democracies and beyond, not to mention gender, religious, and other types of bias. These examples raise the natural question of why collective intelligence falters in these cases, but succeeds so well in so many other contexts?

In this article, we propose a formal mathematical model of the evolution of behavior to understand failures of collective intelligence by answering the following questions: In what environments will polarization and discrimination likely emerge? What are the key drivers behind these phenomena? And, most importantly, how can we avoid

“collective ignorance”<sup>1</sup> and promote collective intelligence instead?

We start by introducing our modeling framework, which builds upon the binary choice model of Brennan and Lo (2011) and Zhang et al. (2014a). We then apply this framework to study the rise of extreme political views, after which we turn our attention to discrimination. We conclude by discussing the broad applicability as well as the limitation of our framework, and provide several practical policy implications for reducing or preventing failures of collective intelligence. Given the breadth of engagement in our chosen topic, we also provide a review of the several distinct literatures related to our work in the [Supplementary Material](#).

## Modeling framework

When any behavior has consequences for fitness, evolutionary principles apply. The actions underneath polarization and bias—which political views to adopt and whether to discriminate against a particular group—yield different economic (or, in an evolutionary context, reproductive) consequences for individuals in different environments. In addition, the nature of risks in the environment also affect what behavior will emerge, and these behaviors may not always agree with individual rationality (Zhang et al., 2014a; 2014b).

Our framework consists of an initial population of hypothetical individuals (not necessarily human) that live for one period of unspecified length, and engage in a single binary decision that has consequences for the random number of offspring they will generate asexually. To the

extent that their behavior is linked to fecundity, only the most reproductively successful behaviors will flourish, due to the dynamics of evolution.<sup>2</sup> Although obvious from an evolutionary biologist’s perspective, this observation yields surprisingly specific implications regarding the types of behavior that are sustainable over time, behaviors that are likely to be innate to most living organisms due to the simplicity and generality of the binary choice framework. The evolved behavior will be collectively intelligent to the extent that it maximizes the population growth rate, but it may also generate other undesirable consequences in certain environments.

To illustrate the basic intuition behind this approach, we first present a simple numerical example before turning to the formal model.<sup>3</sup> Consider a population of individuals, each facing a binary choice between one of two possible actions,  $a$  and  $b$ . Environmental conditions will be positive 70% of the time, and action  $a$  will lead to reproductive success, generating 3 offspring for the individual. Environmental conditions will be negative 30% of the time, and action  $a$  will lead to 0 offspring. Action  $b$  has exactly the opposite outcomes—whenever  $a$  yields 3 offspring,  $b$  yields 0, and whenever  $a$  yields 0,  $b$  yields 3. From the individual’s perspective, always choosing  $a$ , which has the higher probability of reproductive success, will lead to more offspring on average. However, if all individuals in the population behaved in this “rational” manner, the first time that a negative environmental condition occurs, the entire population would become extinct. Assuming that offspring behave identically to their parents, the “always choose  $a$ ” behavior cannot survive over time. For the same reason, “always choose  $b$ ” is also unsustainable.

In fact, in this special case, the behavior with the highest fitness over time is for each individual to choose  $a$  70% of the time, and  $b$  30% of the time, matching the probabilities of reproductive success and failure. The group of individuals exhibiting this probability-matching behavior will achieve the maximum possible growth rate, and eventually, this behavior will dominate the entire population. As a result, it appears as though selection operates at the group level, and that this group—all individuals who randomize their actions with 70% probability—is the fittest from the perspective of reproductive success.<sup>4</sup>

This simple but abstract example illustrates the principle that a given behavior may seem irrational, but when viewed in the broader context of a given environment, can come to dominate the population because individuals engaging in such behavior will reproduce more quickly in that environment than those with other behaviors. To alter such behavior, we must look to the environment that gave rise to this adaptation and change that environment, otherwise the behavior will persist.

We present the formal model in the next section, which is based on Brennan and Lo (2011) and Zhang et al. (2014a).

Table 1 summarizes the key parameters and constraints in our model.

### Formal model

We begin with a population of individuals that live for one period, produce a random number of offspring asexually and only once, and then die. During their lives, individuals make only one decision: they choose from two actions,  $a$  and  $b$ , and this results in one of two corresponding random numbers of offspring,  $x_a$  and  $x_b$ . Note that  $x_a$  and  $x_b$  can be correlated, and their joint distribution represents the entirety of the implications of an individual’s actions for fitness.

We impose a factor structure for  $x_a$  and  $x_b$ , that is, suppose there are two independent environmental factors,  $\lambda_1$  and  $\lambda_2$ , that determine fitness, and  $x_a$  and  $x_b$  are both linear combinations of these two factors

$$\begin{aligned} x_a &= \beta_a \lambda_1 + (1 - \beta_a) \lambda_2 \\ x_b &= \beta_b \lambda_1 + (1 - \beta_b) \lambda_2 \end{aligned} \quad (1)$$

where  $\lambda_1$  and  $\lambda_2$  are nonnegative, and  $\beta_a$  and  $\beta_b$  are between 0 and 1.<sup>5</sup> Because these factors affect the fitness of *all* individuals in the population, we refer to them as *systematic*, and we assume that:

(A1)  $\lambda_1$  and  $\lambda_2$  are independent random variables with some well-behaved distribution functions, such that  $(x_a, x_b)$  and  $\log(px_a + (1 - p)x_b)$  have finite mean and variance for all  $p \in [0, 1]$ ,  $\beta_a \in [0, 1]$ , and  $\beta_b \in [0, 1]$ ; and

(A2)  $(\lambda_1, \lambda_2)$  is independent and identically distributed (IID) over time and identical for all individuals in a given generation.

We shall henceforth refer to  $(\beta_a, \beta_b)$  as an individual’s *characteristics*. For each action, individuals’ fitness involves a tradeoff between exposure to these two factors.

We give two examples of such factor structure to provide intuition for the key idea of the model. In the context of the evolution of hypothetical animals,  $\lambda_1$  might represent weather conditions and  $\lambda_2$  might represent the topography of the terrain. An animal can choose to hunt on the mountain (action  $a$ ) or in the forest (action  $b$ ). The success of hunting on the mountain is highly dependent on the weather, corresponding to a high value of  $\beta_a$ . On the other hand, because the forest provides shelter against extreme weather, the success of hunting in the forest depends mostly on its topography, corresponding to a low value of  $\beta_b$ .

In the context of social evolution in humans,  $\lambda_1$  might represent the degree of globalization in a society, and  $\lambda_2$  might represent the amount of natural resources available locally, such as crude oil. An individual then faces the choice of opening a manufacturing facility (action  $a$ ) or an oil refinery (action  $b$ ). The success of the manufacturing facility depends on the degree of globalization, which provides access to cheap labor globally, corresponding to a

**Table I.** Model parameters and constraints.

Parameters/Constraints	Explanation
$a$ and $b$	Two actions for each individual to choose from.
$x_a \geq 0$ and $x_b \geq 0$	Random numbers of offspring that correspond to choice $a$ and $b$ .
$\lambda_1 \geq 0$ and $\lambda_2 \geq 0$	Two environmental factors that determine fitness, independent and identically distributed (IID) over time and identical for all individuals in a given generation.
$\beta_a \in [0, 1]$ and $\beta_b \in [0, 1]$	Individual characteristics that determine the loading of its fitness on factors.
$p \in [0, 1]$	Individual behavior, defined as the probability to choose action $a$ .
$f \equiv (p, \beta_a, \beta_b)$	An individual's <i>type</i> , which is the unit of selection in evolution because it completely characterizes an individual.
$\mu(p, \beta_a, \beta_b)$	Log-geometric-average population growth rate for individuals of type $f = (p, \beta_a, \beta_b)$ .
$f^* = (p^*, \beta_a^*, \beta_b^*)$	The growth-optimal type that yields the fastest population growth rate.

high value of  $\beta_a$ . However, the success of the oil refinery obviously depends on the availability of crude oil locally, corresponding to a low value of  $\beta_b$ .

Our framework is general, in the sense that we embed in  $x_a$  and  $x_b$ —or equivalently, in factors and individual characteristics—the entire biological machinery that is fundamental to evolution, that is, genetics, but which is of less direct interest to social scientists than the link between behavior and fitness. If action  $a$  leads to higher fecundity than action  $b$  for individuals in a given population, the particular set of genes that predispose individuals to select  $a$  over  $b$  will be favored by natural selection, in which case these genes will survive and flourish, implying that the behavior “choose  $a$  over  $b$ ” will flourish as well.

Using this framework, we show below that the degree of globalization as a factor can affect the emergence of extreme political views, and that the crime rate of racially categorized groups is another factor that can affect the emergence of discriminatory behaviors.

### Individual behavior

Suppose each individual chooses action  $a$  with some probability  $p \in [0, 1]$  and action  $b$  with probability  $1 - p$ , denoted by the Bernoulli random variable  $I^p$ , hence the number of offspring of an individual is given by the random variable

$$x^p = I^p x_a + (1 - I^p) x_b,$$

where

$$I^p = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p. \end{cases}$$

We shall henceforth refer to  $p$  as the individual's *behavior* since it completely determines how the individual chooses between actions  $a$  and  $b$ . Note that  $p$  can be 0 or 1, which corresponds to deterministic behaviors. Generally,  $p$  can also be between 0 and 1, which corresponds to randomizing behaviors.

In this framework, an individual is completely characterized by its behavior  $p$  and characteristics  $(\beta_a, \beta_b)$ . We

shall henceforth refer to  $f \equiv (p, \beta_a, \beta_b)$  as an individual's *type*. To complete the specification of our model, we assume that offspring behave in a manner identical to their parent, that is, they have the same characteristics  $(\beta_a, \beta_b)$ , and choose between  $a$  and  $b$  according to the same  $p$ ; hence, the population may be viewed as comprising many different types, each indexed by the triplet  $f$ . The assumption that offspring from a type- $f$  parent are also of the same type  $f$  implies perfect genetic transmission of behavior from one generation to the next (that is, once a type  $f$ , always a type  $f$ ).

Although clearly unrealistic from a biological perspective, this simplification highlights and clarifies the impact of evolutionary dynamics on behavior, allowing us to derive the growth-optimal behavior explicitly.<sup>6</sup> However, [Brennan et al. \(2018\)](#) have extended this model to allow for mutation, which we shall also consider in our framework below.

In summary, an individual  $i$  of type  $f = (p, \beta_a, \beta_b)$  produces a random number of offspring

$$x_i^{p, \beta_a, \beta_b} = I_i^p x_{a,i}^{\beta_a} + (1 - I_i^p) x_{b,i}^{\beta_b} \quad (2)$$

where

$$\begin{aligned} x_{a,i}^{\beta_a} &= \beta_a \lambda_1 + (1 - \beta_a) \lambda_2 \\ x_{b,i}^{\beta_b} &= \beta_b \lambda_1 + (1 - \beta_b) \lambda_2 \end{aligned} \quad (3)$$

Here, individuals are indexed by  $i$ . In a given generation, individuals with the same characteristics  $\beta_a$  and  $\beta_b$  yield identical fitness as shown in (3), hence we may omit the subscript  $i$  wherever it is unambiguous.

### Population dynamics

Now consider an initial population of individuals that contains an equal number of all types, which we normalize to be 1 each without loss of generality. Suppose the total number of type  $f = (p, \beta_a, \beta_b)$  individuals in generation  $T$  is  $n_T^f$ . Because  $n_T^f$  grows exponentially over time  $T$ , we consider the exponential growth rate of the population size,  $T^{-1} \log n_T^f$ . Under assumptions (A1)

**Table 2.** Growth-optimal type  $f^* = (p^*, \beta_a^*, \beta_b^*)$  for the binary choice model.

	Growth-optimal characteristics	Growth-optimal behavior
If $\alpha_1^* = 1$	$\{(\beta_a, \beta_b) : \beta_a = 1 \text{ or } \beta_b = 1\}$	$p^* = \begin{cases} \frac{\alpha_1^* - \beta_b^*}{\beta_a^* - \beta_b^*} = 1 & \text{if } \beta_a^* = 1, \beta_b^* \neq 1 \\ \frac{\alpha_1^* - \beta_b^*}{\beta_a^* - \beta_b^*} = 0 & \text{if } \beta_a^* \neq 1, \beta_b^* = 1 \\ \text{arbitrary} & \text{if } \beta_a^* = \beta_b^* = 1 \end{cases}$
If $\alpha_1^* = 0$	$\{(\beta_a, \beta_b) : \beta_a = 0 \text{ or } \beta_b = 0\}$	$p^* = \begin{cases} \frac{\alpha_1^* - \beta_b^*}{\beta_a^* - \beta_b^*} = 1 & \text{if } \beta_a^* = 0, \beta_b^* \neq 0 \\ \frac{\alpha_1^* - \beta_b^*}{\beta_a^* - \beta_b^*} = 0 & \text{if } \beta_a^* \neq 0, \beta_b^* = 0 \\ \text{arbitrary} & \text{if } \beta_a^* = \beta_b^* = 0 \end{cases}$
If $0 < \alpha_1^* < 1$	$\{(\beta_a, \beta_b) : (\beta_a - \alpha_1^*)(\beta_b - \alpha_1^*) \leq 0\}$	$p^* = \begin{cases} \frac{\alpha_1^* - \beta_b^*}{\beta_a^* - \beta_b^*} & \text{if } \beta_a^* \neq \beta_b^* \\ \text{arbitrary} & \text{if } \beta_a^* = \beta_b^* \end{cases}$

and (A2), it is easy to show that  $T^{-1} \log n_T^f$  converges in probability to the log-geometric-average growth rate

$$\mu(p, \beta_a, \beta_b) = \mathbb{E} \left[ \log \left( p x_a^{\beta_a} + (1-p) x_b^{\beta_b} \right) \right], \quad (4)$$

as the number of generations and the number of individuals in each generation increase without bound.<sup>7</sup> Note that the term inside the logarithm of (4) is written as a linear combination of  $x_a^{\beta_a}$  and  $x_b^{\beta_b}$ , the fitness of actions  $a$  and  $b$ . Because selection occurs at the level of type  $f = (p, \beta_a, \beta_b)$ , it is also useful to define

$$\begin{aligned} \alpha_1 &= p\beta_a + (1-p)\beta_b \\ \alpha_2 &= p(1-\beta_a) + (1-p)(1-\beta_b) \end{aligned} \quad (5)$$

so that (4) can be rewritten as

$$\mu(p, \beta_a, \beta_b) = \mathbb{E}[\log(\alpha_1 \lambda_1 + (1 - \alpha_1) \lambda_2)], \quad (6)$$

where the term inside the logarithm is a linear combination of factors  $\lambda_1$  and  $\lambda_2$ . It is easy to see that  $\alpha_1 + \alpha_2 = 1$ , and we shall henceforth refer to  $(\alpha_1, \alpha_2)$  as the *factor loadings* of type- $f$  individuals. Equations (4) and (6) characterize the log-geometric-average growth rate of individuals as a function of their type  $f$  in terms of both behavior  $p$  and characteristics  $(\beta_a, \beta_b)$ .<sup>8</sup>

Over time, because the population grows exponentially, individuals with the largest growth rate will dominate the population at a geometric rate, as specified in the following result.<sup>9</sup>

**Proposition 1.** *Under assumptions (A1) and (A2), the optimal factor loading,  $\alpha_1^*$ , that maximizes the log-geometric-average growth rate (6) is given by*

$$\alpha_1^* = \begin{cases} 1 & \text{if } \mathbb{E}[\lambda_1/\lambda_2] > 1 \text{ and } \mathbb{E}[\lambda_2/\lambda_1] < 1 \\ \text{solution to (8)} & \text{if } \mathbb{E}[\lambda_1/\lambda_2] \geq 1 \text{ and } \mathbb{E}[\lambda_2/\lambda_1] \geq 1 \\ 0 & \text{if } \mathbb{E}[\lambda_1/\lambda_2] < 1 \text{ and } \mathbb{E}[\lambda_2/\lambda_1] > 1 \end{cases} \quad (7)$$

where  $\alpha_1^*$  is defined implicitly in the second case of (7) by

$$\mathbb{E} \left[ \frac{\lambda_1}{\alpha_1^* \lambda_1 + (1 - \alpha_1^*) \lambda_2} \right] = \mathbb{E} \left[ \frac{\lambda_2}{\alpha_1^* \lambda_1 + (1 - \alpha_1^*) \lambda_2} \right]. \quad (8)$$

Furthermore, based on (7), the growth-optimal type,  $f^* = (p^*, \beta_a^*, \beta_b^*)$ , is given explicitly in Table 2.

The three possible scenarios in (7) reflect the relative fitness of the two factors.  $\alpha_1^* = 1$  corresponds to behaviors and characteristics with a full loading on  $\lambda_1$ , which is growth-optimal if  $\lambda_1$  exhibits unambiguously higher expected relative fecundity;  $\alpha_1^* = 0$  will be growth-optimal if the opposite is true; and having a balanced loading between  $\lambda_1$  and  $\lambda_2$  will be growth-optimal if neither factor has a clear-cut reproductive advantage.

The growth-optimal characteristics and associated optimal behaviors in Table 2 show that, when  $\alpha_1^*$  is 1 or 0, one of the factors, either  $\lambda_1$  or  $\lambda_2$ , is significantly more important than the other, and the growth-optimal strategy places all the weight on the more important factor. However, when  $\alpha_1^*$  is strictly between 0 and 1, a combination of factors  $\lambda_1$  and  $\lambda_2$  will be necessary to achieve the maximum growth rate. Individual characteristics  $(\beta_a^*, \beta_b^*)$  need to be distributed in such a way that one of the two choices of action puts more weight on one factor, while the other choice puts more weight on the other factor. Eventually, the behavior  $p^*$  will randomize between the two choices and achieve the growth-



optimal combination of factors. This is a generalization of the “adaptive coin-flipping” strategies described by Cooper and Kaplan (1982), who interpret this behavior as a form of altruism, because individuals who engage in this behavior seem to be acting in the interest of the population at the expense of their own individual fitness.<sup>10</sup>

The results in Table 2 also highlight the fact that evolution can lead to multiple coexisting types of individuals. It is mathematically possible that types with different characteristics ( $\beta_a$  and  $\beta_b$ ) and different behaviors ( $p$ ) will lead to the same factor loading ( $\alpha_1^*$ ). They may superficially appear to be doing very different things, but each group of individuals will balance these two actions in its own way, based on its own characteristics. The environmental factor plays an important role in this process, since the ultimate reason that these groups are able to coexist is because they have the same factor loadings. Just as “All roads lead to Rome,” our results show that in evolution, “All sustainable behaviors lead to survival,” that is, those behaviors satisfying the growth-optimality condition in Table 2.

## Binary choice model of political polarization

We first apply our framework to explain the emergence of coordinated groups, groups whose individual members appear to act with a single purpose, such as unions, military alliances, and patient advocacy groups, among others. Here, we focus on extreme political views as an example to illustrate the emergence of political polarization.

The key lies in the fact that the fitness of individuals share several common factors. The consequences of this one feature—which is the evolutionary instantiation of the adage “the enemy of my enemy is my friend”—are enormous, giving rise to seemingly coordinated behavior among subsets of individuals, or groups, purely through evolutionary dynamics.

Consider a hypothetical island isolated from the rest of the world. There are two factors that determine the fitness of any individual on this island. The first factor,  $\lambda_{\text{glob}}$ , represents the degree of globalization where, without loss of generality, we assume that larger values represent higher degrees of globalization.<sup>11</sup> The second factor,  $\lambda_{\text{other}}$ , represents everything else that may be relevant to an individual’s fitness. This is obviously an oversimplification, but more general specifications will become obvious once we present the analysis for this simpler setting.<sup>12</sup>

### A simple example

To develop intuition about the model, we first consider the special case in which the factors are specified by the following Bernoulli distribution

$$\lambda_{\text{glob}} = \begin{cases} 4, & \text{with probability } q \\ 1, & \text{with probability } 1 - q \end{cases}, \quad \lambda_{\text{other}} \equiv 2 \quad (9)$$

In each period, the degree of globalization is either 4 or 1, and the higher values of the probability  $q$  represent a higher average degree of globalization. On the other hand, we may simply assume that all other factors are represented by a constant factor  $\lambda_{\text{other}}$  without loss of generality.

An individual on this island lives for one period, has one opportunity to choose one of two political attitudes (actions)—pro-globalization or anti-globalization—that determines its fitness, and then dies immediately after reproduction. The number of offspring is given by  $x_{\text{anti}}$  if the individual chooses to be anti-globalization, and  $x_{\text{pro}}$  if the individual chooses to be pro-globalization.

$$\begin{aligned} x_{\text{anti}} &= \beta_{\text{anti}} \lambda_{\text{glob}} + (1 - \beta_{\text{anti}}) \lambda_{\text{other}} \\ x_{\text{pro}} &= \beta_{\text{pro}} \lambda_{\text{glob}} + (1 - \beta_{\text{pro}}) \lambda_{\text{other}} \end{aligned} \quad (10)$$

The characteristics  $\beta_{\text{anti}}$  and  $\beta_{\text{pro}}$  determine how an individual’s chosen action affects its fitness through the two factors. Different individuals may possess different characteristics. Here we focus on two specific types of individuals: those who benefit from globalization, and those who are harmed by it. Higher values of  $\lambda_{\text{glob}}$  are more conducive to fitness for those who benefit from globalization, therefore yielding a positive characteristic if the individual chooses to be pro and embrace globalization. We use  $\beta_{\text{pro}}^{\text{benefit}} = 1$  to represent this characteristic. On the other hand, if the individual chooses to be anti, they do not benefit from globalization, and their fitness is purely determined by other factors. We use  $\beta_{\text{anti}}^{\text{benefit}} = 0$  to represent this characteristic.

On the other hand, for those who are harmed by globalization, choosing to be anti and supporting policies that limit globalization can promote their fitness when the level of globalization is high. Therefore, they have a positive characteristic  $\beta_{\text{anti}}^{\text{harm}} = 1$ . In contrast, when they choose to be pro, their fitness is purely determined by other factors:  $\beta_{\text{pro}}^{\text{harm}} = 0$ .

To summarize, we use the superscript “benefit” or “harm” to represent these two types of individuals, and their fitness is determined by

$$\begin{aligned} x_{\text{anti}}^{\text{benefit}} &= \lambda_{\text{other}}, & x_{\text{anti}}^{\text{harm}} &= \lambda_{\text{glob}} \\ x_{\text{pro}}^{\text{benefit}} &= \lambda_{\text{glob}}, & x_{\text{pro}}^{\text{harm}} &= \lambda_{\text{other}} \end{aligned} \quad (11)$$

The behavior  $p$  in this example represents the probability of choosing the “anti-globalization” action. In other words, lower values of  $p$  corresponds to more “pro-globalization” behaviors. We have the following result characterizing the growth-optimal behavior in this example:

**Proposition 2.** Under assumptions (A1) and (A2) and the environment specified by (9) and (10), the population growth rate in (4) can be evaluated explicitly as

$$\begin{aligned} \mu^{\text{benefit}}(p) &= q \log(4 - 2p) + (1 - q) \log(1 + p), \\ \mu^{\text{harm}}(p) &= q \log(2 + 2p) + (1 - q) \log(2 - q), \end{aligned} \quad (12)$$

and the behavior (value of  $p$ ) that maximizes this growth rate is

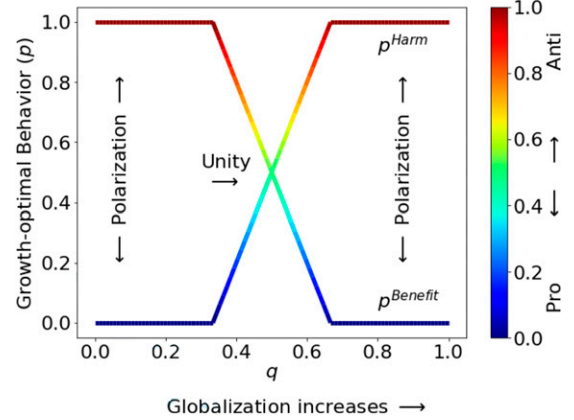
$$p^{\text{benefit}} = \begin{cases} 1, & \text{if } q \leq \frac{1}{3} \\ 2 - 3q, & \text{if } \frac{1}{3} < q < \frac{2}{3} \\ 0, & \text{if } q \geq \frac{2}{3} \end{cases} \quad (13)$$

$$p^{\text{harm}} = \begin{cases} 0, & \text{if } q \leq \frac{1}{3} \\ 3q - 1, & \text{if } \frac{1}{3} < q < \frac{2}{3} \\ 1, & \text{if } q \geq \frac{2}{3} \end{cases} \quad (14)$$

We plot  $p^{\text{benefit}}$  and  $p^{\text{harm}}$  in Figure 1. As the average degree of globalization ( $q$ ) increases, the growth-optimal behavior for individuals who benefit from globalization ( $p^{\text{benefit}}$ ) (that is, leaning pro) decreases, while the growth-optimal behavior for individuals who are harmed by globalization ( $p^{\text{harm}}$ ) (that is, leaning anti) increases. This is due to the fact that as selection pressure on the globalization factor increases, these two groups of individuals are forced by the environment to choose the political views that benefit their respective interests, that is, fitness.<sup>13</sup>

This example illustrates a primitive form of polarization. When the average degree of globalization is either too low or too high, two distinct groups of individuals emerge. They coexist through the evolutionary process, but within each group, individuals share the same characteristics. A particular behavior must be paired with a particular set of characteristics to achieve the optimal growth rate. Note that the individuals in (13) and (14) are optimal only in the group sense. In fact, from any individual's perspective, the survival-maximizing behavior is to always choose the action with higher average fitness ( $p = 0$  or  $1$ ). The continuous spectrum of growth-optimal behaviors in Figure 1 only emerges because a group possesses survival benefits above and beyond an individual. In our framework, these benefits arise purely from stochastic environments with systematic risk.<sup>14</sup>

The usual conception of group selection in the evolutionary biology literature is that natural selection acts at the level of the group, instead of at the more conventional level of the individual (or the gene), and that interaction between members within each group is much more frequent than interaction



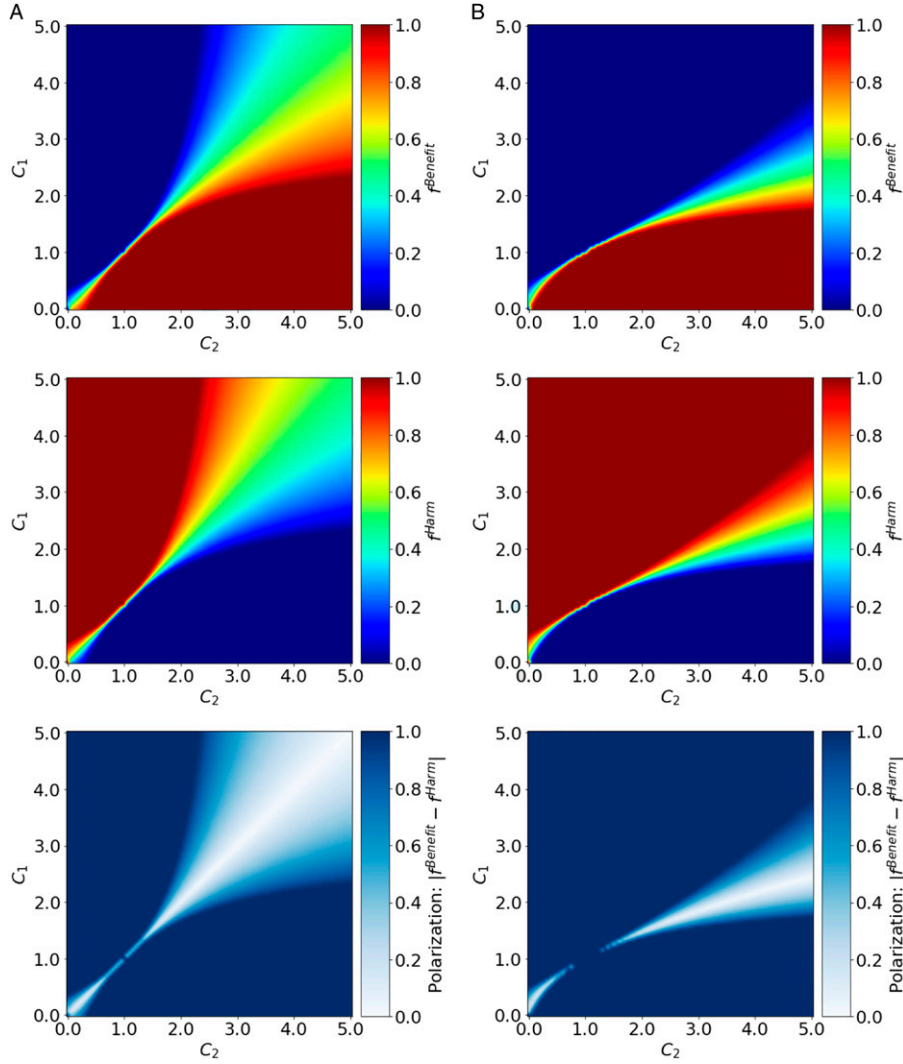
**Figure 1.** Growth-optimal behavior  $p^{\text{benefit}}$  for individuals who benefit from globalization, and  $p^{\text{harm}}$  for individuals who are harmed by globalization. The horizontal axis shows the probability  $q$  in (9). The vertical axis and the color bar show the growth-optimal behavior,  $p^*$ , in different environments parameterized by  $q$ . Blue indicates the “pro-globalization” action, while dark red indicates the “anti-globalization” action.

among individuals across groups. In this case, similar individuals are usually clustered geographically. However, in our model, individuals do not interact at all. Nevertheless, the fact that individuals with the same behavior generate offspring with like behavior makes them more likely to cluster geographically and appear as a “group.”

In reality, the environment is generally nonstationary. Factor distributions change over time, and old factors fade while new factors emerge. In fact, the change in the environment can itself be a consequence of previous adaptations. We see this in the history of globalization itself. From the Silk Road dating back to the 2nd century BCE, to the World Trade Organization established in 1995, the course of globalization has always been fueled by a number of historical factors, such as the desire to trade local goods for exotic products, or to gain access to cheap labor. Imagine that the environment ( $\lambda_{\text{glob}}$ ,  $\lambda_{\text{other}}$ ) experiences a sudden shift. To an outside observer, behaviors among individuals in this population will become increasingly similar after the shift, creating the appearance—but not necessarily the reality—of intentional coordination, communication, and synchronization. If the reproductive cycle is sufficiently short, this change in population-wide behavior may seem highly responsive to environmental changes, giving the impression that individuals are learning about their environment. This is indeed a form of learning, but it occurs at the population level—a form of collective learning—not at the individual level, and not within an individual's lifespan.

### The general case

The factor distribution in (9) can be easily generalized to any arbitrary number of offspring



**Figure 2.** Growth-optimal behaviors for both the “Benefit” group and the “Harm” group,  $f^{\text{Benefit}}$  and  $f^{\text{Harm}}$ , as functions of environmental parameters. (2a): moderate globalization with  $q = 0.5$ . (2b): high globalization with  $q = 0.9$ . The first row shows  $f^{\text{Benefit}}$ ; the second row shows  $f^{\text{Harm}}$ ; the last row shows the absolute difference, that is, polarization:  $|f^{\text{Benefit}} - f^{\text{Harm}}|$ .

$$\lambda_{\text{glob}} = \begin{cases} C_1, & \text{with probability } q \\ 1, & \text{with probability } 1 - q \end{cases} \quad (15)$$

$$\lambda_{\text{other}} = \begin{cases} C_2, & \text{with probability } r \\ 1, & \text{with probability } 1 - r \end{cases}$$

We assume, without loss of generality, that one of the outcomes for each factor yields exactly one offspring while the other is parameterized by  $C_1$  and  $C_2$ , since it is the relative fitness between these two outcomes that matters. In addition, probabilities  $q$  and  $r$  parameterize the average level of the two factors. In Figure 2, we show the growth-optimal behavior for both the “Benefit” group and the “Harm” group,  $f^{\text{Benefit}}$  and  $f^{\text{Harm}}$ , as functions of these environmental parameters.

Figure 2(a) shows the case with a moderate level of globalization over time ( $q = 0.5$ ). The plot in the first row

shows the growth-optimal behavior for those who benefit from globalization ( $f^{\text{Benefit}}$ ). As the fitness for the globalization factor ( $C_1$ ) increases, individuals tend to be pro (blue), but as the fitness for the other factor ( $C_2$ ) increases, individuals tend to be anti (dark red). The plot in the second row shows the growth-optimal behavior for those who are harmed by globalization ( $f^{\text{Harm}}$ ), which are the opposite of the behaviors for the “Benefit” group, in the sense that  $f^{\text{Harm}} = 1 - f^{\text{Benefit}}$ . The plot in the last row shows the absolute difference between the growth-optimal behaviors of the two groups of individuals,  $|f^{\text{Benefit}} - f^{\text{Harm}}|$ , which is a simple measure of polarization. When the “Benefit” group and the “Harm” group show opposing behaviors, the level of polarization is high (dark blue).

Figure 2(b) shows the same set of growth-optimal behaviors when the average level of globalization is



high ( $q = 0.9$ ). Compared to the behaviors in Figure 2(a), when the average globalization shifts toward a higher level, behaviors shift accordingly as well. As a result, the same environmental conditions (the region of the  $(C_1, C_2)$ -plane) that generated unity before may lead to polarization in this environment.

The simple example here considers two groups of individuals: those who benefit from globalization ( $\beta_{\text{pro}}^{\text{benefit}} = 1, \beta_{\text{anti}}^{\text{benefit}} = 0$ ) and those harmed by globalization ( $\beta_{\text{pro}}^{\text{harm}} = 0, \beta_{\text{anti}}^{\text{harm}} = 1$ ). In this stylized example, both groups coexist while different political views emerge. In reality, there is a spectrum of individuals in the population who benefit from or are harmed by globalization to varying degrees. This corresponds to a continuum of characteristics ( $\beta$ ) associated with the globalization and “other” factors. As a result, the population will consist of a more diverse set of political views, spanning the entire range from pro to anti. The ultimate political composition in the population is determined by the mixture of individuals with different characteristics.

## Binary choice model of bias and discrimination

Our framework can also be used to understand the emergence of bias and discrimination, as well as to determine their underlying causes and what can be done to counteract these causes. We use racial discrimination as the main example of bias in this section, but the same principles apply more broadly to other kinds of bias and discrimination, including gender, sexual orientation, religion, socioeconomic strata, and so on.

### A simple example

We consider a hypothetical world with a population composed of two racial groups: a majority group which we refer to as the “Andorians,” and a minority group which we refer to as the “Tellarians.” Group membership is unambiguous, mutually exclusive (an individual is a member of one and only one group), immutable, and observable by all.<sup>15</sup> There are two factors that determine each individual’s fitness:  $\lambda_A$  and  $\lambda_T$ . They represent social interactions with Andorian and Tellarian individuals, respectively. An individual who interacts with Andorian individuals is subject to the Andorian factor,  $\lambda_A$ , whereas an individual who interacts with Tellarian individuals is subject to the Tellarian factor,  $\lambda_T$ .  $\lambda_A$  and  $\lambda_T$  are independent random variables with the following distributions

$$\lambda_A = \begin{cases} 1, & \text{with probability } q \\ 2, & \text{with probability } 1 - q \end{cases}, \quad (16)$$

$$\lambda_T = \begin{cases} 1, & \text{with probability } r \\ 2, & \text{with probability } 1 - r \end{cases}$$

Without loss of generality, we have assumed that each factor only takes two possible values: a low fitness of 1, which happens in the context of an adverse event related to that group,<sup>16</sup> and a high fitness of 2, which represents the normal case. Here, we use  $q$  and  $r$  to represent the probability of the adverse event for the Andorian and the Tellarian groups, respectively, which we refer to as the “adverse probability” for simplicity. For example, with a (small) probability  $r$ , if an adverse event happens in an interaction with the Tellarian individual, anyone with an interaction with that individual will experience low fitness in that period.

Historically, the Tellarian community has been politically underrepresented, with less access to education and economic opportunity. As a result, this greater inequality has led to a higher crime rate for the Tellarian community compared to the average population. Note that the higher crime rate is not *because* of race, but the result of a complicated set of determinants, including less access to resources historically. However, in this model, individuals observe only each other’s race, modeled here as group membership, which they use as a marker in the absence of any other information. The true underlying causes of higher crime rates, such as a lack of educational opportunity or socioeconomic status, are assumed to be unobservable, a key assumption.

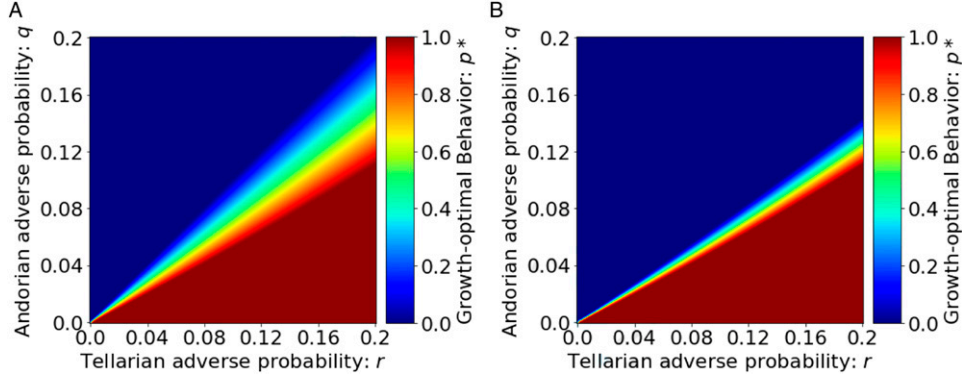
We now focus on the perspective of an Andorian, who faces a decision between one of two actions—whether or not to discriminate against a Tellarian—which determines their fitness. We assume that an Andorian’s number of offspring is given by  $x_{\text{discriminate}}$  if the individual chooses to discriminate, and  $x_{\text{not discriminate}}$  if the individual chooses not to discriminate

$$\begin{aligned} x_{\text{discriminate}} &= \lambda_A \\ x_{\text{not discriminate}} &= \beta\lambda_T + (1 - \beta)\lambda_A \end{aligned} \quad (17)$$

If an Andorian chooses to discriminate against a Tellarian, it avoids any interactions with that individual, and therefore, its fitness will be subject only to  $\lambda_A$ . On the other hand, if an Andorian does not discriminate, its fitness is subject to both  $\lambda_T$  and  $\lambda_A$ . Here,  $\beta$  represents the percentage of Tellarians in the population, hence the weight on the factor  $\lambda_T$ .<sup>17</sup>

For a particular behavior  $p$  (the probability to discriminate against a Tellarian), the population growth rate in (4) is a function of the environment (that is, the adverse probabilities,  $q$  and  $r$ ) and the characteristic ( $\beta$ ). In this simple case, as in the example of political polarization in the previous section, we can characterize the growth-optimal behavior explicitly:

**Proposition 3.** *Under assumptions (A1) and (A2) and the environment specified by (16) and (17), the population growth rate can be evaluated explicitly as*



**Figure 3.** Growth-optimal behaviors,  $p^*$ , as a function of environmental parameters. (3a): percentage of Tellarians in the population  $\beta = 0.5$ . (3b): percentage of Tellarians in the population  $\beta = 0.2$ .

$$\begin{aligned} \mu(p) = & q(1-r)\log(1+\beta-p\beta) \\ & + (1-q)r\log(1-\beta+p\beta) \\ & + (1-q)(1-r)\log(2-p), \end{aligned} \quad (18)$$

and the behavior (that is, the value of  $p$ ) that maximizes this growth rate is

$$p^* = \begin{cases} 1, & \text{if } r \geq \frac{2q}{1+q} \\ 1 - \frac{qr - 2q + r}{(2qr - q - r)\beta}, & \text{if } \frac{(2-\beta)q}{(1-2\beta)q + (1+\beta)} < r < \frac{2q}{1+q} \\ 0, & \text{if } r \leq \frac{(2-\beta)q}{(1-2\beta)q + (1+\beta)} \end{cases} \quad (19)$$

Equation (19) is the behavior that yields the highest growth rate and therefore characterizes the behavior favored by natural selection. Recall that  $p^* = 1$  corresponds to fully discriminatory behavior. We plot  $p^*$  in Figure 3 with two different population group percentages. Figure 3(a) shows a world with an equal number of Andorian and Tellarian individuals ( $\beta = 0.5$ ), and Figure 3(b) shows a world with only 20% Tellarians ( $\beta = 0.2$ ).

In both cases, when the adverse probability associated with Tellarians ( $r$ ) is low compared to the adverse probability associated with Andorians ( $q$ ), no discrimination emerges. As  $r$  increases relative to the adverse probability for Andorians ( $q$ ), discrimination emerges, that is,  $p^*$  increases from 0 to 1. This is because individuals who choose to avoid interactions with Tellarians gain an evolutionary advantage by reducing their exposure to the factor  $\lambda_T$  and the higher adverse probability  $r$  on average. This effect emerges from the fact that in our model, race is the only observable marker of the individuals in the population and the true underlying causes of the higher adverse probability are not observable. This phenomenon is also referred to as statistical discrimination (Phelps, 1972; Arrow, 1973).

In addition, we can observe from the first case of (19) that the environment leading to full discrimination ( $p^* = 1$ ) does

not depend on the percentage of Tellarians in the population ( $\beta$ ). It is only a function of the adverse probability,  $q$  and  $r$ . This is also clear by comparing Figure 3(a) and (b). In both cases, when the adverse probability associated with Tellarians is high compared to that for Andorians ( $r \geq \frac{2q}{1+q}$ ), full discrimination emerges.

On the other hand, when Tellarians are the minority ( $\beta = 0.2$ ), the region where individuals have partially discriminatory behavior shrinks (given by the middle case in (19), where  $p^*$  is strictly between 0 and 1). This implies that when the group in consideration consists of a small fraction of the entire population, the boundary of the environmental conditions leading to no discrimination and full discrimination is sharper.

In our simple example, the key to the emergence of discrimination is the fact that race is the only observable feature of individuals. However, these implications will likely remain true even if other attributes of the individuals are partially observable, given the insight of the memory/prediction framework by Hawkins and Blakeslee (2004), who argue that we store memory patterns and use them to predict what will happen in the future. When individuals experience a random adverse event in association with a Tellarian, they tend to attribute it to the Tellarian's race because it is the most easily observable marker, leading to discrimination against Tellarians. Based on a similar hypothesis, Bordalo et al. (2016) develop a model of stereotyping based on the representativeness heuristic (Tversky and Kahneman, 1983): agents overweight the prevalence of a trait in a group when that trait appears to be highly representative of the group in question. This is, however, not the root cause of the adverse event. In other words, it is much too easy to confuse correlation with causation.

We have seen that the difference in relative adverse probabilities,  $q$  and  $r$ , can lead to serious biases and discriminatory practices. Next, we are able to strengthen our results by showing that even when the two groups have equal probabilities of adverse events, or even in certain cases when Tellarian individuals have a lower probability of

adverse events than their Andorian counterparts, discrimination can still emerge.

### Feedback loops

Discrimination against Tellarians in the general population affects the Tellarian community adversely. For example, those individuals who participate in discriminatory behavior against Tellarians may contact law enforcement more often, leading to a higher incidence of false accusations against the Tellarian community. They may develop more hostile behaviors toward the Tellarian community, reducing educational and economic opportunities for the Tellarian community, which further increases the probability of an adverse event associated with Tellarians.

Another less obvious type of feedback comes from the increasing popularity and prevalence of engagement-based recommender systems on news and social media platforms. When presented with new information (which may be a news broadcast or a social media post), humans tend to anchor towards what they originally believe (Tversky and Kahneman, 1974). As a result, even a small initial bias acquired randomly can be reinforced and amplified through feedback based on a recommender algorithm.

To incorporate this feedback loop into our model, we make the following assumption:

(A3) Factor  $\lambda_T$ 's distribution in generation  $T$  is given by

$$\lambda_T = \begin{cases} 1, & \text{with probability } \tilde{r} := r(1 + \tau\bar{p}_{T-1}) \\ 2, & \text{with probability } 1 - \tilde{r} \end{cases} \quad (20)$$

where  $\bar{p}_{T-1}$  represents the average behavior in the population in the previous generation,  $T - 1$ .

When the level of bias is higher in the population (that is, when  $\bar{p}_{T-1}$  is higher), the adverse probability associated with Tellarians ( $\tilde{r}$ ) is higher. Here,  $\tau$  represents the intensity of the feedback effect. For example, when  $\tau = 1$ , the adverse probability is, at most, twice when everyone discriminates against Tellarians, compared to when no one discriminates. A higher value of  $\tau$  implies higher multiples of this effect.

Note that the factors in (16) are identically distributed over time. In other words, they do not depend on time, nor on realizations of the past evolution of results. In contrast, the factor in (20) introduces path dependency into the evolutionary process, because it depends on the past realizations of population behavior. As a result,  $\lambda_T$  is no longer stationary over time. This simple change generates a surprisingly rich set of new implications.

We first use simulation methods to develop an intuition for the effect of different intensities of negative feedback. We consider a world that starts from an equal number of

individuals in the population with 11 different behaviors:  $p \in \{0, 1/10, 2/10, \dots, 1\}$ . Figure 4 shows the evolution of the relative frequency of these behaviors over 10,000 generations, given different environmental conditions.

Figure 4(a)–(c) depict simulations of an environment with equal adverse probabilities for Tellarians and Andorians ( $q = r = 0.2$ ), with the feedback intensity,  $\tau$ , increasing from 0 (no feedback) to 1 (the adverse probability is doubled with full discrimination in the population).<sup>18</sup> Figure 4(a) corresponds to an environment with no feedback, and the behavior  $p^* = 0$  (no discrimination) quickly dominates the population. This also corresponds to the growth-optimal behavior in the upper right corner of Figure 3(a). As the feedback intensity increases to  $\tau = 0.6$ , as shown in Figure 4(b), positive  $p^*$  (partial discrimination) emerges. Finally, as the feedback intensity increases to  $\tau = 1$ , as shown in Figure 4(c),  $p^* = 1$  (full discrimination) quickly dominates the population.

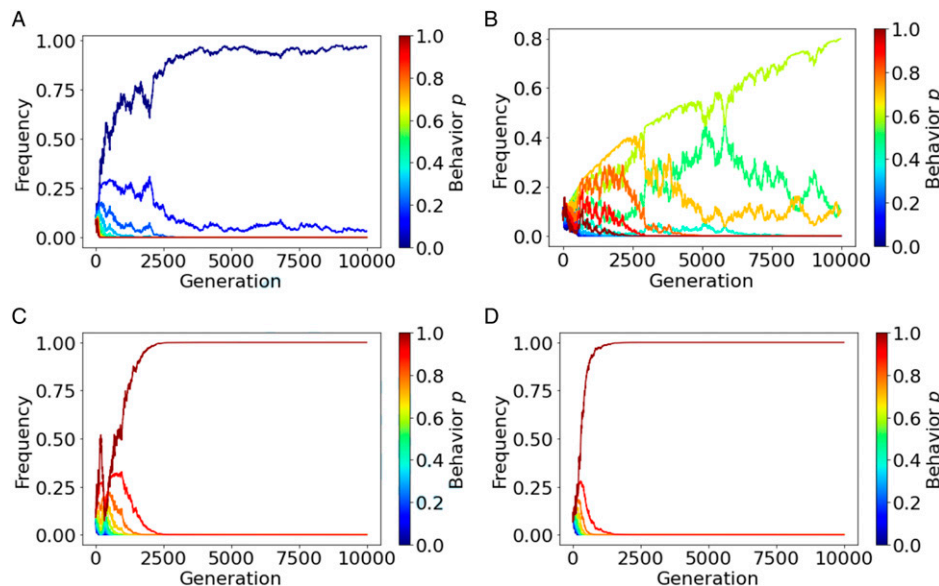
In addition, Figure 4(d) illustrates an environment in which Tellarians have a lower probability of an adverse event than Andorians ( $r < q$ ). Given conditions of strong feedback ( $\tau = 2$ ), fully discriminatory behavior ( $p^* = 1$ ) still dominates the population. This is because the feedback intensity is so high that discrimination quickly worsens the adverse probability for the Tellarian population, leading to severe discrimination against the population, despite the fact that the Tellarian population starts with a more favorable adverse probability.<sup>19</sup>

More generally, despite the challenging complexities of a nonstationary and path-dependent environment created by the feedback mechanism, we can analytically quantify the growth-optimal behavior,  $p^*$ , implicitly. The factor with feedback in (20) is mathematically equivalent to the simple environment we considered in (16), except that the adverse probability associated with Tellarians,  $r$ , is replaced by the feedback-adjusted adverse probability,  $\tilde{r}$ . Therefore, a behavior can survive in the long run only when it satisfies the growth-optimal condition (19), with  $r$  replaced by the feedback-adjusted  $\tilde{r}$ , hence we have:

**Proposition 4.** *Under assumptions (A1)–(A3) and the environment specified by (17), the growth-optimal behavior,  $p^*$ , with feedback must satisfy the following fixed-point condition*

$$\begin{aligned} p^* &= \text{Bound}_0^1 \left( 1 - \frac{q\tilde{r} - 2q + \tilde{r}}{(2q\tilde{r} - q - \tilde{r})\beta} \right) \\ &= \text{Bound}_0^1 \left( 1 - \frac{(q+1)r(1 + \tau p^*) - 2q}{[(2q-1)r(1 + \tau p^*) - q]\beta} \right) \end{aligned} \quad (21)$$

where  $\text{Bound}_0^1(x) = \max(0, \min(1, x))$  represents a function that bounds the behavior to lie within the closed unit interval.



**Figure 4.** The evolution of 11 behaviors,  $p \in \{0, 1/10, 2/10, \dots, 1\}$ , over 10,000 generations. The vertical axis represents the relative frequency of each behavior, and the horizontal axis represents time. (4a): equal adverse probability ( $q = r = 0.2$ ), no feedback ( $\tau = 0$ ); (4b): equal adverse probability ( $q = r = 0.2$ ), mild feedback ( $\tau = 0.6$ ); (4c): equal adverse probability ( $q = r = 0.2$ ), more feedback ( $\tau = 1$ ); (4d): lower Tellarian adverse probability ( $q = 0.2, r = 0.15$ ), even more feedback ( $\tau = 2$ ).

Equation (21) is a necessary, but insufficient, condition for any behavior to survive in the long run. Due to its nonlinearity, the growth-optimal behavior,  $p^*$ , implied by (21) may not be unique for some environments. However, without intervention, only one behavior is stable and able to persist in each environment, for which we need to define the new notion of a locally evolutionarily stable strategy.

### Locally evolutionarily stable strategies

An evolutionarily stable strategy (ESS), first introduced by Maynard Smith and Price (1973),<sup>20</sup> is a strategy that is impermeable to other strategies when adopted by a population in adaptation to a specific environment. In other words, it cannot be displaced by an alternative strategy, which may be novel or initially rare. In game-theoretical terms, an ESS is an equilibrium refinement of the Nash equilibrium concept, given that a Nash equilibrium is also “evolutionarily stable.” Once fixed in a population, natural selection alone is sufficient to prevent alternative (or mutant) strategies from replacing it.

We define a locally evolutionarily stable strategy (L-ESS) to be one that is stable locally. In other words, it is a strategy that cannot be displaced by any local perturbation of that strategy.<sup>21</sup>

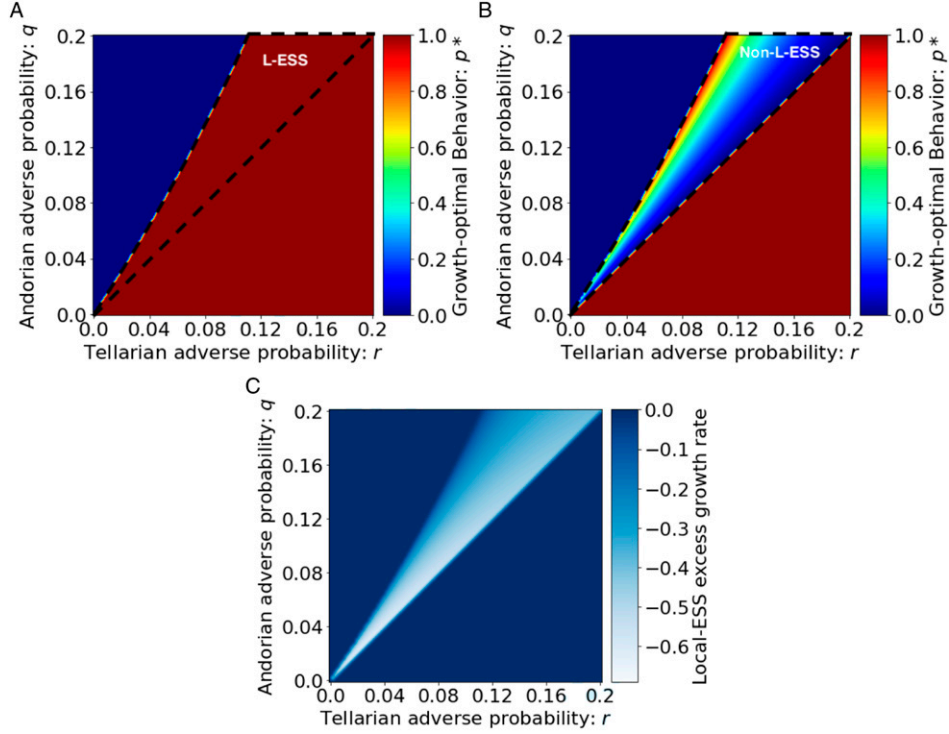
**Definition 1.** A *L-ESS behavior*,  $p^*$ , is one for which any local perturbation in the average population behavior  $\{\bar{p} : \bar{p} = p^* + \varepsilon\}$  leads to a growth-optimal behavior,  $p'$ , given by (19) that is closer to  $p^*$  than  $\bar{p}$ ,  $|p' - p^*| < |\varepsilon|$ .

In other words, when randomness in the environment causes the average behavior of the population,  $\bar{p}$ , to change around the growth-optimal behavior,  $p^*$ , the perturbed  $\bar{p}$  will lead to a new behavior that is very close to the original growth-optimal behavior,  $p^*$ . As a result, evolutionary dynamics itself will always bring the population back to the original growth-optimal behavior,  $p^*$ . In this sense, such behaviors are locally stable from an evolutionary perspective. The L-ESS is an additional requirement to the growth-optimal behavior,  $p^*$ , implied by the fixed-point condition (21). Without intervention, only L-ESS behaviors can persist in the long run. When there is no or little feedback in the environment (that is, when  $\tau$  is small), behaviors implied by (21) are always L-ESS. However, as the feedback intensity increases, non-L-ESS behaviors can emerge.

Figure 5 shows an environment with strong feedback intensity ( $\tau = 2$ ). Recall that the nonlinearity of the fixed-point condition (21) can lead to multiple solutions of  $p^*$ , and we compare the L-ESS (Figure 5(a)) and non-L-ESS behaviors (Figure 5(b)). The dashed triangular regions<sup>22</sup> represent the set of environments where the fixed-point condition (21) leads to one L-ESS and one non-L-ESS behavior. In this region, the non-L-ESS behaviors are less discriminatory, and the strong feedback intensity nudges the population to evolve towards fully discriminatory behaviors.

In addition, Figure 5(c) shows the differences in population growth rates between the L-ESS behavior and





**Figure 5.** Comparison of L-ESS and non-L-ESS behaviors for an environment with strong feedback intensity ( $\tau = 2$ ). (5a): L-ESS growth-optimal behaviors implied by the fixed-point equation (21); (5b): non-L-ESS growth-optimal behaviors if the fixed-point equation (21) yields multiple solutions, otherwise we plot the unique solution from (21) which is L-ESS; (5c): L-ESS excess growth rate as defined in (22), which is the difference in growth rates between the L-ESS behavior and the non-L-ESS behavior.

non-L-ESS behavior. We refer to this as the “L-ESS excess growth rate”

ignorance” that could otherwise be improved with greater diversity in the population.<sup>23</sup>

$$\text{L-ESS excess growth rate} = \begin{cases} \mu(p_{\text{L-ESS}}^*) - \mu(p_{\text{non-L-ESS}}^*), & \text{if (21) yields multiple solutions} \\ 0, & \text{otherwise} \end{cases} \quad (22)$$

**Proposition 5.** Under assumptions (A1)–(A3) and the environment specified by (17), if (21) yields multiple solutions where the L-ESS behavior is more discriminative than the non-L-ESS behavior ( $p_{\text{L-ESS}}^* > p_{\text{non-L-ESS}}^*$ ), the L-ESS excess growth rate is always negative, which means that the L-ESS behavior will always yield a lower growth rate than non-L-ESS behavior.

This example demonstrates that path dependency can lead to evolutionary outcomes with slower growth rates than otherwise achievable, and the population ends up with a suboptimal growth rate compared to a world without feedback.

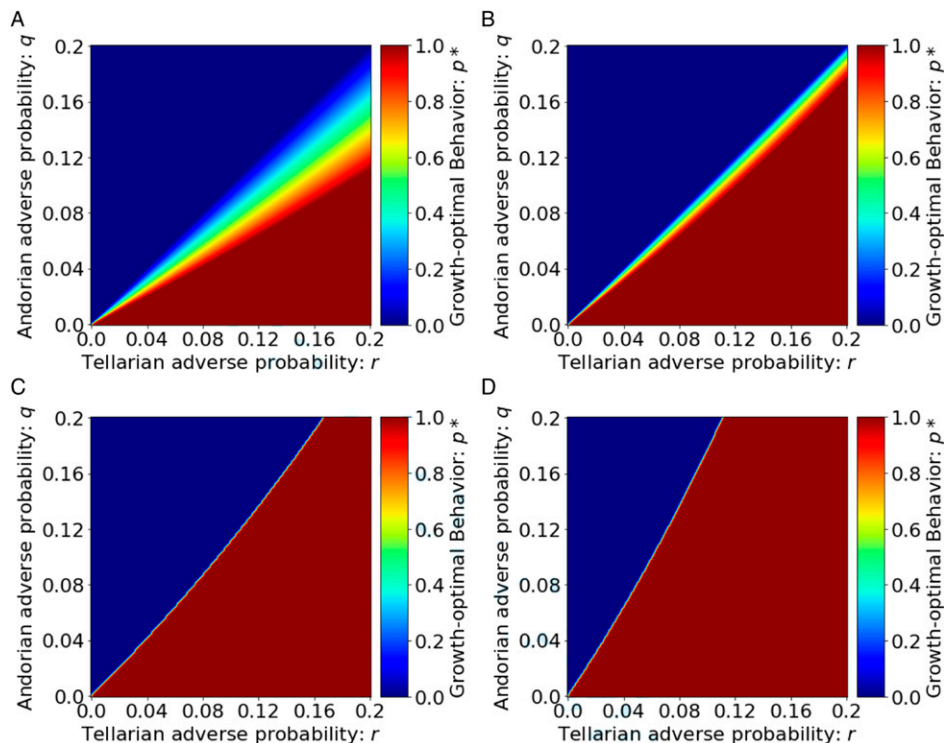
In the context of our model, L-ESS behavior implies that the Andorian individual will always avoid any interaction with the Tellerian individual, a state of “collective

### Feedback can lead to greater bias

With our understanding of L-ESS behavior, we can now finally show the variation in growth-optimal behavior in environments with different feedback intensities. We have the following intuitive but important result:

**Proposition 6.** Under assumptions (A1)–(A3) and the environment specified by (17), as the feedback intensity,  $\tau$ , increases, discriminatory behaviors are more likely to emerge, in the sense that they dominate the population for increasingly larger regions of environmental conditions, as parameterized by the adverse probabilities,  $q$  and  $r$ , for the Andorian and the Tellerian groups, respectively.





**Figure 6.** L-ESS behaviors,  $p^*$ , as a function of environmental parameters, when there is feedback. The feedback intensity,  $\tau$ , increases from 0 in (Figure 6(a)) to 2 in (Figure 6(d)).

Figure 6 shows L-ESS behaviors for different levels of feedback intensity and demonstrates Proposition 6. As  $\tau$  increases from 0 (Figure 6(a)) to 2 (Figure 6(d)), discriminatory behaviors dominate the population for increasingly larger regions of environmental conditions. When feedback is absent from these evolutionary dynamics (Figure 6(a)), discrimination only emerges when Tellarians have a higher probability of adverse events than Andorians. However, when the feedback intensity is high (Figures 6(c) and (d)), full discrimination prevails, even in environments where the adverse probability for Tellarians is lower than that for Andorians.

These results emphasize the central role feedback plays in the emergence of bias and discrimination. By combining the observation that individuals tend to attribute the occurrence of random adverse events to the only observable characteristic, race (Hawkins and Blakeslee, 2004), with the negative feedback from those random adverse events, our model has demonstrated the power of these forces in generating widespread bias and discrimination in the population.

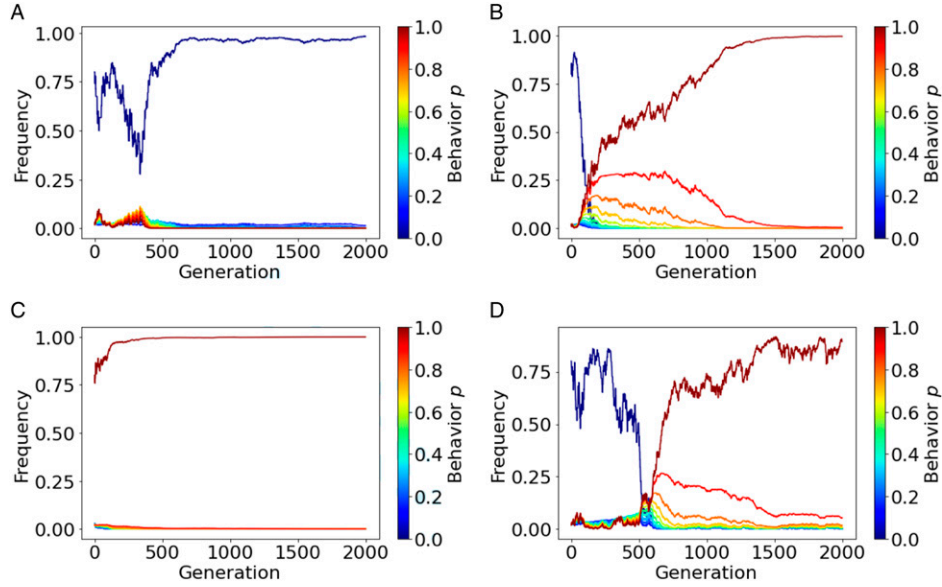
These results shed light on the evolutionary dynamics behind the emergence of biases not only toward the Tellarian community (which is of course fictional), but also other forms of bias and discrimination. From the policy perspective, these results emphasize the importance of preventing the effects of negative feedback in the greater population. One example is to proactively provide

more educational and economic opportunities among disadvantaged groups. This does not directly eliminate the negative feedback, but will indirectly help to reduce its impact by elevating their socioeconomic status and reducing their adverse probabilities. Another example is to enforce regulations that cut through such (sometimes unconscious) negative feedback mechanisms. These actions together will create more favorable environments for collective intelligence to emerge rather than allowing collective ignorance to propagate, and can potentially reduce, and eventually reverse, selection pressure behind the emergence of bias and discrimination.

### Path-dependent evolution and initial conditions

When feedback loops exist in the environment, evolution may become path dependent. Therefore, the dominant behavior that emerges in a given population will sometimes depend on the initial composition of that population. We consider evolution in populations that begin with non-uniform initial distributions of behaviors in this section.

Figure 7 demonstrates that two realizations of an evolutionary system under the same environment can lead to different growth-optimal behaviors, and different initial populations can also lead to different growth-optimal behaviors. Like the simulations illustrated in Figure 4, we simulate the evolution of 11 behaviors,  $p \in \{0, 1/10, 2/10,$



**Figure 7.** Path dependency of evolution in an environment with equal adverse probability ( $q = r = 0.2$ ) and feedback  $\tau = 1$ . We show the evolution of 11 behaviors,  $p \in \{0, 1/10, 2/10, \dots, 1\}$ , over time, with different starting populations. The vertical axis represents the relative frequency of each behavior, and the horizontal axis represents time. (7a): the initial population has low discrimination,  $n_0 = (0.8, 0.02, 0.02, \dots, 0.02)$ ; (7b): a different simulation run with the same conditions as in (7a); (7c): the initial population has high discrimination,  $n_0 = (0.02, 0.02, \dots, 0.02, 0.8)$ ; (7d): the initial population has low discrimination,  $n_0 = (0.8, 0.02, 0.02, \dots, 0.02)$ , and behaviors have a 0.1% mutation rate.

$\dots, 1\}$ , for an environment with equal adverse probabilities for the Tellarian and Andorian populations ( $q = r = 0.2$ ), and with a feedback intensity  $\tau = 1$ .

We use  $n_0$  to denote the frequency of different behaviors in the initial population. Figure 7(a) and (b) show two simulation runs of the evolution for an initial population with little bias:  $n_0 = (0.8, 0.02, 0.02, \dots, 0.02)$ . In other words, 80% of the initial population starts with no discrimination ( $p = 0$ ). After 2000 generations,  $p = 0$  dominates the population in the first case, whereas  $p = 1$  dominates in the second case.

In contrast, Figure 7(c) shows the evolution for an initial population with a substantial amount of bias:  $n_0 = (0.02, 0.02, \dots, 0.02, 0.8)$ , hence 80% of the initial population starts with fully discriminatory behavior ( $p = 1$ ). Not surprisingly,  $p = 1$  dominates.

When the initial population is non-uniform, some behaviors may quickly become extinct before they have a chance to spread. In fact, if we allow a small amount of mutation in each generation—modeled as in Brennan et al. (2018), that is, with some small probability, for example, 0.1%, that offspring of type- $p$  parents will be, in fact, type  $p' \neq p$  where  $p'$  is uniformly distributed in  $[0, 1] - \{p\}$ —discrimination will again dominate, even if the initial population begins with very little bias. Figure 7(d) shows such an example.<sup>24</sup>

This result underscores the fact that public policy may be able to guide a society towards different outcomes by

purposefully imposing a strong prior belief onto the population. This may be achievable by encouraging fairer beliefs through early education, and by providing more accurate portrayals of other cultures to counteract inaccurate stereotypes. From the perspective of our binary choice model, these policies would nudge the initial population such that its subsequent evolution may lead to a less discriminatory society collectively.

## Discussion

We present an evolutionary framework based on a binary choice model subject to evolutionary dynamics and stochastic environments that affects the fitness of a differentiated population. This framework yields collective intelligence in the form of sophisticated rational behaviors that emerge out of an initial population in which all possible behaviors are equally represented (Brennan and Lo, 2011, 2012). Within the same model, we can also specify conditions under which this collective intelligence breaks down, especially under conditions where agents face correlated fitness, or in the presence of path-dependent feedback. This offers one explanation of the emergence of political polarization, bias, and racial discrimination.

The root cause of these failures is complexity, particularly with respect to population heterogeneity, stochastic environments, and feedback mechanisms. Yet it is precisely

in such complex environments that we are in most need of collective intelligence. Our results show that it is the complexity within the evolutionary process—not the complexity of the task (the task in our model is a simple binary choice)—that can undermine collective intelligence, which is far more subtle and challenging a problem.<sup>25</sup>

Of course, our model has several limitations and is by no means a complete description of reality. Even a partial description would involve the interplay between sophisticated human behavior and highly complex nonstationary environments with multiple unknown factors. However, our approach offers a starting point for describing and understanding the fundamental principles behind the emergence of these failures of collective intelligence. A natural next step for future research is to develop more realistic models and conditions under which such failures can be expected.

Some of the biggest challenges facing humanity can only be solved through a collective and global effort. They include not only dealing with political polarization and discrimination, but also climate change, various life-threatening diseases, economic and social inequality, and the spread of disinformation. Extensions of our framework may help to explain the spread of disinformation and belief polarization, another example of the failure of collective intelligence (Haghtalab et al., 2021). This is closely related to political polarization and racial discrimination because the spread of disinformation facilitates the formation of these biases. With the advent and popularity of engagement-based recommender systems on news and social media platforms, disinformation has a much greater chance of propagating across the population. One of the great insights of Tversky and Kahneman (1974) is that humans tend to anchor towards their original beliefs. When first presented with new information, either through a news service, or simply a Twitter post, regardless of its authenticity, there will always be a group of people who happen to share a similar belief, even if that belief is false. Regardless of the small size of this initial group, through engagement-based recommendations their beliefs can be amplified rapidly throughout the population. This effect, in turn, will cause recommender algorithms to serve up similar information more frequently, reinforcing these false beliefs in a vicious cycle.

So how can we prevent failures of collective intelligence? Our evolutionary perspective suggests that the key is to foster environments under which the desired behavior—collective intelligence—will emerge naturally through evolutionary dynamics, instead of simply regulating against the undesired outcome which could create selective pressures that make matters worse. In our example of globalization, the fundamental cause of the emergence of polarization is the sharp difference in personal outcomes that comes with global integration: some individuals benefit,

while others suffer. Constructing the right tools for those who are harmed by the polarizing factor—options such as extended education and providing employment opportunities in the new industrial landscape—is likely to be more effective than simply “shutting down” globalization.

More generally, proactively providing educational, social, and economic opportunities to counteract negative feedback loops, encouraging more accurate beliefs among current and future generations through early exposure, and shaping the environment to favor collective intelligence are likely to be more successful policies than attempting to outlaw undesirable behaviors. As long as the environmental factors giving rise to these behaviors are still in force, the banned behaviors will re-emerge in one form or another.

Continuing with our example of Andorians and Tellarians, if bias and discrimination already exist against the Tellarians, an obvious policy may be to simply criminalize such discrimination. This can lead to more forced interactions between the Andorians and the Tellarians, which in turn causes everyone to have a higher factor exposure to the Tellarians. However, since bias already exists in the population (since the Tellarians will have a higher probability of adverse events either initially, or through negative feedback loops), this will lead to more Andorians experiencing adverse events from their interactions with Tellarians, inevitably leading to even stronger negative feedback (and even higher adverse probabilities) for the Tellarians—a cognitive tendency that is difficult to change (Hawkins and Blakeslee, 2004). As a result, direct attempts to outlaw bias and discrimination against the Tellarians may actually make matters worse. In this sense, our society needs not only more integration among different groups (Anderson, 2010) but, more importantly, measures to ensure that negative feedback does not reinforce itself after the integration.

These simple examples illustrate how seemingly well-intended policies can create more selective pressure for collective ignorance to emerge. The fundamental reason is that they are addressing the symptoms, not the root cause, of these failures of collective intelligence. We do not model the objective function that policy makers should use for managing societal issues such as polarization and discrimination, but implicit in our framework is the fitness of different types of individuals that determines their survival. Therefore, as representatives of a given group of constituents, policy makers can reasonably be expected to focus on what improves the long-term fitness (in the economic sense) of those constituents. Our evolutionary framework provides a lens through which the underlying causes—the environment in which these failures emerge—can be identified so as to construct more productive policies.

Using history as a mirror, these implications are even more relevant now as we experience the Artificial Intelligence (AI) revolution (Makridakis, 2017; Diamandis and Kotler, 2020). Just as in the Industrial Revolution 200 years

ago, and modern globalization over the past 50 years, the AI revolution will increase aggregate productivity while inevitably leading to another major shift in the industrial landscape and composition of the labor market. In this process, some individuals will benefit while others may be harmed. The policy suggestions outlined in this article, including extended education for those whose jobs have been replaced by AI, and providing children with equal access to education, particularly in STEM and AI-related subjects, are more pressing than ever.

### Acknowledgements

Research support from the MIT Laboratory for Financial Engineering is gratefully acknowledged. We thank Zach Church, Jessica Flack (editor), Steven A. Frank, Wendy Liu, David C. Schmittlein, Harriet A. Zuckerman, and an anonymous reviewer for helpful comments and discussion, and Jayna Cummings for editorial assistance. The views and opinions expressed in this article are those of the authors only, and do not necessarily represent the views and opinions of any institution or agency, any of their affiliates or employees, or any of the individuals acknowledged above.

### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### ORCID iD

Ruixun Zhang  <https://orcid.org/0000-0002-7670-8393>

### Supplemental Material

Supplemental material for this article is available online.

### Notes

- Here, we use “collective ignorance” to broadly refer to either lack of knowledge or simply bad judgment, which includes the phenomena of polarization and discrimination as special cases.
- Of course, certain “traits” are more likely to be biologically inherited than what is usually referred as “rational economic behaviors.” However, recent developments in neuroeconomics (Camerer et al., 2005; Glimcher and Fehr, 2013) suggest that these economic behaviors may also have a biological origin. For example, studies have found that the activity of a specific brain region correlates with risk-taking and risk-averse behavior (Tom et al., 2007).
- See Brennan and Lo (2011) for a more detailed discussion of this example, and Zhang et al. (2014a) for extensions of the binary choice model to explicitly allow for factor structures.
- The mathematical intuition behind why probability matching is the growth-optimal behavior lies in the fact that the population growth rate,  $3 \times p^{70\%} \times (1-p)^{30\%}$ , is maximized when  $p = 70\%$ . When the assumption of 3 offspring is replaced by, for example, 1.5, the population becomes extinct regardless of  $p$ . See Brennan and Lo (2011) for a more general discussion on how probability matching emerges in evolution and Lo et al. (2021) for experimental evidence on probability matching in financial decision-making.
- Zhang et al. (2014a) provide a general model with multiple factors. For expositional simplicity and without loss of generality, we consider a two-factor model here.
- We use the term “growth-optimal” instead of the shorter term “optimal” to distinguish our focus on the behavior that emerges through evolutionary dynamics from the optimizing behavior of rational agents that economists typically take for granted. Growth-optimal behavior is generally *not* optimal from the individual’s perspective.
- See Supplemental Material for proof.
- This result corresponds to the well-known principle of geometric mean fitness (Seger and Brockmann, 1987) in evolutionary biology. For evolutionary systems that do not necessarily maximize geometric mean fitness, see, for example, Frank (1990) and Lo et al. (2018) (relative success), McNamara (1995) (accounting for actions of kin), and an excellent review in Frank (2011a). Another possible direction to extend this framework is to consider fidelity in transmission (Lewis and Laland, 2012; Montrey and Shultz, 2020) and allow for mutation (King, 1972; Taddei et al., 1997; Drake et al., 1998; Brennan et al., 2018). We thank Steve A. Frank for bringing this important point to our attention.
- This proposition is proved in Brennan and Lo (2011) and Zhang et al. (2014a), which we reproduce here for completeness. Proofs of all propositions are provided in the online Supplemental Material.
- The risk-spreading behavior is also closely related to kin selection. Yoshimura and Clark (1991) show that a risk-spreading polymorphism can exist only for groups. Yoshimura and Jansen (1996) argue that a risk-spreading adaptation is a form of kin selection (Cooper and Kaplan 1982), in which the strategies of kin are important in stochastic environments even if no interactions exist. McNamara (1995) introduces the profile of a strategy and relates the geometric-mean fitness to a deterministic game.
- By globalization, we mean the growing interaction and integration among individuals, institutions, and economies worldwide.
- See Rodrik (2018, 2020) for more extensive discussions about globalization and the emergence of populism. As Rodrik states, “Globalization is probably not the only force at play in the rise of extreme political views. Changes in technology, rise



of winner-take-all markets, erosion of labor-market protections, and decline of norms restricting pay differentials all have played their part. These developments are not entirely independent from globalization, insofar as they both fostered globalization and were reinforced by it.” With that awareness, we use globalization as a single factor in our model to illustrate its role in shaping political views, using it to develop the intuition to apply our model to other factors.

13. Although it appears that groups of individuals emerge, evolutionary dynamics operate on the behaviors or strategies of these individuals. Individuals with the same behavior share the same fitness, and therefore rise and fall together via evolution.
14. See Brennan and Lo (2011) and Zhang et al. (2014a) for more discussions about individual versus group optimality, where the stochastic nature of the environment is key. This also relates to the extensive literature on the role of stochastic environments in evolutionary biology (Lynch and Lande, 1993; Burger and Lynch, 1995; Pekalski, 1998, 1999, 2002; De Blasio, 1999; Burger and Gimelfarb, 2002) and behavioral ecology (Real and Caraco, 1986; Stephens and Krebs, 1986; Deneubourg et al., 1987; Harder and Real, 1987; Pasteels et al., 1987; Mangel and Clark, 1988; Hölldobler and Wilson, 1990; Kirman, 1993; Thuijsman et al., 1995; Smallwood, 1996; Kesar et al., 2002; Ben-Jacob, 2008). In particular, Fretwell (1972), Cooper and Kaplan (1982), and Frank and Slatkin (1990) observe that randomizing behavior can be advantageous in the face of stochastic environmental conditions. See also Frank (2011a, 2011b, 2012a) for an excellent review.
15. We have deliberately chosen to use fictitious races borrowed from science fiction to lower the tension that accompanies a discussion of these highly emotionally charged issues, and also to illustrate the generality of our analysis. In particular, our framework can be applied to any marginalized group.
16. Examples of adverse events might include crimes, disease, or economic hardship, among others, as long as the adverse event represents a systematic factor for the particular group in consideration.
17.  $\beta$  is exogenous in our framework. However, one can consider extensions where the weight on the two factors,  $\lambda_T$  and  $\lambda_A$ , are in turn determined by the percentage of Tellarians in the population. This may generate time-varying patterns and cycles of discrimination levels in the population. We thank an anonymous reviewer for this point.
18. For this set of simulations, we fix the fraction of the Tellarian population at  $\beta = 0.5$ .
19. Readers may wonder why the non-discriminatory behavior ( $p^* = 0$ ) cannot persist in this case, which will become clear as we discuss the notion of locally evolutionarily stable strategies below.
20. See also Maynard Smith (1982).
21. There is also a large literature in evolutionary biology on the local properties and stability for ESSs, including convergence stability (Christiansen 1991), neighborhood invader strategy

(Apaloo 1997), frequency-dependent ESS (Pohley and Thomas 1983), evolutionarily singular strategies (Geritz et al., 1998), and sets of equilibrium strategies (Thomas 1985). See Apaloo et al. (2009) for a comparison of these concepts. We thank Steve A. Frank for bringing this to our attention.

22. The left boundary is not, strictly speaking, linear, but we refer to the region as triangular for simplicity.
23. For example, Jones et al. (2013) document that sharing varied perspectives, talents, and worldviews is beneficial to human interaction and institutional performance. They also demonstrate the resistance elicited in response to diversity, and the benefits that arise when it is overcome.
24. See Brennan et al. (2018) for a more detailed discussion of the role mutation plays in the context of the binary choice model.
25. We thank an anonymous reviewer for bringing this important point to our attention.

## References

- Alexander RD (1974) The evolution of social behavior. *Annual Review of Ecology and Systematics* 5: 325–383.
- Anderson E (2010) *The Imperative of Integration*. Princeton: Princeton University Press.
- Apaloo J (1997) Revisiting strategic models of evolution: the concept of neighborhood invader strategies. *Theoretical Population Biology* 52(1): 71–77.
- Apaloo J, Brown JS and Vincent TL (2009) Evolutionary game theory: ess, convergence stability, and nis. *Evolutionary Ecology Research* 11(4): 489–515.
- Arnold D, Dobbie WS and Hull P (2021) *Towards a Non-discriminatory Algorithm in Selected Data*. Technical report, National Bureau of Economic Research.
- Arrow K (1973) The theory of discrimination. In: Aschenfelter O and Rees A (eds) *Discrimination in Labor Markets*. Princeton, NJ: Princeton University Press, 3–33.
- Arrow KJ (1974) *The Limits of Organization*. New York: WW Norton & Company.
- Barkow JH, Cosmides L and Tooby J (1992) *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. New York, NY: Oxford University Press.
- Becker GS (1957) *The Economics of Discrimination*. Chicago, IL: University of Chicago Press.
- Ben-Jacob E (2008) Social behavior of bacteria: from physics to complex organizations. *European Physics Journal B* 65: 315–322.
- Bénabou R and Tirole J (2002) Self-confidence and personal motivation. *The Quarterly Journal of Economics* 117(3): 871–915.
- Blume L and Easley D (2006) If you’re so smart, why aren’t you rich? Belief selection in complete and incomplete markets. *Econometrica* 74(4): 929–966.
- Bohren JA, Haggag K, Imas A, et al. (2019) *Inaccurate Statistical Discrimination*. Technical report, National Bureau of Economic Research.



- Bordalo P, Coffman K, Gennaioli N, et al. (2016) Stereotypes. *The Quarterly Journal of Economics* 131(4): 1753–1794.
- Brennan TJ and Lo AW (2011) The origin of behavior. *Quarterly Journal of Finance* 1: 55–108.
- Brennan TJ and Lo AW (2012) An evolutionary model of bounded rationality and intelligence. *Plos One* 7(11): e50310.
- Brennan TJ, Lo AW and Zhang R (2018) Variety is the spice of life: irrational behavior as adaptation to stochastic environments. *Quarterly Journal of Finance* 8(3): 1850009.
- Brocas I and Carrillo JD (2000) The value of information when preferences are dynamically inconsistent. *European Economic Review* 44(4–6): 1104–1115.
- Burger R and Gimelfarb A (2002) Fluctuating environments and the role of mutation in maintaining quantitative genetic variation. *Genetical Research* 80: 31–46.
- Burger R and Lynch M (1995) Evolution and extinction in a changing environment: a quantitative-genetic analysis. *Evolution; International Journal of Organic Evolution* 49(1): 151–163.
- Buss DM (2004) *Evolutionary Psychology: The New Science of the Mind*. Boston, MA: Pearson.
- Camerer C, Loewenstein G and Prelec D (2005) Neuroeconomics: how neuroscience can inform economics. *Journal of Economic Literature* 43(1): 9–64.
- Carrillo JD and Mariotti T (2000) Strategic ignorance as a self-disciplining device. *The Review of Economic Studies* 67(3): 529–544.
- Charness G, Rustichini A and Van de Ven J (2018) Self-confidence and strategic behavior. *Experimental Economics* 21(1): 72–98.
- Christiansen FB (1991) On conditions for evolutionary stability for a continuously varying character. *The American Naturalist* 138(1): 37–50.
- Coate S and Loury GC (1993) Will affirmative-action policies eliminate negative stereotypes? *American Economic Review* 83(5): 1220–1240.
- Compte O and Postlewaite A (2004) Confidence-enhanced performance. *American Economic Review* 94(5): 1536–1557.
- Cooper WS and Kaplan RH (1982) Adaptive “coin-flipping”: a decision-theoretic examination of natural selection for random individual variation. *Journal of Theoretical Biology* 94(1): 135–151.
- Cosmides L and Tooby J (1994) Better than rational: evolutionary psychology and the invisible hand. *American Economic Review* 84: 327–332.
- Darwin C (1859) *On the Origin of Species*. London: Routledge.
- Dawkins R (1976) *The Selfish Gene*. Oxford, UK: Oxford University Press.
- De Blasio FV (1999) Diversity and extinction in a lattice model of a population with fluctuating environment. *Physical Review* 60: 5912–5917.
- Deneubourg JL, Aron S, Goss S, et al. (1987) Error, communication and learning in ant societies. *European Journal of Operational Research* 30(2): 168–172.
- Diamandis PH and Kotler S (2020) *The Future Is Faster than You Think: How Converging Technologies Are Transforming Business, Industries, and Our Lives*. New York: Simon & Schuster.
- Drake JW, Charlesworth B, Charlesworth D, et al. (1998) Rates of spontaneous mutation. *Genetics* 148(4): 1667–1686.
- Ehrlich PR and Levin SA (2005) The evolution of norms. *Plos Biology* 3: e194.
- Fama EF (1970) Efficient capital markets: a review of theory and empirical work. *The Journal of Finance* 25(2): 383–417.
- Flew T and Iosifidis P (2020) Populism, globalisation and social media. *International Communication Gazette* 82(1): 7–25.
- Frank SA (1990) When to copy or avoid an opponent’s strategy. *Journal of Theoretical Biology* 145(1): 41–46.
- Frank SA (2011a) Natural selection. i. variable environments and uncertain returns on investment. *Journal of Evolutionary Biology* 24(11): 2299–2309.
- Frank SA (2011b) Natural selection. ii. developmental variability and evolutionary rate. *Journal of Evolutionary Biology* 24: 2310–2320.
- Frank SA (2012a) Natural selection. iii. selection versus transmission and the levels of selection. *Journal of Evolutionary Biology* 25: 227–243.
- Frank SA and Slatkin M (1990) Evolution in a variable environment. *The American Naturalist* 136(2): 244–260.
- Fretwell SD (1972) *Populations in a Seasonal Environment*. Princeton, NJ: Princeton University Press.
- Fryer R and Jackson MO (2008) A categorical model of cognition and biased decision-making. *The B.E. Journal of Theoretical Economics* 8(1): 1–42.
- Fuligni AJ (2007) *Contesting Stereotypes and Creating Identities: Social Categories, Social Identities, and Educational Participation*. New York: Russell Sage Foundation.
- Geritz SA, Mesze G, Metz JA, et al. (1998) Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evolutionary Ecology* 12(1): 35–57.
- Gigerenzer G (2000) *Adaptive Thinking: Rationality in the Real World*. New York, NY: Oxford University Press.
- Glimcher PW and Fehr E (2013) *Neuroeconomics: Decision Making and the Brain*. Academic Press.
- Gorski P (2008) The myth of the “culture of poverty”. *Educational Leadership* 65(7): 32.
- Gregg D (2009) Developing a collective intelligence application for special education. *Decision Support Systems* 47(4): 455–465.
- Gregg DG (2010) Designing for collective intelligence. *Communications of the ACM* 53(4): 134–138.
- Haghtalab N, Jackson MO and Procaccia AD (2021) Belief polarization in a complex world: a learning theory perspective. *Proceedings of the National Academy of Sciences* 118(19): e2010144118.
- Hamilton WD (1963) The evolution of altruistic behavior. *The American Naturalist* 97(896): 354–356.

- Hamilton WD (1964) The genetical evolution of social behavior. i and ii. *Journal of Theoretical Biology* 7(1): 1–52.
- Harder LD and Real LA (1987) Why are bumble bees risk averse? *Ecology* 68(4): 1104–1108.
- Hawkins J and Blakeslee S (2004) *On Intelligence*. London: Macmillan.
- Heller Y and Winter E (2020) Biased-belief equilibrium. *American Economic Journal: Microeconomics* 12(2): 1–40.
- Hirshleifer J (1977) Economics from a biological viewpoint. *Journal of Law and Economics* 20: 1–52.
- Hölldobler B and Wilson EO (1990) *The Ants*. Cambridge, MA: Belknap Press.
- Ilon L (2012) How collective intelligence redefines education. In: *Advances in Collective Intelligence 2011*. Berlin: Springer, pp. 91–102.
- Jean E, Perroux M, Pepin J, et al. (2020) How to measure the collective intelligence of primary healthcare teams? *Learning Health Systems* 4(3): e10213.
- Johnson-Laird PN (1983) *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Cambridge: Harvard University Press, Vol. 6.
- Jones JM, Dovidio JF and Vietze DL (2013) *The Psychology of Diversity: Beyond Prejudice and Racism*. Hoboken: John Wiley & Sons.
- Keasar T, Rashkovich E, Cohen D, et al. (2002) Bees in two-armed bandit situations: foraging choices and possible decision mechanisms. *Behavioral Ecology* 13: 757–765.
- King JL (1972) The role of mutation in evolution. In: Le Cam LM, Neyman J and Scott EL (eds) *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume V*. Berkeley, CA: University of California Press, 69–100.
- Kirman A (1993) Ants, rationality, and recruitment. *Quarterly Journal of Economics* 108(1): 137–156.
- Kogan L, Ross SA, Wang J, et al. (2006) The price impact and survival of irrational traders. *The Journal of Finance* 61(1): 195–229.
- Kogan L, Ross SA, Wang J, et al. (2017) Market selection. *Journal of Economic Theory* 168: 209–236.
- Kubota JT, Li J, Bar-David E, et al. (2013) The price of racial bias: intergroup negotiations in the ultimatum game. *Psychological Science* 24(12): 2498–2504.
- Lai C and Banaji M (2021) The psychology of implicit intergroup bias and the prospect of change. In: *Difference without Domination*. Chicago, IL: University of Chicago Press, 115–146.
- Leimeister JM (2010) Collective intelligence. *Business & Information Systems Engineering* 2(4): 245–248.
- Lewis HM and Laland KN (2012) Transmission fidelity is the key to the build-up of cumulative culture. *Philosophical Transactions of the Royal Society B: Biological Sciences* 367(1599): 2171–2180.
- Lo AW (2004) The adaptive markets hypothesis. *Journal of Portfolio Management* 30(5): 15–29.
- Lo AW (2005) Reconciling efficient markets with behavioral finance: the adaptive markets hypothesis. *Journal of Investment Consulting* 7(2): 21–44.
- Lo AW (2012) Adaptive markets and the new world order. *Financial Analysts Journal* 68(2): 18–29.
- Lo AW (2013) Fear, greed, and financial crises: a cognitive neurosciences perspective. In: Fouque J and Langsam J (eds) *Handbook of Systemic Risk*. Cambridge, UK: Cambridge University Press, 622–662.
- Lo AW (2017) *Adaptive Markets: Financial Evolution at the Speed of Thought*. Princeton, NJ: Princeton University Press.
- Lo AW, Marlowe KP and Zhang R (2021) To maximize or randomize? an experimental study of probability matching in financial decision making. *Plos One* 16(8): e0252540.
- Lo AW, Orr HA and Zhang R (2018) The growth of relative wealth and the kelly criterion. *Journal of Bioeconomics* 20(1): 49–67.
- Lu Y, Kaushal N, Huang X, et al. (2021) Priming covid-19 salience increases prejudice and discriminatory intent against asians and hispanics. *Proceedings of the National Academy of Sciences* 118(36): e2105125118.
- Luo GY (1995) Evolution and market competition. *Journal of Economic Theory* 67(1): 223–250.
- Lynch M and Lande R (1993) Evolution and extinction in response to environmental change. In: Karieva PM, Kingsolver JG and Huey RB (eds) *Biotic Interactions and Global Change*. Sunderland, MA: Sinauer Associates, 235–250.
- Makridakis S (2017) The forthcoming artificial intelligence (ai) revolution: its impact on society and firms. *Futures* 90: 46–60.
- Malone TW (2018) *Superminds: The Surprising Power of People and Computers Thinking Together*. New York: Little, Brown Spark.
- Malone TW, Laubacher R and Dellarocas C (2010) The collective intelligence genome. *MIT Sloan Management Review* 51(3): 21.
- Mangel M and Clark CW (1988) *Dynamic Modeling in Behavioral Ecology*. Princeton, NJ: Princeton University Press.
- Maynard Smith J (1982) *Evolution and the Theory of Games*. Cambridge, UK: Cambridge University Press.
- Maynard Smith J (1984) Game theory and the evolution of behaviour. *Behavioral and Brain Sciences* 7: 95–125.
- Maynard Smith J and Price GR (1973) The logic of animal conflict. *Nature* 246(5427): 15–18.
- McNamara JM (1995) Implicit frequency dependence and kin selection in fluctuating environments. *Evolutionary Ecology* 9(2): 185–203.
- Merton RK (1960) The ambivalences of LeBon's The Crowd. In: *Introduction to the Compass Books Edition of Gustave LeBon, the Crowd*. New York: Viking.
- Montrey M and Shultz TR (2020) The evolution of high-fidelity social learning. *Proceedings of the Royal Society B* 287(1928): 1–8.
- Nowak MA (2006) Five rules for the evolution of cooperation. *Science* 314(5805): 1560–1563.

- Oster E (2020) Health recommendations and selection in health behaviors. *American Economic Review: Insights* 2(2): 143–160.
- Pasteels JM, Deneubourg JL and Goss S (1987) Self-organization mechanisms in ant societies. i: trail recruitment to newly discovered food sources. *Experientia. Supplementum*.
- Pastor L and Veronesi P (2020) *Inequality Aversion, Populism, and the Backlash against Globalization*. Technical report, National Bureau of Economic Research.
- Pekalski A (1998) A model of population dynamics. *Physica A: Statistical Mechanics and Its Applications* 252: 325–335.
- Pekalski A (1999) Mutations and changes of the environment in a model of biological evolution. *Physica A: Statistical Mechanics and Its Applications* 265: 255–263.
- Pekalski A (2002) Evolution of population in changing conditions. *Physica A: Statistical Mechanics and Its Applications* 314: 114–119.
- Phelps ES (1972) The statistical theory of racism and sexism. *American Economic Review* 62(4): 659–661.
- Pinker S (1979) Formal models of language learning. *Cognition* 7: 217–283.
- Pinker S (1991) Rules of language. *Science* 253: 530–535.
- Pinker S (1994) *The Language Instinct: How the Mind Creates Language*. New York, NY: William Morrow and Company.
- Pohley HJ and Thomas B (1983) Non-linear ess-models and frequency dependent selection. *Bio Systems* 16(2): 87–100.
- Real L and Caraco T (1986) Risk and foraging in stochastic environments. *Annual Review of Ecology and Systematics* 17: 371–390.
- Reinhart CM and Rogoff KS (2009) *This Time Is Different: Eight Centuries of Financial Folly*. Princeton, NJ: Princeton University Press.
- Riedl C, Kim YJ, Gupta P, et al. (2021) Quantifying collective intelligence in human groups. *Proceedings of the National Academy of Sciences* 118(21): e2005737118.
- Roberts SO and Rizzo MT (2021) The Psychology of American Racism. *American Psychologist* 76(3), 475–487. <https://doi.org/10.1037/amp0000642>
- Robson AJ (1996) A biological basis for expected and non-expected utility. *Journal of Economic Theory* 68(2): 397–424.
- Rodrik D (2018) Populism and the economics of globalization. *Journal of International Business Policy* 1(1): 12–33.
- Rodrik D (2020) Why does globalization fuel populism? economics, culture, and the rise of right-wing populism. *Annual Review of Economics* 13: 133–170.
- Rogers AR (1994) Evolution of time preference by natural selection. *American Economic Review* 84(3): 460–481.
- Samuelson PA (1965) Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review* 6(2): 41–49.
- Schneider DJ (2005) *The Psychology of Stereotyping*. New York, NY: Guilford Press.
- Segaran T (2007) *Programming Collective Intelligence: Building Smart Web 2.0 Applications*. Newton: O’Reilly Media, Inc.
- Seger J and Brockmann HJ (1987) What is bet-hedging? *Oxford Surveys in Evolutionary Biology* 4: 182–211.
- Smallwood P (1996) An introduction to risk sensitivity: the use of Jensen’s Inequality to clarify evolutionary arguments of adaptation and constraint. *American Zoologist* 36: 392–401.
- Stephens DW and Krebs JR (1986) *Foraging Theory*. Princeton, NJ: Princeton University Press.
- Surowiecki J (2005) *The Wisdom of Crowds*. New York, NY: Anchor.
- Swank D and Betz HG (2003) Globalization, the welfare state and right-wing populism in western europe. *Socio-Economic Review* 1(2): 215–245.
- Taddei F, Radman M, Maynard-Smith J, et al. (1997) Role of mutator alleles in adaptive evolution. *Nature* 387: 700–702.
- Thomas B (1985) On evolutionarily stable sets. *Journal of Mathematical Biology* 22(1): 105–115.
- Thuijssman F, Peleg B, Amitai M, et al. (1995) Automata, matching and foraging behavior of bees. *Journal of Theoretical Biology* 175: 305–316.
- Tom SM, Fox CR, Trepel C, et al. (2007) The neural basis of loss aversion in decision-making under risk. *Science* 315(5811): 515–518.
- Tooby J and Cosmides L (1995) Conceptual foundations of evolutionary psychology. In: Barkow JH, Cosmides L and Tooby J (eds) *The Handbook of Evolutionary Psychology*. Hoboken, NJ: John Wiley & Sons, 5–67.
- Trivers RL (1971) The evolution of reciprocal altruism. *The Quarterly Review of Biology* 46(1): 35–57.
- Trivers RL (1985) *Social Evolution*. Menlo Park, CA: Benjamin/Cummings.
- Trivers RL (2002) *Natural Selection and Social Theory: Selected Papers of Robert L. Trivers*. Oxford, UK: Oxford University Press.
- Tversky A and Kahneman D (1974) Judgment under uncertainty: heuristics and biases. *Science* 185(4157): 1124–1131.
- Tversky A and Kahneman D (1983) Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment. *Psychological Review* 90(4): 293.
- Vomfell L and Stewart N (2021) Officer bias, over-patrolling and ethnic disparities in stop and search. *Nature Human Behaviour* 5: 566–575.
- Waldman M (1994) Systematic errors and the theory of natural selection. *American Economic Review* 84(3): 482–497.
- Waller I and Anderson A (2021) Quantifying social organization and political polarization in online platforms. *Nature* 600(7888): 264–268.
- Westley PA, Berdahl AM, Torney CJ, et al. (2018) Collective movement in ecology: from emerging technologies to conservation and management. *Philosophical Transactions of the Royal Society B: Biological Sciences* 373(1746): 20170004.

- Wilson EO (1975) *Sociobiology: The New Synthesis*. Cambridge, MA: Harvard University Press.
- Woolley AW, Aggarwal I and Malone TW (2015) Collective intelligence and group performance. *Current Directions in Psychological Science* 24(6): 420–424.
- Woolley AW, Chabris CF, Pentland A, et al. (2010) Evidence for a collective intelligence factor in the performance of human groups. *Science* 330(6004): 686–688.
- Wynne-Edwards VC (1963) Intergroup selection in the evolution of social systems. *Nature* 200: 623–626.
- Yoshimura J and Clark CW (1991) Individual adaptations in stochastic environments. *Evolutionary Ecology* 5(2): 173–192.
- Yoshimura J and Jansen VA (1996) Evolution and population dynamics in stochastic environments. *Researches on Population Ecology* 38(2): 165–182.
- Zhang R, Brennan TJ and Lo AW (2014a) Group selection as behavioral adaptation to systematic risk. *Plos One* 9(10): e110848.
- Zhang R, Brennan TJ and Lo AW (2014b) The origin of risk aversion. *Proceedings of the National Academy of Sciences* 111(50): 17777–17782.