

RESEARCH ARTICLE

Threshold dynamics of corrective action in a population model of structural discrimination

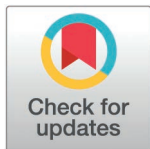
Masanori Takano ^{1,2*}

1 CyberAgent, Inc., Tokyo, Japan, **2** Keio University, Tokyo, Japan

* takano_masanori@cyberagent.co.jp

Abstract

Structural discrimination can sustain underrepresentation even when no actor explicitly excludes a group. We develop a simple population model in which current representation changes each type's per-capita effective rate of maintaining or expanding its presence, while corrective action responds to the gap from a reference-population share. The model yields three main results. First, an exclusionary state can remain stable when corrective action is present but weak relative to representation-dependent structural feedback. Second, the corrective strength required to destabilize exclusion is greater for a focal type that is rarer in the reference population. Third, avoiding complete exclusion is not the same as attaining proportional representation: with finite corrective strength, the stable long-run share can remain below the reference share. These results connect qualitative accounts of cumulative structural disadvantage to explicit conditions for long-run stability. They also suggest that corrective policies should be evaluated by their effective dynamic strength relative to the structural feedback they are intended to change, rather than by their formal presence alone.



OPEN ACCESS

Citation: Takano M (2026) Threshold dynamics of corrective action in a population model of structural discrimination. PLOS Complex Syst 3(9): e0000132. <https://doi.org/10.1371/journal.pcsy.0000132>

Editor: Aming Li, Peking University, CHINA

Received: June 30, 2026

Accepted: August 19, 2026

Published: September 15, 2026

Copyright: © 2026 Masanori Takano. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: All data are in the manuscript and/or [Supporting Information](#) files. No empirical datasets were generated or analyzed in this study. The code used to generate all model-derived numerical values and figures is provided as [S1 Code](#).

Funding: The author(s) received no specific funding for this work.

Author summary

We study how underrepresentation can persist even when explicit exclusion is absent. In many social systems, current representation affects what happens next: people may have better access to information, support, role models, and networks when people like them are already present. We use a simple population model in which these mechanisms change the per-person rate at which each group maintains or expands its representation. The model shows that a corrective measure can exist yet remain too weak to prevent exclusion. A lasting nonzero share appears only when corrective action exceeds a threshold relative to the structural feedback. That threshold is higher when the focal group is rarer in a broader reference population. Even above the threshold, the long-run share in the target population can remain below the reference share. Thus, preventing

Competing interests: The author has declared that no competing interests exist.

exclusion and approaching proportional representation are distinct policy goals, and both the strength of corrective action and the strength of the underlying structural feedback matter.

1 Introduction

Disparities in social attributes, such as gender, race, ethnicity, age, and social background, are observed across employment, education, political representation, and participation in scientific and technological fields. Classical theories of discrimination have considered discrimination based on individual preferences and discrimination based on incomplete information about groups. Becker formulated taste-based discrimination in economic terms [1], whereas Phelps [2], Arrow [3], and Aigner and Cain [4] developed statistical theories of discrimination. These studies provided a basis for analyzing discrimination using formal and economic models. More recent work has also treated discrimination as a dynamic process. Bohren et al. [5], for example, developed a dynamic model showing that the evolution of discrimination can help identify its underlying source. However, many real disparities cannot be explained only by explicit discriminatory intentions, imperfect information, or individual-level prejudice.

The concept of structural discrimination is useful for this problem. Pincus [6] distinguished individual, institutional, and structural discrimination. Structural discrimination refers to cases in which policies, customs, procedures, or evaluation criteria that appear neutral in intent nevertheless have negative effects on particular groups. For example, selection criteria may be formally identical for all individuals, but the opportunities to acquire the required experience, information, networks, or role models may differ by attribute. Such discrimination is difficult to identify because it can occur without a clear perpetrator or explicit hostility [6–8]. Small and Pager [9] argue that institutional discrimination should be analyzed independently of taste-based and statistical discrimination, because past discrimination can have contemporary consequences even when the current actor's discriminatory decision is immaterial. Ray [10] further emphasizes that organizations should not be regarded as race-neutral containers, but can operate as structures that stabilize and institutionalize inequality.

A key feature of structural discrimination is its self-reinforcing nature. When people with a given attribute are rare in an organization, occupation, educational program, or political body, they may have fewer role models, less informal support, and weaker access to relevant networks. This logic is close to the representation dynamic of Carvalho and Pradelski [11]: when a group is more represented in an activity, the activity is more likely to be viewed as normal or appropriate for that group, which affects participation and feeds back into representation. A similar representation-dependent mechanism appears in models of mentoring. Müller-Ippen and Öry [12] analyze a workforce model in which mentoring quality depends on group representation and show that persistent interventions may be needed to avoid majority overrepresentation. Empirical evidence on mentoring also supports the importance of role models

and support; for example, female peer mentors in engineering education can have persistent positive effects on female students [13]. Thus, a low current share can produce a higher future barrier, and past disadvantage can be reproduced through present opportunities.

This idea is closely related to cumulative advantage and cumulative disadvantage. Merton's self-fulfilling prophecy [14], the Matthew effect [15], and later work on cumulative inequality [16] emphasize that initial advantages or disadvantages can be amplified over time. Kanter's study of tokenism also shows that group proportions affect experience and evaluation within organizations [17]. In mathematical social science, Schelling showed that weak individual-level tendencies can produce strong aggregate segregation [18], and Granovetter proposed a threshold model in which individual actions depend on the proportion of others already acting [19]. These studies indicate that aggregate inequality can emerge from feedback between current proportions and subsequent behavior.

Corrective interventions, including affirmative action, mentoring, targeted information provision, and institutional revisions, are often introduced to reduce such disadvantages. However, their effects are not necessarily simple. Lundberg and Startz [20] considered discrimination and social intervention in competitive labor markets. Coate and Loury [21] showed that affirmative action can eliminate negative stereotypes under some conditions but can preserve them under others. Further studies have discussed how the effects of affirmative action depend on institutional design, incentives, initial conditions, and market structure [22–24]. The mentoring model of Müller-Iten and Öry [12] also shows that representation-dependent mechanisms can make persistent interventions welfare-relevant. Therefore, it is important to examine not only whether an intervention exists, but also how strong it must be relative to the structural forces that maintain underrepresentation.

This problem is also related to fairness in decision-making. Recent studies on algorithmic fairness distinguish individual fairness from group-based criteria such as equality of opportunity [25–27]. When structural discrimination exists, applying the same rule to all individuals does not necessarily eliminate group-level disadvantage, because the opportunities required to satisfy that rule may have been unequally distributed over time. Thus, structural discrimination should be treated not only as a static evaluation problem, but also as a dynamic process in which group shares evolve.

Mathematical models are useful for clarifying such processes. Real societies contain many interacting factors, and a model cannot represent all of them. However, a simple model can isolate a basic feedback mechanism and reveal the conditions under which it changes long-term outcomes. In particular, when current representation affects a type's subsequent rate of entry, continuation, promotion, or exit, a dynamic model is needed. Replicator dynamics provide a standard framework in evolutionary game theory and population games: they describe a population process in which a type's frequency increases when its per-capita payoff exceeds the population-average payoff [28–31]. Population dynamics have also been used to analyze discrimination and prejudice through social contagion and external shocks [32]. The present study focuses on a different mechanism: representation-dependent structural feedback and corrective action intended to counter it.

Formal models of social categories and inequality have also been developed. O'Connor [33] analyzed how social categories, such as gender and race, can lead to unfairness in coordination and resource-allocation games. Carvalho and Pradelski [11] provide a particularly close point of comparison by studying a representation dynamic in which group representation affects perceived suitability and future participation. Our model builds on the same general idea that representation feeds back into future participation, but it asks a different question: how strong must an explicit corrective force be to change the stability of exclusion and coexistence? This dynamic-stability framework complements classical taste-based and statistical-discrimination models. Those baseline theories center, respectively, on discriminatory preferences and inference under imperfect information, whereas the present model directly endogenizes feedback from current representation to subsequent type-specific rates. It can therefore describe initial-condition dependence, stable boundary states, and policy-induced stability changes without treating the classical theories as incapable of representing every structural mechanism.

We propose a minimal two-type population model. The parameter p is the share of focal type A in a reference population, and x is its current share in a target population. We use payoff in the replicator-dynamic sense of a type-specific per-capita effective growth or retention rate, rather than a monetary utility or an aggregate number of opportunities. The parameter α measures the strength with which reference prevalence and current representation generate structural feedback in these rates, while β measures the effective strength with which corrective action translates the representation gap into an opposing rate advantage. Both parameters are reduced-form quantities that can summarize multiple institutional mechanisms rather than one-to-one measures of a particular policy.

The analysis makes three contributions. First, it derives a threshold above which a small positive representation can grow rather than decline toward exclusion. Second, it shows that the required corrective-strength threshold becomes higher as the focal type becomes rarer in the reference population. Third, it distinguishes destabilizing complete exclusion from approaching the reference share. For the baseline model and $0 < p \leq 1/2$, a locally asymptotically stable interior equilibrium (a long-run state with $0 < x < 1$, in which both types remain represented) exists only if

$$\frac{\beta}{\alpha} > \frac{1-p}{2p}. \quad (1)$$

For $p < 1/2$, the stable interior equilibrium remains below p for every finite β/α , although it approaches p as corrective strength becomes large. The threshold therefore marks the avoidance of stable exclusion, not the attainment of proportional representation.

To assess whether the central mechanism depends on the two-type assumption, we also analyze a direct separable symmetric N -type extension and illustrate its equilibrium transitions with three- and four-type examples.

2. Materials and methods

2.1. Model setting and notation

We consider a target population consisting of two attribute types, denoted by A and B , with A as the focal type. The target population may be a profession, organization, educational program, or political body in which representation evolves over time. Let $p \in (0, 1)$ denote the share of type A in a reference population, and let $x(t) \in [0, 1]$ denote its current share in the target population. The corresponding shares of type B are $1 - p$ and $1 - x(t)$. We write x for $x(t)$ when no confusion arises.

In the baseline model, the same reference share p has two roles: it is an exogenous prevalence factor in the structural component and the proportional benchmark in the corrective component. In recruitment settings where potential entrants are drawn from the reference population, p may additionally be interpreted as the baseline probability that a potential entrant is of type A . This recruitment interpretation is optional rather than the general definition of p .

2.2. Per-capita payoff specification

We use payoff in the replicator-dynamic sense. Specifically, $f_A(x)$ and $f_B(x)$ are type-specific per-capita effective growth or retention rates. They are not aggregate numbers of opportunities and need not be monetary utilities. Rather, they summarize how readily each type maintains or expands its representation through the combined effects of entry, continuation, promotion, and exit. Relative dynamics are determined by the difference $f_A - f_B$.

2.3. Structural and corrective components

For type A , we specify the structural component of the per-capita rate as $\alpha p x$, where $\alpha > 0$. This reduced-form term increases jointly with the reference-population share p and current target-population representation x . The variable x summarizes representation-dependent mechanisms such as access to role models, information, mentoring, informal support, networks, and institutional fit. The model does not identify these pathways separately or claim that each is exactly linear in representation.

The bilinear term is an analytically tractable baseline rather than a uniquely derived empirical law. If a smooth representation-dependent component $S_A(p, x)$ vanishes when the focal type is absent from the reference population or when its target-population representation is zero, so that $S_A(0, x) = S_A(p, 0) = 0$, then px is the lowest-order nonzero interaction between p and x . By the same baseline logic, type B receives the structural component $\alpha(1 - p)(1 - x)$.

This symmetric specification is adopted as a simple analytical baseline that isolates the core mechanism. It should not be interpreted as an empirical claim that dominant and marginalized groups face identical institutional or network mechanisms. We discuss the implication of asymmetric structural strengths below and provide the formal derivation in [S1 Appendix](#).

Corrective action can include targeted recruitment or support, information provision, mentoring, and institutional revision. Its terms are also changes in per-capita rates rather than aggregate policy resources. When $x < p$, $\beta(p - x)$ raises the rate of type A , while $\beta(x - p)$ changes the rate of type B in the opposite direction, where $\beta > 0$. Thus, corrective action contributes $2\beta(p - x)$ to $f_A - f_B$. The parameter β is the effective intensity with which a representation gap is converted into a relative dynamic advantage.

$$f_A(x) = \alpha px + \beta(p - x), \quad (2)$$

$$f_B(x) = \alpha(1 - p)(1 - x) + \beta(x - p). \quad (3)$$

2.4. Replicator dynamics

The population-average payoff is $\bar{f}(x) = xf_A(x) + (1 - x)f_B(x)$. The two-type replicator equation can then be written as

$$\dot{x} = x\{f_A(x) - \bar{f}(x)\} = x(1 - x)\{f_A(x) - f_B(x)\}.$$

The prefactor x is the current frequency of type A and weights the per-capita payoff difference by that frequency. By contrast, the x inside f_A changes the per-capita rate itself through representation-dependent feedback. These two appearances of x therefore have distinct roles. In particular, xf_A is a frequency-weighted contribution to the population-average payoff, not a measure of total growth or total opportunity for type A .

From [Eqs. \(2\) and \(3\)](#), the payoff difference is

$$f_A(x) - f_B(x) = \alpha(x + p - 1) + 2\beta(p - x).$$

The baseline dynamical equation is therefore

$$\dot{x} = x(1 - x) [\alpha(x + p - 1) + 2\beta(p - x)]. \quad (4)$$

The states $x=0$ and $x=1$ are invariant boundaries of this standard replicator system. The analysis below identifies boundary and interior equilibria, determines their local stability from the sign of the vector field, and follows the equilibrium branches as the relative corrective strength changes.

2.5. Analytical strategy and robustness extensions

The closed-form analysis uses the baseline bilinear and linear terms. To examine which threshold conclusions depend on these forms, we also consider the generalized per-capita rates

$$\begin{aligned} f_A(x) &= \alpha p s(x) + \beta c(p - x), \\ f_B(x) &= \alpha(1 - p)s(1 - x) + \beta c(x - p). \end{aligned}$$

Here $s : [0, 1] \rightarrow [0, 1]$ is continuous and increasing, with $s(0) = 0$ and $s(1) = 1$, and is differentiable in the interior when derivatives are used. The function $c : [-1, 1] \rightarrow \mathbb{R}$ is continuous, increasing, and odd, with $c(0)=0$ and $c'(z) > 0$ over the examined range. We derive local boundary conditions analytically and examine $s(x) = x^\gamma$, $\gamma \in \{0.8, 1.0, 1.2\}$, and $c(z) = z + \kappa z^3$, $\kappa \in \{-0.2, 0, 0.2\}$, numerically. Detailed derivations, numerical settings, and sensitivity plots are provided in Sections 2–6 (Local boundary thresholds through Illustrative five-equilibrium phase line) of [S1 Appendix](#).

2.6. Separable multi-type extension

To examine whether the qualitative mechanism depends on the two-type assumption, we consider a direct separable symmetric extension to N mutually exclusive types. Let $p_i > 0$ and $x_i \geq 0$ denote the reference-population and current target-population shares of type i , respectively, with

$$\sum_{i=1}^N p_i = 1, \quad \sum_{i=1}^N x_i = 1.$$

The type-specific per-capita rate is

$$f_i(\mathbf{x}) = \alpha p_i x_i + \beta(p_i - x_i) = \alpha\{v p_i + (p_i - v)x_i\}, \quad (5)$$

and the corresponding replicator dynamics are

$$\dot{x}_i = x_i(f_i - \bar{f}), \quad \bar{f} = \sum_{j=1}^N x_j f_j, \quad (6)$$

where $v = \beta/\alpha$ is the corrective-to-structural strength ratio. Setting $N=2$, $p_1=p$, $p_2=1-p$, $x_1=x$, and $x_2=1-x$ recovers [Eqs. \(2\)](#) and [\(3\)](#) exactly. This categorical extension retains common structural and corrective strengths and excludes cross-type spillovers; detailed equilibrium and stability analyses are provided in [S1 Appendix](#).

3. Results

We analyze [Eq. \(4\)](#) using the relative corrective strength v . Unless otherwise stated, the primary case is $0 < p \leq 1/2$, in which focal type A is no more prevalent than type B in the reference population. We later give the corresponding condition for $p > 1/2$.

3.1. Equilibria

The boundary equilibria $x=0$ and $x=1$ always exist. An interior equilibrium is an equilibrium with $0 < x < 1$, at which both types remain represented. For $v \neq 1/2$, the candidate interior equilibrium is

$$x^* = \frac{\alpha(1-p) - 2\beta p}{\alpha - 2\beta} = \frac{1-p-2pv}{1-2v}. \quad (7)$$

This candidate is an actual interior equilibrium only when $0 < x^* < 1$. Defining $g(x) = \alpha(x+p-1) + 2\beta(p-x)$ gives $\dot{x} = x(1-x)g(x)$ and $g'(x) = \alpha(1-2v)$.

3.2. Local stability

Near $x=0$, $\dot{x} \simeq xg(0)$. Hence, $x=0$ is locally asymptotically stable when $g(0) = \alpha(p-1) + 2\beta p < 0$, or equivalently

$$v < \frac{1-p}{2p}. \quad (8)$$

Thus, the exclusionary state of type *A* is stable when corrective action is weaker than this relative threshold. Similarly, $x = 1$ is locally asymptotically stable when $g(1) = \alpha p - 2\beta(1-p) > 0$, which gives

$$v < \frac{p}{2(1-p)}. \quad (9)$$

At an interior equilibrium,

$$\left. \frac{dx}{dt} \right|_{x=x^*} = x^*(1-x^*)\alpha(1-2v),$$

so an existing interior equilibrium is locally asymptotically stable for $v > 1/2$ and unstable for $v < 1/2$. For $0 < p \leq 1/2$, combining stability with the requirement $0 < x^* < 1$ yields the main threshold

$$v = \frac{\beta}{\alpha} > \frac{1-p}{2p}. \quad (10)$$

The right-hand side diverges as $p \rightarrow 0$, making explicit why a rarer focal type requires stronger corrective action relative to the structural feedback.

3.3. Parameter regimes and bifurcation structure

For the strict minority case $0 < p < 1/2$, define

$$v_1 = \frac{p}{2(1-p)}, \quad v_0 = \frac{1-p}{2p}, \quad v_1 < \frac{1}{2} < v_0.$$

At $v = v_1$, the equilibrium branch $x^*(v)$ intersects the boundary branch $x = 1$, and the two branches exchange stability; hence, v_1 is a transcritical bifurcation. At $v = v_0$, $x^*(v)$ intersects the boundary branch $x = 0$, and the two branches exchange stability; hence, v_0 is also a transcritical bifurcation. [Table 1](#) summarizes the open parameter intervals; the equality cases are nonhyperbolic and are excluded from the table.

For $0 < v < v_1$, the two boundaries are stable and the interior equilibrium is unstable, so the outcome depends on the initial representation. For $v_1 < v < v_0$, only $x = 0$ is stable and every interior trajectory approaches exclusion. For $v > v_0$, both boundaries are unstable and every interior trajectory approaches the stable interior equilibrium. The policy-relevant stability exchange occurs at v_0 : immediately above it, a small positive share of type *A* grows rather than declines toward exclusion.

Table 1. Parameter regimes for $0 < p < 1/2$.

Condition	Stable long-run state	Interpretation
$0 < v < v_1$	$x = 0$ or $x = 1$	Initial-state dependence
$v_1 < v < v_0$	$x = 0$	Stable exclusion
$v > v_0$	$x^* \in (0, p)$	Sub-proportional coexistence

<https://doi.org/10.1371/journal.pcsy.0000132.t001>

For $0 < p < 1/2$, $v = 1/2$ lies between v_1 and v_0 . At this value, the payoff difference is independent of x , and [Eq. \(4\)](#) reduces to

$$\dot{x} = \alpha(2p - 1)x(1 - x). \quad (11)$$

Hence, every interior trajectory approaches $x = 0$, and no equilibrium branches intersect and exchange stability. Thus, $v = 1/2$ is not a transcritical bifurcation. At the doubly degenerate point $p = 1/2$, $v = 1/2$, every $x \in [0, 1]$ is an equilibrium.

For general $p \in (0, 1)$, a stable interior equilibrium requires

$$v > \max \left\{ \frac{1-p}{2p}, \frac{p}{2(1-p)} \right\}. \quad (12)$$

Accordingly, when $p > 1/2$, stable interior coexistence requires $v > p/[2(1-p)]$ and the stable branch satisfies $x^* > p$, the mirror image of the $p < 1/2$ case. The smaller value $(1-p)/(2p)$ is the point at which $x = 0$ loses stability; between these two thresholds, interior trajectories instead approach $x = 1$. A focal type can therefore be a majority in a reference population yet initially underrepresented in a particular occupation. For example, men constituted 52.9% of all employed people but only 12.7% of employed registered nurses in the United States in 2025 [\[34\]](#). This is an illustrative context for the labeling convention, not an empirical test of the present model.

3.4. Location of the stable interior equilibrium

The existence of a stable interior equilibrium does not imply proportional representation. For $0 < p < 1/2$ and $v > v_0$, [Eq. \(7\)](#) gives

$$p - x^* = \frac{1 - 2p}{2v - 1} > 0. \quad (13)$$

Thus, exceeding v_0 prevents stable exclusion but leaves a positive long-run gap for every finite v . The gap vanishes only in the limit $v \rightarrow \infty$. More generally, if $0 < \epsilon < p$ is an acceptable gap, then $p - x^* \leq \epsilon$ holds exactly when

$$v \geq \frac{1}{2} \left(1 + \frac{1 - 2p}{\epsilon} \right). \quad (14)$$

This lower bound exceeds v_0 . The intervention needed to make exclusion unstable is therefore weaker than the intervention needed to move the stable equilibrium within a specified tolerance of the reference share.

3.5. Comparative numerical cases and visualizations

[Table 2](#) compares the requested reference shares. At $p = 0.1$, the exclusion threshold is $v_0 = 4.5$; at $p = 0.4$, it is 0.75; and at $p = 0.5$, both boundary thresholds equal 0.5.

For $p = 0.4$, the stable interior branch is $x^* = (0.6 - 0.8v)/(1 - 2v)$; for example, $x^* = 0.2$ at $v = 1$ and $x^* = 1/3$ at $v = 2$. For $p = 0.5$, $x^* = 0.5$ is stable when $v > 0.5$, unstable when $v < 0.5$, and every state is an equilibrium at $v = 0.5$.

[Fig 1](#) shows phase lines for $p = 0.4$ across the bistable, stable-exclusion, and stable-interior regimes. [Fig 2](#) follows all three equilibrium branches and shows that $x^*(v)$ intersects and exchanges stability with the respective boundary branches at $v_1 = 1/3$ and $v_0 = 3/4$; these are the two transcritical bifurcations. The value $v = 1/2$ is shown separately as the point at which the payoff difference loses its dependence on current representation. [Fig 3](#) expands the vertical range to include the $p = 0.1$, $v_0 = 4.5$ case and marks all three comparative examples.

Table 2. Comparative baseline thresholds.

p	v_1	v_0	Implication
0.1	$1/18 \approx 0.0556$	$9/2 = 4.5$	High exclusion threshold
0.4	$1/3$	$3/4$	Illustrative main case
0.5	$1/2$	$1/2$	Balanced reference share

<https://doi.org/10.1371/journal.pcsy.0000132.t002>

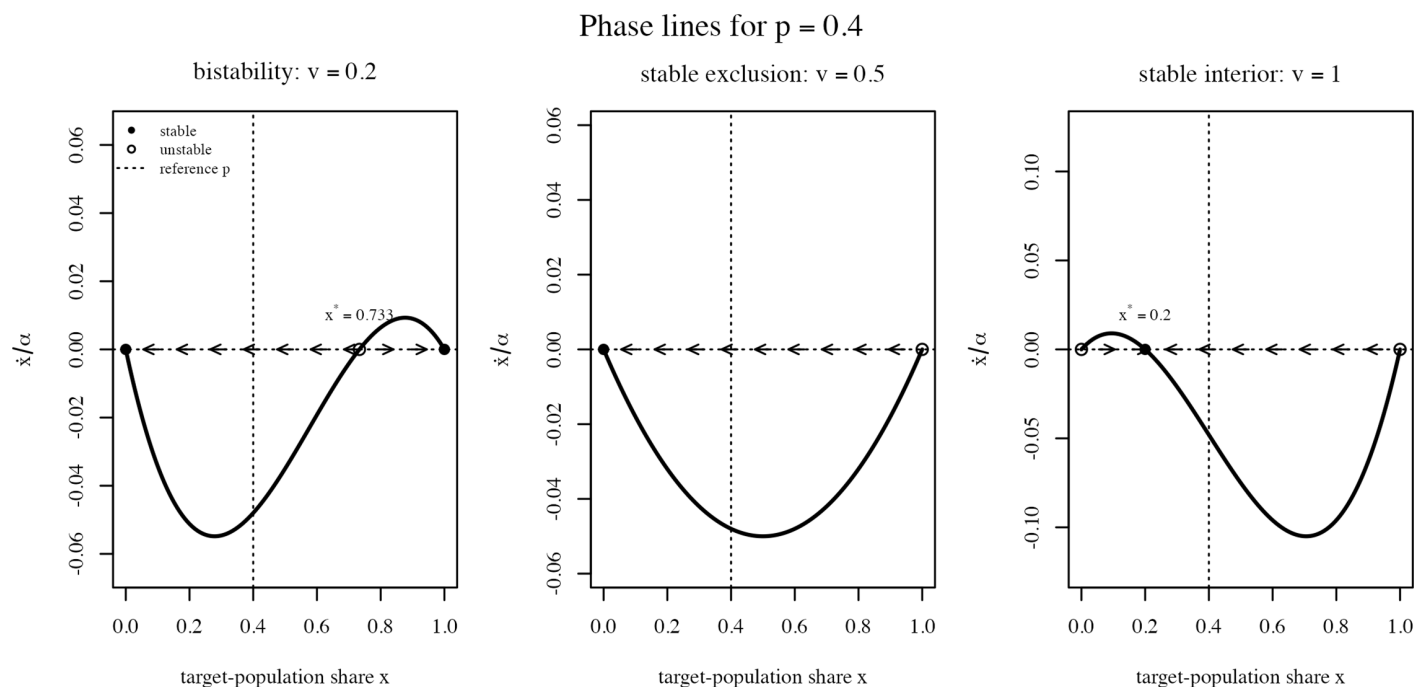


Fig 1. Phase-line diagrams for the one-dimensional dynamics. The panels use $p=0.4$ and $v=0.2, 0.5$, and 1.0 . Curves show $\dot{x}/\alpha$ as a function of x . Filled points are locally stable equilibria, and open points are unstable equilibria. Corrective action is present in all panels, but $x=0$ remains stable until v exceeds $v_0=0.75$.

<https://doi.org/10.1371/journal.pcsy.0000132.g001>

3.6. Robustness to nonlinear feedback

For the generalized per-capita rates, the payoff difference at the exclusion boundary is

$$f_A(0) - f_B(0) = -\alpha(1-p) + 2\beta c(p).$$

Therefore, a small positive representation of type A grows, and $x=0$ loses local stability, when

$$\frac{\beta}{\alpha} > \frac{1-p}{2c(p)}. \quad (15)$$

Similarly, $x=1$ loses local stability when

$$\frac{\beta}{\alpha} > \frac{p}{2c(1-p)}. \quad (16)$$

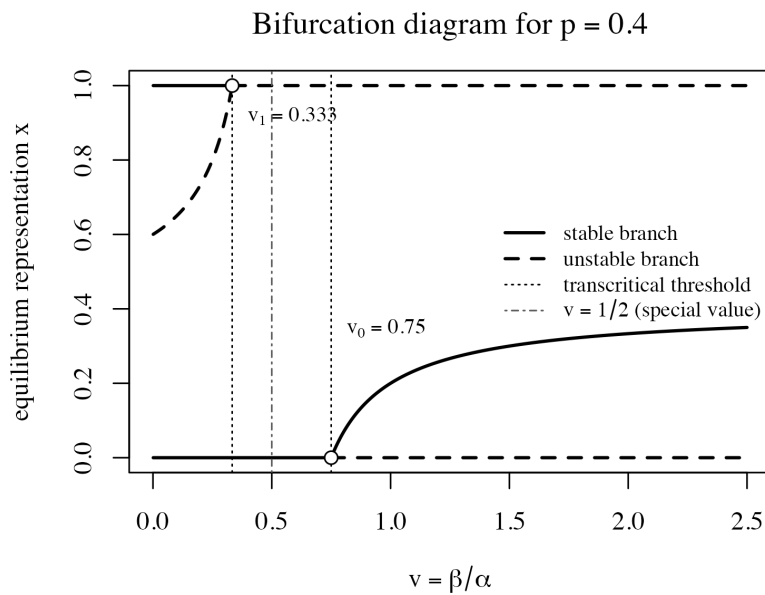


Fig 2. Bifurcation diagram for $p=0.4$. Solid branches are locally stable and dashed branches are unstable. The equilibrium branch $x^*(v)$ intersects and exchanges stability with $x=1$ at $v_1=1/3$ and with $x=0$ at $v_0=3/4$; these are the two transcritical bifurcations. The vertical line at $v=1/2$ marks a special value at which the payoff difference is independent of x ; it is not a transcritical bifurcation. Above v_0 , a small positive representation of type A grows toward the stable interior branch rather than declining to exclusion.

<https://doi.org/10.1371/journal.pcsy.0000132.g002>

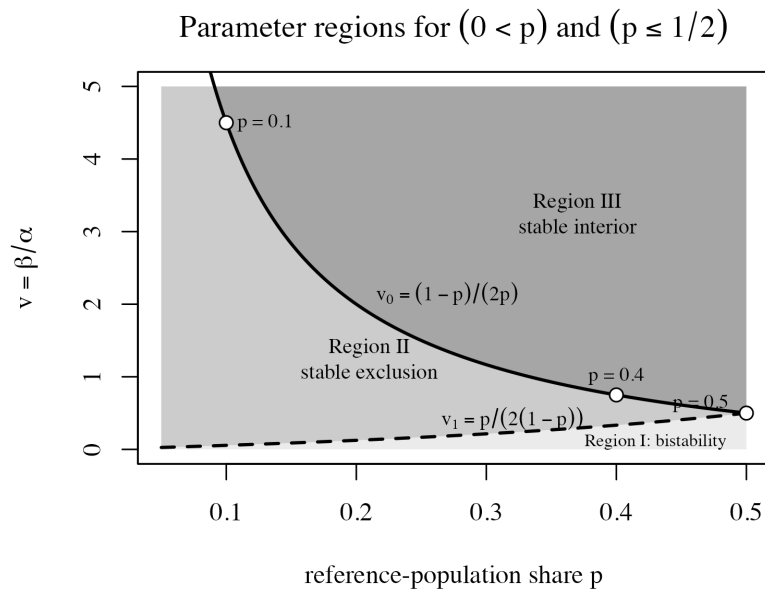


Fig 3. Parameter regions for $0 < p \leq 1/2$. The stable-interior regime lies above $v_0 = (1-p)/(2p)$, and $v_1 = p/[2(1-p)]$ separates bistability from stable exclusion. Markers show $p=0.1$, 0.4 , and 0.5 at their exclusion thresholds, illustrating the sharp increase in the required relative corrective strength as p decreases.

<https://doi.org/10.1371/journal.pcsy.0000132.g003>

These expressions reduce to v_0 and v_1 when $c(z) = z$. They are local boundary conditions and do not by themselves establish the number, uniqueness, or stability of interior equilibria.

Because $c(p) > 0$ and $c'(p) > 0$, the generalized exclusion threshold $(1 - p)/[2c(p)]$ is strictly decreasing in p . For $c(z) = z + \kappa z^3$, changing κ from -0.2 to 0.2 shifts this threshold from 4.509 to 4.491 at $p = 0.1$ and from 0.775 to 0.727 at $p = 0.4$. The high threshold for a scarce focal type therefore persists in these monotone alternatives, although its exact value changes. For the illustrative strict-minority values $p = 0.1$ and $p = 0.4$, the stated sign-change scan detected at most one interior equilibrium on the examined grid. The separate near-balanced scan detected additional interior equilibria. Because an even-multiplicity equilibrium at a bifurcation value need not change sign, these scans are diagnostic rather than exhaustive. The local threshold logic is thus retained in the alternatives examined, whereas closed-form location and uniqueness are features of the baseline specification. Full derivations and numerical details are in Sections 2–6 (Local boundary thresholds through Illustrative five-equilibrium phase line) of [S1 Appendix](#).

3.7. Multi-type equilibria and illustrative examples

At a multi-type equilibrium, all represented types have a common per-capita rate. The direct separable system also admits the mathematical potential function

$$\Phi(\mathbf{x}) = \frac{\alpha}{2} \sum_{i=1}^N p_i x_i^2 - \frac{\beta}{2} \sum_{i=1}^N (x_i - p_i)^2, \quad \dot{\Phi} = \sum_{i=1}^N x_i (f_i - \bar{f})^2 \geq 0. \quad (17)$$

The potential is a mathematical device for characterizing the dynamics, not a measure of social welfare or normative fairness. The structural term favors concentration through positive representation feedback, whereas the corrective term lowers the potential in proportion to the squared distance from the reference composition and thereby favors states closer to that composition. Under the sufficient strong-correction condition $v > \max_i p_i$, the potential is strictly concave on the simplex, so the full simplex has a unique locally asymptotically stable equilibrium, although unstable rest points may remain on invariant boundary faces.

Suppose that the reference shares are distinct and ordered as $p_1 > p_2 > \dots > p_N$. Under the sufficient strong-correction condition $v > \max_i p_i$, the stable support (the set of types with positive equilibrium shares) has the form $\{1, \dots, m\}$ and expands as v crosses the ordered support-entry thresholds $\tau_2 < \dots < \tau_N$. At $v = \tau_m$, type m , which is excluded just below the threshold, has the same per-capita rate as the represented types $1, \dots, m - 1$. Substituting their common equilibrium rate into this equality yields, for $m = 2, \dots, N$, the following equation, whose unique solution above p_1 is τ_m :

$$\sum_{i=1}^{m-1} \frac{\tau_m (p_i - p_m)}{\tau_m - p_i} = 1. \quad (18)$$

Thus, the single two-type coexistence threshold generalizes to a hierarchy of support-entry thresholds: increasing corrective strength progressively adds lower-prevalence types to the stable support, and partial coexistence is possible between successive thresholds. At $v = \tau_m$, the entering type has zero equilibrium share and the support-entry transition is non-hyperbolic. If some reference shares are tied, types with the same reference share have identical per-capita rates when absent, so their invasion equalities coincide and they may enter simultaneously. The general-support formulas and proof of the ordering are given in [S1 Appendix](#).

[Table 3](#) summarizes three- and four-type examples.

At $v = 1$, the stable equilibria are $\mathbf{x}^* = (93, 29, 9)/131 \simeq (0.7099, 0.2214, 0.0687)$ for the three-type example and $\mathbf{x}^* = (29, 17, 8, 1)/55 \simeq (0.5273, 0.3091, 0.1455, 0.0182)$ for the four-type example. All types coexist, but the

Table 3. Support-entry thresholds in the separable multi-type extension.

Example	Support-entry thresholds	Stable-support sequence
$N=3$ $\mathbf{p} = (0.5, 0.3, 0.2)$	$\tau_2 = 0.6250$ $\tau_3 = 0.7791$	$\{1\} \rightarrow \{1, 2\} \rightarrow \{1, 2, 3\}$
$N=4$ $\mathbf{p} = (0.4, 0.3, 0.2, 0.1)$	$\tau_2 = 0.4444$, $\tau_3 = 0.5392$ $\tau_4 = 0.8579$	$\{1\} \rightarrow \{1, 2\} \rightarrow \{1, 2, 3\}$ $\rightarrow \{1, 2, 3, 4\}$

Thresholds refer to the unique stable equilibrium under the sufficient strong-correction condition $v > \max_i p_i$. Equality cases are nonhyperbolic support-entry transitions.

<https://doi.org/10.1371/journal.pcsy.0000132.t003>

lowest-prevalence shares remain below their reference values: $0.0687 < 0.2$ and $0.0182 < 0.1$, respectively. Full coexistence therefore does not imply proportional representation, preserving the distinction identified in the two-type model. Fig B in Section 10 of [S1 Appendix](#) plots the stable shares and support-entry transitions over the sufficient strong-correction region $v > \max_i p_i$.

These results apply to the direct separable symmetric extension. Outside this sufficient strong-correction region, multiple boundary or partial-coexistence equilibria may coexist. Models with cross-type spillovers, shared-attribute effects, or type-specific parameters need not preserve the same potential or ordered threshold structure.

4. Discussion

We constructed a minimal population model in which representation-dependent structural feedback competes with corrective action. The central objects are type-specific per-capita effective growth or retention rates. This interpretation makes it possible to connect social mechanisms to a tractable threshold without treating the payoff terms as aggregate opportunity counts. The model distinguishes three outcomes: stable exclusion, persistent but sub-proportional representation, and an equilibrium that approaches the reference share as corrective strength becomes large.

4.1. From discrimination mechanisms to dynamical stability

Taste-based models center on discriminatory preferences, while statistical-discrimination models center on inference about individuals under imperfect information [1–4]. These approaches remain essential for analyzing discrimination in markets and selection. The present model addresses a complementary mechanism that their classical baseline forms do not directly endogenize: current representation changes subsequent type-specific per-capita rates. It therefore does not require an explicitly discriminatory preference or a contemporaneous belief error to produce persistent inequality. Recent dynamic models likewise show that the time path of discrimination can reveal its source [5]; our focus is specifically on the stability consequences of representation feedback.

This formulation contributes three dynamic objects to the comparison. It represents historical dependence through basins of attraction, persistent exclusion through a stable boundary equilibrium, and intervention through a change in equilibrium stability. If $x=0$ is stable, the absence of type A is not merely a temporary imbalance but a locally self-reinforcing state. Corrective action is then evaluated by whether its strength changes that stability, not merely by whether a policy has been adopted.

4.2. Cumulative disadvantage, thresholds, and representation effects

The structural term is related to cumulative advantage and disadvantage: earlier differences in representation can alter later access to relevant resources [14–16]. Representation-dynamic models similarly show that participation can change group-specific norms [11], while mentoring models show that support quality can depend on group composition [12]. In our reduced form, the x inside a payoff captures this representation-dependent change in the per-capita rate. It is distinct from the prefactor x in the replicator equation, which frequency-weights the payoff difference.

The model is also related to threshold and segregation models. Schelling's model of segregation and Granovetter's threshold model showed that local rules or threshold responses can generate large aggregate changes [18,19]. Our model follows the same general logic that proportions matter. However, it differs in two respects. First, it describes entry or continuation in a target population using replicator dynamics rather than spatial segregation or collective action. Second, it introduces both a reference share p and a corrective strength β , allowing the threshold for destabilizing underrepresentation to be derived explicitly.

Organizational and educational studies provide examples of mechanisms that may contribute to such rates. Group proportions can affect experience and evaluation [17], and peer mentoring can affect persistence [13]. Information, role models, mentoring, informal networks, and institutional fit may each contribute to a representation-dependent rate, but the baseline model does not identify their separate effects. Its purpose is to derive the stability consequences of their combined feedback.

4.3. Corrective action and fairness as dynamic intervention problems

Prior research shows that the effects of affirmative action depend on institutional design, incentives, initial conditions, and stereotypes [20–24]. Our result adds a relative-strength interpretation. The parameter β can be raised by more effective targeted recruitment, support, mentoring, or information provision. The ratio β/α can also be raised by reducing α through structural changes that weaken dependence on existing networks, informal sponsorship, or inherited institutional fit. These are distinct policy pathways even when they have the same mathematical effect on the ratio.

At v_0 , the system changes from one in which a small positive representation declines to one in which it grows. Yet Eq. (13) shows that this first policy level only avoids stable exclusion. A second, generally stronger level is required to bring the equilibrium within a chosen distance of p . Because a finite corrective force leaves a residual structural difference in the per-capita rates, continued intervention may be needed until the underlying feedback has weakened.

The model does not directly define individual fairness or static group fairness [25–27]. Instead, it examines whether the deviation between x and p decreases or persists over time. In this sense, fairness is treated as a dynamic stability problem. When structural barriers are present, applying identical criteria at a single point in time may not remove accumulated disadvantages. A dynamic model is therefore useful for examining how interventions affect long-term representation.

4.4. A tractable threshold formulation

Replicator equations are standard in evolutionary game theory and population games [28–31]. The present model uses that framework to convert a qualitative claim about structural reproduction into explicit conditions under which exclusion remains stable, becomes unstable, or is replaced by coexistence. Its closed-form expressions also separate conclusions that follow from boundary effects from conclusions that depend on the baseline functional form.

O'Connor analyzed how social categories can generate unfairness through evolutionary interactions [33]. Our model complements this approach. We do not model the emergence of social categories. We assume that categories already exist and ask how the representation of one type in a target population is maintained or changed. The key object is the relative strength β/α , which determines whether the exclusionary state loses stability and whether an interior equilibrium emerges. The value of the model is therefore not only that it uses a replicator equation, but that it identifies the relative intervention strength required to change the stability of long-run representation.

Persistent representation and proportional representation are therefore different. For $p < 1/2$ and finite v , a stable interior equilibrium satisfies $x^* < p$. The tolerance condition in Eq. (14) quantifies how much stronger the corrective force must be to move beyond nonzero representation and approach the reference share. The share p is a model benchmark, however, not a claim that one proportion is the uniquely appropriate normative allocation in every institution.

4.5. Asymmetric and multi-type extensions

Relaxing structural symmetry can shift the exclusion threshold. When the two types have distinct structural strengths, the local condition under which a small positive representation of type *A* grows depends on the structural strength associated with the locally dominant type *B*. Institutional or network asymmetry can therefore alter the corrective strength required to destabilize exclusion. The asymmetric per-capita rates and boundary derivation are provided in [S1 Appendix](#).

The direct separable symmetric *N*-type extension preserves the comparison between structural feedback and corrective action while introducing a richer equilibrium structure. Under the sufficient strong-correction condition $v > \max_i p_i$, the single two-type coexistence threshold becomes an ordered hierarchy of support-entry thresholds, so increasing corrective strength expands the stable support in stages. Partial-coexistence states are therefore possible: some intersectional categories may remain excluded even when others are represented. For distinct ordered reference shares, lower-prevalence types enter the stable support later in this direct extension.

Treating gender-by-race combinations as distinct types is only a categorical approximation. The current extension treats these categories as mutually exclusive and omits spillovers between categories that share gender, race, or another attribute. Such effects could be represented by a cross-type interaction matrix, for example

$$f_i = \alpha p_i \sum_{j=1}^N M_{ij} x_j + \beta (p_i - x_i).$$

For a general matrix M_{ij} , the mathematical potential and ordered threshold hierarchy derived here need not be preserved. Cross-type interactions, shared-attribute effects, and type-specific structural or corrective parameters remain directions for future work.

4.6. Limitations and future work

First, the bilinear structural term and linear corrective term are reduced-form baseline approximations, not exact empirical laws. The nonlinear analysis shows that local boundary thresholds can be generalized, but their numerical values change with the corrective function and additional interior equilibria can arise. Closed-form location and uniqueness of the interior equilibrium are therefore specification-dependent. The baseline also uses the same p as both a prevalence factor in the structural term and the benchmark for corrective action; future work could separate these roles with distinct parameters.

Second, each per-capita payoff aggregates entry, continuation, promotion, and exit together with mechanisms such as information, role models, mentoring, informal networks, and institutional fit. The model cannot identify which pathway produces an observed feedback. Its symmetric treatment of types is also a simplifying baseline rather than an empirical claim about institutions faced by dominant and marginalized groups. Asymmetric parameters and mechanism-specific transition models would be needed for such inference.

Third, the main analysis remains two-type because it provides transparent closed-form thresholds and a complete one-dimensional bifurcation diagram. The direct separable symmetric *N*-type extension derives general support equilibria and an ordered support-entry hierarchy under common structural and corrective strengths. It also represents intersectional categories as distinct mutually exclusive types, consistent with work showing that a policy defined along one attribute can miss inequality at combinations of attributes [\[35–37\]](#). However, it omits cross-type spillovers, shared-attribute effects, attribute-specific corrective policies, and type-specific parameters; these richer models may not admit the same potential representation or threshold ordering. The baseline population is also deterministic and well mixed; it omits finite-population noise, network structure, local interaction, and heterogeneity across organizations.

Fourth, p , α , and β are constant over time. Real reference populations, institutions, and interventions can evolve, and feedback may operate with delays. Time-varying, stochastic, network, and organization-specific extensions may change transient dynamics and the relevant policy horizon.

Finally, the parameters and nonlinear functions have not been empirically estimated. Longitudinal data separating entry, retention, promotion, and exit would be needed to identify their contributions and to test the proposed feedback. The choice of reference population and the use of p as a proportional benchmark are also normative and context dependent. The model clarifies the consequences of a chosen benchmark; it does not determine which benchmark an institution ought to adopt.

5. Conclusions

We proposed a replicator model in which each type's per-capita effective growth or retention rate contains a reduced-form structural component that increases jointly with its share in a reference population and its current representation in a target population. In the baseline, the same reference share is also the benchmark for corrective action.

The analysis clarifies three points. A small positive representation grows only when corrective action exceeds the structural boundary effect; that exclusion-avoidance threshold is higher for a rarer focal type; and crossing it does not by itself produce proportional representation. For $p < 1/2$, every finite corrective strength leaves the stable baseline equilibrium below p , and a stronger condition is required to approach the benchmark within a chosen tolerance.

Nonlinear and asymmetric extensions preserve the comparison between corrective and structural boundary effects but can change exact thresholds and equilibrium structure. The direct separable symmetric multi-type extension replaces the single coexistence threshold with a hierarchy of support-entry thresholds and permits partial coexistence. More general interaction-rich and intersectional models remain an important direction for future work. The baseline is therefore best understood as a transparent reference model for testing when corrective action changes the local stability of exclusion, rather than as a complete empirical law of structural discrimination.

Supporting information

S1 Appendix. Robustness and extensions of the baseline model. Analytical derivations and numerical analyses for nonlinear two-type feedback, asymmetric structural strengths, and the direct separable symmetric multi-type extension. (PDF)

S1 Code. Code used to reproduce the figures and model-derived numerical values. The script generates Figs 1–3, Figs A and B in S1 Appendix, threshold-table values, nonlinear grid-scan summaries, and the multi-type equilibrium calculations. (R)

Author contributions

Conceptualization: Masanori Takano.

Formal analysis: Masanori Takano.

Investigation: Masanori Takano.

Methodology: Masanori Takano.

Project administration: Masanori Takano.

Resources: Masanori Takano.

Supervision: Masanori Takano.

Validation: Masanori Takano.

Visualization: Masanori Takano.

Writing – original draft: Masanori Takano.

Writing – review & editing: Masanori Takano.

References

1. Becker GS. The economics of discrimination. Chicago: University of Chicago Press. 1957.
2. Phelps ES. The statistical theory of racism and sexism. *Am Econ Rev*. 1972;62(4):659–61.
3. Arrow KJ. The Theory Of Discrimination. *Discrimination in Labor Markets*. Princeton University Press. 2015. p. 1–33. <https://doi.org/10.1515/9781400867066-003>
4. Aigner DJ, Cain GG. Statistical Theories of Discrimination in Labor Markets. *Industrial and Labor Relations Review*. 1977;30(2):175. <https://doi.org/10.2307/2522871>
5. Bohren JA, Imas A, Rosenberg M. The dynamics of discrimination: theory and evidence. *Am Econ Rev*. 2019;109(10):3395–436. <https://doi.org/10.1257/aer.20171829>
6. Pincus FL. Discrimination comes in many forms: individual, institutional, and structural. *Am Behav Sci*. 1996;40(2):186–94. <https://doi.org/10.1177/0002764296040002009>
7. Reskin BF. Including mechanisms in our models of ascriptive inequality. *Am Sociol Rev*. 2003;68(1):1–21. <https://doi.org/10.2307/3088900>
8. Pager D, Shepherd H. The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets. *Annu Rev Sociol*. 2008;34:181–209. <https://doi.org/10.1146/annurev.soc.33.040406.131740> PMID: 20689680
9. Small ML, Pager D. Sociological perspectives on racial discrimination. *J Econ Perspect*. 2020;34(2):49–67. <https://doi.org/10.1257/jep.34.2.49>
10. Ray V. A theory of racialized organizations. *Am Sociol Rev*. 2019;84(1):26–53. <https://doi.org/10.1177/0003122418822335>
11. Carvalho J-P, Pradeliski BSR. The representation dynamic and the “normalization” of group differences. *The Journal of Law, Economics, and Organization*. 2024;41(3):996–1024. <https://doi.org/10.1093/jleo/ewae014>
12. Müller-Itten M, Öry A. Mentoring and the Dynamics of Affirmative Action. *American Economic Journal: Economic Policy*. 2022;14(2):402–44. <https://doi.org/10.1257/pol.20190579>
13. Wu DJ, Thiem KC, Dasgupta N. Female peer mentors early in college have lasting positive impacts on female engineering students that persist beyond graduation. *Nat Commun*. 2022;13(1):6837. <https://doi.org/10.1038/s41467-022-34508-x> PMID: 36369261
14. Merton RK. The Self-Fulfilling Prophecy. *The Antioch Review*. 1948;8(2):193. <https://doi.org/10.2307/4609267>
15. Merton RK. The Matthew Effect in Science. *Science*. 1968;159(3810):56–63. <https://doi.org/10.1126/science.159.3810.56>
16. DiPrete TA, Eirich GM. Cumulative advantage as a mechanism for inequality: a review of theoretical and empirical developments. *Annu Rev Sociol*. 2006;32:271–97. <https://doi.org/10.1146/annurev.soc.32.061604.123127>
17. Kanter RM. Some effects of proportions on group life: skewed sex ratios and responses to token women. *Am J Sociol*. 1977;82(5):965–90. <https://doi.org/10.1086/226425>
18. Schelling TC. Dynamic models of segregation. *J Math Sociol*. 1971;1(2):143–86. <https://doi.org/10.1080/0022250X.1971.9989794>
19. Granovetter M. Threshold models of collective behavior. *Am J Sociol*. 1978;83(6):1420–43. <https://doi.org/10.1086/226707>
20. Lundberg SJ, Startz R. Private discrimination and social intervention in competitive labor markets. *Am Econ Rev*. 1983;73(3):340–7.
21. Coate S, Loury GC. Will affirmative-action policies eliminate negative stereotypes?. *Am Econ Rev*. 1993;83(5):1220–40.
22. Moro A, Norman P. Affirmative action in a competitive economy. *Journal of Public Economics*. 2003;87(3–4):567–94. [https://doi.org/10.1016/S0047-2727\(01\)00121-9](https://doi.org/10.1016/S0047-2727(01)00121-9)
23. Holzer H, Neumark D. Assessing affirmative action. *J Econ Lit*. 2000;38(3):483–568. <https://doi.org/10.1257/jel.38.3.483>
24. Fryer RG Jr, Loury GC. Affirmative action and its mythology. *J Econ Perspect*. 2005;19(3):147–62. <https://doi.org/10.1257/089533005774357888>
25. Dwork C, Hardt M, Pitassi T, Reingold O, Zemel R. Fairness through awareness. In: *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, 2012. 214–26. <https://doi.org/10.1145/2090236.2090255>
26. Hardt M, Price E, Srebro N. Equality of opportunity in supervised learning. *Advances in Neural Information Processing Systems*. 2016;29:3315–23.
27. Kleinberg J, Mullainathan S, Raghavan M. Inherent trade-offs in the fair determination of risk scores. In: *Proceedings of the 8th Innovations in Theoretical Computer Science Conference*, 2017. 43:1–43:23. <https://doi.org/10.4230/LIPIcs.ITCS.2017.43>
28. Taylor PD, Jonker LB. Evolutionarily stable strategies and game dynamics. *Math Biosci*. 1978;40(1–2):145–56. [https://doi.org/10.1016/0025-5564\(78\)90077-9](https://doi.org/10.1016/0025-5564(78)90077-9)
29. Weibull JW. *Evolutionary game theory*. Cambridge (MA): MIT Press. 1995.
30. Hofbauer J, Sigmund K. *Evolutionary games and population dynamics*. Cambridge: Cambridge University Press. 1998.
31. Sandholm WH. *Population games and evolutionary dynamics*. Cambridge (MA): MIT Press. 2010.
32. Takano M, Nakazato K, Taka F. Dynamics of discrimination and prejudice via two types of social contagion. *Appl Math Comput*. 2023;448:127916. <https://doi.org/10.1016/j.amc.2023.127916>
33. O'Connor C. *The origins of unfairness: social categories and cultural evolution*. Oxford: Oxford University Press. 2019.
34. U.S. Bureau of Labor Statistics. Employed people by detailed occupation, sex, race, and Hispanic or Latino ethnicity, 2025 annual averages. U.S. Bureau of Labor Statistics. <https://www.bls.gov/cps/cpsaat11.htm> 2026. Accessed 2026 July 29.

35. Carvalho J-P, Pradelski BS. Identity and underrepresentation: Interactions between race and gender. *Journal of Public Economics*. 2022;216:104764. <https://doi.org/10.1016/j.jpubeco.2022.104764>
36. Carvalho JP, Pradelski BSR, Williams C. Intersectionality: affirmative action with multidimensional identities. *Manag Sci*. 2025;71(5):4495–509. <https://doi.org/10.1287/mnsc.2023.03839>
37. Mehrotra A, Pradelski BSR, Vishnoi NK. Selection in the Presence of Implicit Bias: The Advantage of Intersectional Constraints. In: 2022 ACM Conference on Fairness Accountability and Transparency, 2022. 599–609. <https://doi.org/10.1145/3531146.3533124>