Theoretical Approaches in Stratification Economics

Brendan Brundage[*], Dan McGee[†], and Daniele Tavani[‡]

October 1, 2025

Abstract

We review new developments in formal theory as applied to racial stratification. Across this literature, we discuss how discrimination produces benefits for advantaged groups at a cost to marginalized groups, how groups resolve conflicts between individual and collective interest, how direct discrimination is transformed into persistent inequality, and the implications of these forces for racial disparities, efficiency, and social welfare. For researchers active in Stratification Economics, we highlight parallels between intuitions about dimensions of racial discrimination and inequality developed by SE and insights from formal models by theorists outside SE. For formal theorists interested in working on questions of race and inequality, we provide a overview of how SE conceptualizes racial stratification and identify open questions where formal theory can provide more rigorous micro-foundations to these intuitions.

[*]Morehouse College
[†]University of Toronto
[‡]Colorado State University

1 Introduction

Stratification Economics (SE) began as a research program on intergroup conflict, with empirical studies capturing the dimensions of racial inequality in the United States and worldwide, and with descriptive work critiquing mainstream approaches to the study of race and discrimination as failing to capture crucial mechanisms in the creation, maintenance, and transmission of group-based inequality across generations.[1] More recently, various papers have drawn explicitly on insights from the SE tradition to develop formal models illustrating the mechanisms underlying racial divides. In this chapter, we survey theoretical work, both explicitly within Stratification Economics and within 'mainstream' economic theory but consistent with ideas studied in SE, and highlight four broad questions where formal theory can contribute to the field's understanding of racial inequality and specific open issues within these questions. We emphasize insights from formal theory that economists working in the SE tradition can incorporate into their work, even without developing formal models themselves, and insights from SE that may provide interesting avenues for research for mainstream formal theorists.

In a recent review essay, Lefebvre (2025) identifies three core ideas shared across Stratification Economics. First, racism is best understood through models of rational, self-interest behavior. SE argues that social rewards are fundamentally rivalrous, so an increase in rewards for one group will often come at a cost to others. Hence, prejudice and discrimination are strategies to build and perpetuate privilege in this world of unequal rewards. As such, racial inequality in the past and present exists due to purposeful actions from those who benefit from racial hierarchy. Second, social groups, not just individuals, are central objects for economic analyses. As we argue below, since groups can offer material and psychological benefits to their members, collective interests are relevant for understanding individuals' motivations. SE is primarily interested in groups which are at most semi-permeable as any groups which confer substantial benefits and can be joined at low or zero cost will quickly see their benefits become diluted by new entrants and

[1] For example, Darity (2005) argues that models which rely on the continued dysfunction of marginalized groups fail to explain why racial inequality persists because the successful acquisition of human capital does not protect minorities from discrimination and is not supported by the empirical data.

groups which suffer substantial disadvantages will quickly disappear if exiting the group is costless. Finally, the inter-generational transmission of advantage and disadvantage through inheritances of wealth, institutions, and power cannot be overlooked. As individuals within the dominant group receive benefits from their position at the top of the social hierarchy, they have an incentive to maintain this superior position. Thus, both market and non-market institutions are structured to maintain the existing racial hierarchy.

We structure this chapter around four fundamental questions that we believe benefit from formal theory, two at the 'microeconomic' level (Section 2) and two at the 'macroeconomic' level (Section 3). At the micro level, we investigate the specific channels through which discrimination creates or preserves benefits for the advantaged group and how social groups coordinate effort towards collective goals, even when this conflicts with individual interests. In these analyses, we take seriously the challenges posed by notions of group agency and collective interests with the goal of providing rigorous microfoundations to our intuition that racial identities matter greatly for individual behavior. At the macro level, we focus on the aggregate consequences of strategic discrimination for the distribution of income and wealth and the effects on growth and efficiency, as well as the channels through which discrimination becomes persistent and systemic. Given the work at the micro level, in models focusing on the macro consequences of discrimination, we argue that it can be acceptable to simplify and focus on group agency, understanding that groups have mechanisms to resolve intra-group conflicts.[2]

Before we turn to these four questions, it behooves us to elaborate for non-theorists, especially those active in Stratification Economics, what formal theory can add to their work. SE emerged and developed in conversation with other disciplines within the social sciences, such as sociology, history, and critical legal studies, where formal theory is less prevalent than in economics. As such, formal models are far less common in SE than in economics more broadly.[3] However, bringing formal theory to bear on questions of race

[2]In the spirit of Debortoli and Gali (2025), one could think of a two-group macroeconomic model as a simplification of models with both intra- and inter-group conflict that can still capture some of the relevant insights and dynamics.

[3]As two rare exceptions, Darity et al. (2006) and Mason et al. (2022) present models of racial identity formation; however, as agents are automata following a simple replicator dynamic, the models are limited

is likely to aid the further advancement of SE. Tools developed in game theory and mechanism design allow economists to conduct sophisticated analyses of social interactions, and given that race is a social construct, i.e., an equilibrium phenomenon, it is well-suited to the application of these tools with several benefits as we outline here:

- First, by expressing our ideas in formal models, we are forced to be precise about the meanings of objects we study. Although this may cut down on the breadth of what we are able to say, we gain much from this precision. For example, precise statements make it easier to root out inconsistencies in our thinking and to identify testable and measurable predictions from theory.

- Second, formal tools also allow us to push models further and identify more non-obvious implications of our theories than informal arguments often allow. Hence, we can understand all the logical consequences of our assumptions embedded in the model.

- Third, theory pushes us to strive for both *general* and *generalizable* mechanisms, beyond a single motivating example. For example, racial boundaries are constructed differently in the Caribbean than in the United States, but the same underlying process of labor conflict underlies both racial systems.[4] These general mechanisms help us understand the historical contingency of race; given an alternate sequence of events, racial divisions could have been constructed differently and the hierarchical ranking of groups could have been different. Yet without an understanding of the process by which they were formed, we cannot see how our history created today's inequalities.

- Lastly, if we do not construct an explicit theory of race, it is easy to fall back on 'folk' understandings of race, often shaped by ideological narratives in society about what race *is* and what race *does*, such as the biologically or culturally deterministic views

in their ability to illustrate the mechanisms behind discriminatory behavior and many of the insights in the papers are only discussed informally and are not captured in the models.

[4] Additionally, identities like Native, Asian, Hispanic are also socially constructed but not identically to the Black/White racial distinction and not for identical purposes. See also chapters by Deshpande, Paul and Smith, and Aja et al. in this volume.

criticized by Darity (2005). Instead, by applying these theoretical tools, we can make clear what it means for race to be a socio-political tool used to mark individuals for different treatment in a world where groups are ranked hierarchically and receive greater or lesser benefits and access to resources than each other.

2 Stratification in the Micro

At the 'micro' level, theoretical work on SE contributes to two fundamental questions about racial group behavior: first, through what mechanisms does discrimination operate to create and/or perpetuate advantages for socially dominant groups and disadvantages for marginalized groups? Second, how do racial groups 'cohere' so that individuals take actions that benefit their groups, even if those actions may be individually costly, and how do these solutions to the collective action problem differ between advantaged and disadvantaged groups?

2.1 How Discrimination Benefits the Advantaged

A core tenet of SE is that discrimination arises to provide material benefits to advantaged groups at the expense of marginalized groups. As such, this sidesteps the traditional dichotomy in economic theory between discrimination on the basis of preferences and discrimination on the basis of beliefs, as surveyed by Onuchic (2024). This focus on the material and distributional consequences of discrimination allows us to distinguish between the proximate cause of a specific discriminatory act, which may include an agent's prejudice or their biased beliefs, and the ultimate causes of discrimination and the resulting inequality between socially-stratified groups. Based on this notion of discrimination to create and protect material advantages, we define such behavior as *strategic discrimination*, which McGee (2025c) describes as a pattern where agents discriminate to reduce competition, to divert benefits to themselves, or to exploit the labor of discriminated groups. Importantly, these discriminatory gains for the advantaged group come at the expense of the marginalized group and often at a cost to total welfare. Additionally,

we note that while strategic discrimination benefits the advantaged group as a whole, these benefits may be unevenly distributed within this group, so gains for the group in aggregate do not preclude that some in the advantaged group can be worse off than in the absence of discrimination. Furthermore, while purely hedonic or psychological benefits can serve as a proximate cause of discrimination (see Section 2.2.2) and some members of the advantaged group may only receive these psychological benefits, we argue that discrimination which creates material harms for all members of the advantaged group is unlikely to persist.

We characterize two forms of strategic discrimination, one found primarily in 'horizontal' environments where agents compete or cooperate directly and one found in 'vertical' environments where a mechanism designer can discriminate to affect the behavior of various agents whose actions affect the designer's payoffs. In horizontal environments, strategic discrimination dissuades marginalized groups from competing against the advantaged groups and redirects their effort into tasks where their labor benefits the advantaged group, which we term the 'divert and exploit' pattern of strategic discrimination. In vertical environments, the designer typically wants to elicit uniformly higher levels of a relevant action (e.g., effort from employees) or uniformly lower levels of the action (e.g., attacks from revolutionaries) and discriminating against one group in favor of the other can shift both group's actions in the designer's desired direction, a form of 'divide and conquer.' In both cases, strategic discrimination is an equilibrium phenomenon arising from the interactions and interdependencies between agents. Agents from the advantaged group engage in strategic discrimination because it causes the marginalized group (and other members of the advantaged group) to change their behavior in ways that benefit the advantaged group as a whole.

Finally, strategic discrimination is frequently, albeit not universally, costly to total welfare. That is, the gains to the advantaged group are outweighed by the magnitude of the losses suffered by the marginalized group. Often, these losses emerge because discrimination creates misallocation of effort across individuals by group identity (Ashraf et al., 2023). Hence, strategic discrimination can exhibit a pattern reminiscent of classic rent seeking: groups exert wasteful effort to divert benefits to themselves, harming the oppos-

ing group and social welfare (Chelwa et al., 2022). However, there can exist two countervailing forces that reduce this welfare cost and in special cases, may even outweigh the cost of misallocation. First, there may exist large gains from specialization alongside large costs from mis-coordination. The former implies that assigning asymmetric roles to different groups of agents produces significantly more surplus than having each group behave symmetrically, while the latter implies that within-group randomization between each asymmetric role cannot produce the same welfare gains. As such, the coordinating role of group identity can produce large enough gains for the advantaged group that social welfare is higher in the presence of strategic discrimination, despite the cost to the disadvantaged group. Second, Eeckhout (2006) highlights that if a group anticipates discrimination from an opposing group, this may lead them to trust their ingroup more and invest more strongly in ingroup ties. He shows that when a lack of trust is sufficiently costly, the gains from greater cooperation and investment during within-group interactions outweighs the losses from reduced between-group cooperation. However, consistent with the typical pattern of strategic discrimination, these gains are larger for the dominant group than for the marginalized group, so these benefits of within-group trust must be sufficiently large for between-group discrimination to be a Pareto improvement.

2.1.1 Divert and Exploit

To illustrate the basic mechanics of 'divert and exploit' in horizontal environments, let there be two agents, 1 and 2 who choose effort $e \in \{L, M, H\}$, where an agent can represent an individual or a group engaging in coordinated action. Each agent has a group identity, either Black, B, or White, W and payoffs for each agent are determined by the joint effort provision, $\pi_i(e_i, e_j)$. As this pattern arises most prevalently in environments of strategic substitutes, we assume that the payoff function exhibits decreasing differences:

$$\pi_i(e_i, e_j) - \pi(e_i', e_j) \leq \pi_i(e_i, e_j') - \pi_i(e_i', e_j') \text{ for } e_i > e_i', e_j > e_j'$$

with some inequalities strict. Hence, the marginal return to higher effort is lower when the opposing agent exerts more effort. Within this environment, we consider both settings where the opposing agent's effort is harmful, which occurs when:

$$\pi_i(e_i, e_j) < \pi_i(e_i, e'_j) \text{ for } e_j > e'_j$$

and settings where the other agent's effort is beneficial, when the prior inequality is reversed. For example, the consumption/use of a common good exhibit strategic substitutes with negative payoff spillovers as the productivity of the common good falls as it is more intensively used. By contrast, the production of a public good represents strategic substitutes with positive spillovers as both agents benefit from greater production but this production exhibits diminishing returns to effort.[5] Finally, we say that an agent discriminates if her strategy conditions on the opposing agent's group identity.[6]

Suppose first that payoffs are such that low effort and high effort are mutual best response and medium effort is best response to itself. Hence, the game admits two asymmetric equilibria in pure strategies, (L, H) and (H, L), and one symmetric equilibrium (M, M). These asymmetric equilibria are familiar outcomes in canonical games of strategic substitutes; in the game of chicken, one agent swerves and the other does not, in a public goods game, one agent contributes and the other does not. However, in the absence of group identities, there may not be a clear coordination device for two randomly matched agents to use to determine who exerts high effort and who chooses low effort. Instead, the symmetric equilibrium presents an intuitively appealing outcome for these one-shot interactions.[7] For this reason, (M, M) is a natural prediction for within-group interactions, those where agents 1 and 2 share a group identity, irrespective of whether both agents are Black or are White. Therefore, if an agent chooses M when interacting with the other group, that agent is following a non-discriminatory strategy, and they dis-

[5] In principle, payoffs functions need not be symmetric, such as in team production between a single manager and many employees; in this chapter, we treat them as symmetric for expositional purposes.

[6] Hence, in this section, we focus on *direct discrimination* as defined by Bohren et al. (2025); we discuss *systemic discrimination* in Section 3.7.

[7] Whether the symmetric equilibrium offers higher expected utility than an equilibrium in mixed strategies will depend both on the efficiency gains or losses from specialization, i.e., the payoffs to (L, H) and (H, L) versus (M, M), and on the costs of mismatch, i.e., the losses from playing (L, L) or (H, H).

criminate if they choose either L or H. Unlike within-group interactions which require an external coordinate device, group identities themselves can serve to coordinate behavior on an asymmetric equilibrium when agents have different group identities. In this sense, racial identity acts as a social norm as per Mason (2023) since codes of intergroup interactions are determined by agents' identities.

As a first example, consider the Nash Demand game: each agent submits a claim for a fraction of the surplus, $e_i \in \{1/4, 1/2, 3/4\}$. If the claims are compatible, $e_i + e_j \leq 1$, agents receive their claim, but if the two claims are incompatible, both receive zero. Hence, this captures a simple form of competition over a scarce resource.[8] It is straightforward to observe that this game exhibits decreasing differences with negative payoff spillovers and has three pure-strategy equilibria as described above. In this setting, the symmetric equilibrium involves an equal split of the surplus—each agent receiving their 'fair share'—while in each of the asymmetric equilibria, one agent claims more than their opponent. If White agents submit claims of $3/4$ when interacting with Black agents, but only request $1/2$ with other White agents, we observe that these White agents are engaging in strategic discrimination.

In equilibrium, White agents are made better off by this discrimination as they receive more surplus and Black agents are worse off. Nevertheless, both agents are behaving rationally; given that White agents are making more aggressive claims, Black agents rationally reduce their requests when faced with a White opponent, and since Black agents make smaller claims in between-group interactions, Whites know that they can press these claims. If viewed in isolation, a naive observer might conclude that Blacks are behaving dysfunctionally by claiming less than otherwise-identical Whites. Yet as this simple example makes clear, their choices would differ if not for Whites' strategic discrimination. We also note that in contrast to the predictions of many taste-based discrimination models, Whites would prefer to interact with a Black opponent than with a member of their own group as they receive exploitative gains that are unavailable with

[8]Many generalizations of the Nash Demand game are possible, such as where larger demands on the surplus decrease the total surplus produced, as in a tragedy of the commons, or where the probability that the claims are paid out is gradually decreasing in the amount by which the joint claims exceed some threshold, rather than dropping immediately to zero.

another White agent.[9]

Another archetypical environment for divert and exploit under negative spillovers are competitive filters as described in McGee (2025c), where individuals must exert effort and/or invest in specialized skills before the competition resolves and the return to these investments is higher for those who win the competition compared to those who lose. The presence of this filter implies that human capital investments by different social groups are not separable, unlike the standard setting of Arrow (1973). Put another way, when explaining identity-group disparities, the notion of individuals as pure price-takers does not accurately capture the market, especially when there is labor market congestion and social, political, or economic power. For example, agents could make human capital investments before entering a Tullock contest, where the winner gains admittance to a top university or employment in a high-productivity firm, while the loser cannot. Unlike the standard Tullock contest, if an agent's payoffs are sufficiently responsive to their pre-market human capital investment, the competitive filter will exhibit global strategic substitutes, incentivizing strategic discrimination. Just as with the Nash Demand game, in equilibrium, the marginalized group reduces their effort and investment, anticipating aggressive competition from the advantaged group. Viewed in isolation, fewer members of the marginalized group are admitted to top universities, top firms, and other competitive positions due to their performance at the competitive filter. However, this view again neglects that this behavior is the rational best response to the advantaged group's discrimination. If the advantaged group were to treat the marginalized group as no different from their own group, equilibrium effort would rise to equalize performance between the two groups.

As these two examples show, when there are negative payoff spillovers, the advantaged group discriminates by increasing the intensity of their action, inducing the marginalized group to back off. As Lewis (1985) notes, when social or market interactions are competitive, so others' efforts are harmful to one's own payoffs, one strategy to maintain dominance is to render marginalized groups non-competing with the advantaged

[9]Logan (2022) mentions that racist Whites would seek out opportunities to mistreat marginalized groups, experiencing utility from the act of domination. See also Imas and Madarasz (2024).

group (Davis, 2024). Interpreting this as strategic discrimination implies that formal and informal tools which make coordination on the asymmetric equilibrium more likely are the practices used to make a group non-competing (Chelwa et al., 2022). First, they can *de jure* discriminate by restricting the action space for the marginalized group. In the Nash Demand game, if Black agents are only legally permitted to claim $1/4$ of the surplus when interacting with a White agent, the unique equilibrium asymmetrically favors White agents and if they are not permitted to vie for the reward under a competitive filter, they will clearly never win the competition. Yet the advantaged group need not use as blunt a tool as de jure discrimination. A second strategy is to reduce the returns to higher effort for the marginalized group, so that even if they were to exert the same effort as the dominant group, they would not earn equal returns. For example, if predominantly Black schools are deprived of resources so the same effort builds less human capital, Whites advantage themselves even if the competitive filter itself is race-neutral. Additionally, if the dominant group controls the credentialing process, they can define 'merit' in their own image to ensure that marginalized individuals with ostensibly equal human capital receive lower rewards (Darity, 2001; Darity et al., 2015). Finally, cultural narratives about the appropriate place for the dominant and marginalized groups can create common expectations about what actions each will take, easing coordination on an asymmetric equilibrium.[10]

Strategic discrimination can also arise when the interaction exhibits positive spillovers such as in the production of public goods. However, as the advantaged group benefits from the marginalized group's efforts, discrimination involves less effort from the advantaged and more from the disadvantaged. By reducing their effort when interacting with the outgroup, advantaged group members can free ride on the outgroup's productive labor while contributing less to the general pool themselves. As before, the dominant group can use both formal and informal methods to elicit greater input from the marginalized group. Often, this form of strategic discrimination coincides with a 'separate spheres' ideology where the advantaged group is believed to have exclusive claim to the more

[10]Relatedly, cultural expectations about when and where women should compete with men can similarly serve to exclude them from certain valuable environments.

remunerative, high-status realm, especially when that realm's output is more easily captured by the agent. By contrast, the marginalized group is relegated to a second sphere whose outputs cannot as easily be privatized, so they receive less even as they work as much or more than the dominant group.[11] This terminology is most frequently used in the context of gender segregation, where women were supposed to occupy a private sphere within the home, conducting domestic labor to support the family, while men went out into the public sphere to labor for cash wages and exercise political power. Nevertheless, this concept also applies to ethnic and racial stratification as marginalized groups are often pushed into subservient roles and are especially prevented from overseeing the labor of a dominant-group member.[12]

In addition to prior work in SE, earlier theoretical work has identified settings where this pattern of 'divert and exploit' occurs. An early contribution is Moro and Norman (2004) who introduce a conflict of group interests into the Coate and Loury (1993) model of statistical discrimination. As the authors note, models which treat discrimination as a pure coordination failure cannot rationalize why the dominant group would resist measures to reduce discrimination against marginalized groups, as doing so would represent a Pareto improvement. However, in their model, the groups are linked through competition in the labor market. As such, the dominant group receives exploitative gains at the expense of the marginalized group, so a reduction in discrimination can only occur by eliminating the advantaged group's benefits. Similarly, Dewan and Wolton (2022) embed a model of norm-based discrimination as in Peski and Szentes (2013) into a stylized labor market (see also Section 2.2.1). They demonstrate that a norm where both employers and workers discriminate against the marginalized group is sustainable in equilibrium and yields better employment outcomes for dominant group workers at the expense of the marginalized group.

[11] See also Choy (2024) who endogenizes task assignment between social groups.

[12] For example, Black railway workers were relegated to roles as porters or serving as White engeineers' valets and were excluded from the unions of brakemen and firemen (Sundstrom, 1990; Arnesen, 1994).

2.1.2 Divide and Conquer

The second pattern of strategic discrimination is the strategy of divide-and-conquer. In broad terms, agents employ this strategy when facing an opposition whose coordinated action (intentional or not) would make it harder for that agent to achieve their goals. However, by treating different members of the opposition differently, the agent can break this coordination, allowing them to extract greater rewards.[13] Morrock (1973) describes ethnic divide and rule as a common strategy employed by colonizers to reduce opposition by creating a favored group within colonized peoples, which often creates lasting ethnic inequality even if the colonizer eventually cedes control over the territory. Drawing on theories of the social construction of race, McGee (2025a) shows that the racial categories of 'Black' and 'White,' as they are understood across the Americas, arose from labor conflict during the colonial era of plantation slavery.[14] If the plantation elite exploited all workers equally, they would have faced a serious revolt threat. However, by creating a privileged class with greater social and political rights, ensuring that this class could protect themselves from exploitation, the elite could divide workers and enable greater exploitation of the disfavored class.[15] In the Americas, racial categories of Black and White were created to divide those who could be enslaved from those who were protected from this most extreme form of exploitation. Critically, plantation owners in the United States faced distinct demographic conditions from elites in the Caribbean or South America, as the United States was colonized primarily by European families moving as a unit, unlike the almost uniformly-male colonial populations elsewhere. Thus, to form a minimum viable coalition, elites in the Caribbean and South America had to expand the boundaries of Whiteness to include mixed-race descendants of European men and native or African women, whereas elites in the North American colonies could implement a strict rule of

[13] See also the online appendix in Onuchic (2024) which discusses discriminatory institutional design from a mechanism design perspective, focusing on incentives under private information or actions.

[14] Cox (1948) provides an early articulation of this thesis, Lopez (1994) and Mills (1997) discuss the social construction of race, and Fields (1990) and Fields and Fields (2022) focus on its creation in the USA.

[15] Esteban and Ray (2008) and Bhattacharya and Mukherjee (2023) provide additional arguments as to why ethnic divisions may be more salient than class divides: in the former, conflict requires both labor and capital and as ethnic groups are more heterogeneous in wealth, can more easily provide both than class groups; in the latter, increasing economic mobility reduces agents' certainty over their future economic class without increasing uncertainty over their future ethnicity.

hypodescent where any non-European ancestry relegated one to the oppressed Black category.

The notion of the minimum viable coalition has long been present in economics and political science. Buchanan and Tullock (1962) and Ferejohn (1974) argue that politicians will tend to promise rewards to the smallest set of voters sufficient to win the election or other political contest. However, these coalitions can be unstable as those excluded from an existing coalition will accept less than what existing coalition members receive to be brought into the fold, creating an opportunity for politicians to engage in divide and conquer (Frohlich, 1975; Ferejohn, 1986). SE emphasizes that although group identities are political and social constructions, they become meaningful divisions amongst those on either side of the color line. Thus, divide-and-conquer strategies that use these constructed color lines are more likely to persist than divisions orthogonal to the social divisions that people think matter.

A similar pattern occurs in markets even in the absence of explicit coercion. Lagerlof (2020) develops a model of labor market segmentation where a pair of firms compete over workers and this duopsony implies each has wage-setting power. He shows that if one firm commits to discriminating against the minority group, both firms can benefit through lower overall wage bills, echoing the argument in Darity and Williams (1985) on segregated markets. Intuitively, once one firm discriminates, the other has monopsony power over minority workers, allowing them to lower wages. Then, as wage setting exhibits strategic complements in his setting, the discriminating firm can also reduce wages it offers to the advantaged group. Hence, even though the advantaged group is better off than the marginalized group in relative terms, both because they have greater employment opportunity and because the discriminating firm has weaker monopsony power, the bulk of the material gains from this strategic discrimination flows to the employers, just as the material gains from the colonial construction of race flowed to the landowning elites.[16] This pattern is also not restricted to labor markets: Jones et al. (2025) study an

[16]Muller and Fabian (2021) and Siddique et al. (2023) provide evidence that employer discrimination is weaker in more competitive labor markets. Similarly, Boulware and Kuttner (2019) and Challe et al. (2024) show that hiring discrimination is counter-cyclical with less discrimination in tight labor markets, consistent with a counter-cyclical racial gap in unemployment (Hoynes et al., 2012; Cajner et al., 2017).

environment of segmented consumer markets where the same dynamic occurs. If some firms choose to discriminate, the remaining non-discriminatory firms enjoy market power over minority consumers, allowing them to raise prices. They show both theoretically and empirically that in contrast to the predictions of a model of consumer discrimination under perfect competition, advantaged-group consumers may face lower prices than the marginalized group if sufficiently many firms discriminate that the non-discriminatory firms have significant market power.

Finally, connecting back to divert-and-exploit strategies in competitive filters, the contest designer may also benefit from embedding discrimination into the contest, especially when the designer benefits from the aggregate effort employed. Kawamura and Moreno de Barreda (2014) and Drugov and Ryvkin (2017) demonstrate that even if the contestants are ex-ante identical when entering the contest, the designer can benefit from biasing the contest in favor of one competitor, increasing the total effort exerted by both participants. As noted by Chelwa et al. (2022), when the advantaged group is unable to secure their advantage through discrimination in the pre-market stage, the fallback is to simply discriminate in the market. Therefore, considering both pre-market and market interactions, we observe that the two strategies of divert-and-exploit and divide-and-conquer can occur alongside each other, yielding benefits for both employers (or other contest designers) and the advantaged group participants in these filters.

2.2 How Groups Act in Individuals

A second tenet of SE is that the processes of identity formation allows groups to 'act in individuals' (Darity et al., 2006; Darity, 2022). A fundamental challenge is that in a large society, actions which produce the greatest benefits for the group need not be the individually optimal choice, so each individual within a group has an incentive to free ride on the efforts of others. For example, within the advantaged group, strategic discrimination may yield aggregate benefits for them at the marginalized group's expense, but may also require individuals within the advantaged group to incur individual costs when discriminating. This conflict between individual and collective interests is exactly the logic

behind the claim that market competition will eliminate discrimination by employers or employees: even if discrimination is collectively beneficial, every agent has an incentive to undercut their discriminatory rivals, yielding a complete unraveling of any discrimination in the marketplace. However, as we observe that discrimination does persist, SE emphasizes that identity formation creates incentives for individuals to align their behavior with the group's collective interest.

These incentives are not limited to the advantaged group; Stewart (1995) argues that economists should examine the mechanisms through which there exists a 'natural solidarity' amongst Black people (and other groups with a shared experience of oppression). Just as individual members of the advantaged group have an incentive to free ride on the discriminatory efforts of the rest of their group, members of the marginalized group have an incentive to assimilate into the advantaged group, either by 'passing' (physical assimilation) or 'Tomming' (ideological assimilation) (Darity et al., 2017).[17] Yet, as these strategies of assimilation impose costs on those members of their group who either cannot or will not assimilate, marginalized groups want to create barriers to assimilation to ensure their group continues to cohere. In essence, we observe that people treat their racial identity as highly meaningful, whether Black or White, and their actions are shaped by the norms, beliefs, and material consequences associated with one identity or another.

For both the advantaged and marginalized groups, the processes of forming, maintaining, and transmitting group identity are endogenous to these incentives which help to resolve collective action problems within groups. Thus, theoretical approaches within SE can model this formation and transmission of racial identity, the incentives that enable racial groups to engage in effective collective action, and the maintenance of group boundaries. Importantly, although racial identities may make use of biological markers, such as skin color or hair texture, *race* is a socio-political category, not a biological one. As noted above, racial categories were not constructed identically across the Americas, so the same individual may become a different race merely by crossing a national border, demonstrating the inadequacy of biological theories to explain the nuances of race. In-

[17]See also Caselli and Coleman (2013) and Bhattacharya et al. (2015) who develop formal models of group conflict with the possibility of passing between the marginalized and advantaged groups.

stead, group identities are formed both by the socio-political action of dominant groups as in the racialization of slavery in the Americas, and by the collective social effort of racialized individuals (Stewart, 1995). We note that even though race was created for the purposes of division, exclusion, and exploitation, this does not preclude that collective identity has value to marginalized individuals and that those individuals will actively contribute to the maintenance and transmission of their racial identity. Thus, the formation of racial identity is an ongoing, iterative process with input from both the advantaged and the marginalized groups.

For the purposes of this chapter, we divide these incentives into two broad categories: material rewards and punishments that groups provide to influence individual behavior, and psychological influences that lead individuals to take actions which do not maximize their immediate material utility. We note that these material rewards are distinct from the material gains from strategic discrimination; here, the focus is on the rewards that a group can provide its own members to incentivize the group's desired behavior. In situations where strategic discrimination is individually rational in addition to being collectively beneficial, groups need not provide these additional rewards as self-interest suffices to achieve the desired end. Thus, we focus on the more interesting settings where there exist conflicts between self- and group-interest, so these incentives matter.

2.2.1 Material Incentives

In some situations, groups can provide material rewards or punishments to incentivize the desired behavior. The most overt forms are formal rules and barriers which bar marginalized groups from taking actions, joining institutions, or accessing resources that are permitted to the advantaged group. Similarly, in principle, advantaged groups could provide direct material rewards to their members who engage in discriminatory acts in the groups interest. Yet the bluntness of these tools are often their downfall as although they are not perfectly enforced, legal protections often exist which prevent barriers from being explicitly based on race and prevent groups from explicitly rewarding discrimination. However, there are three indirect channels through which groups provide material rewards or punishments to their members: discrimination as a social norm, identity

groups as clubs, and investments in identity capital.

The notion of discrimination as a social norm embeds group identities into an environment of community enforcement as in Kandori (1992). Common to these models is that behavior which violates the norm in one interaction leads to punishment from the broader community, even those who were not directly affected by the violation of the norm.[18] Peski and Szentes (2013) present a model where discrimination against the marginalized group can be sustained through this community enforcement. In their model, agents have both a physical color, representing their own group identity, and a social color, representing their reputation. Crucially, if an agent cooperates with another who does not share their physical color, there is a risk that their social color will change to match the outgroup member's color. Intuitively, if a social norm prohibits interactions across the color line, individuals who violate this norm risk having their reputations 'polluted' by their interactions. Then, the authors show that a stable social norm exists where individuals refuse interactions with those who do not share their physical color to avoid ostracism by their ingroup.

Building on this basic framework, Choy (2018) shows that a hierarchical social norm, where higher-ranked groups are prohibited from cooperating with lower-ranked groups, but not vice versa, is easier to sustain than an equilibrium where groups discriminate against one another symmetrically. As such, the paper provides an argument for why discriminatory social norms are likely to coincide with social stratification. Similarly, Dewan and Wolton (2022) identify norm-based discrimination that creates a divert-and-exploit pattern in labor markets. In their model, workers and employers match to engage in production. They show that a norm in which both advantaged group workers and employers discriminate against the marginalized group can be supported in equilibrium and provide conditions under which workers from the advantaged group gain from this discrimination, either through higher wages or higher employment probabilities. Thus, their model demonstrates how these norms can be used to support strategic discrimination in equilibrium, even as individual employers would prefer to deviate and hire the most-qualified candidate.

[18] Online Appendix A in Onuchic (2024) provides a closely related survey.

Also related is Harbaugh and To (2014) and Bramoulle and Goyal (2016), who focus on environments where the advantaged group can divert benefits, either by cheating marginalized group customers (Harbaugh and To, 2014) or by favoring ingroup members for jobs and contracts over more-qualified candidates from the outgroup (Bramoulle and Goyal, 2016). In both papers, discrimination diverts benefits from the outgroup towards one's ingroup at a cost to total welfare, consistent with the typical pattern under strategic discrimination. Furthermore, both papers study how the potential for cheating is disciplined by the threat of ostracism by some members of society. However, the groups' relative size matters, both numerical and in terms of their control of economic resources. Harbaugh and To (2014) shows that opportunistic discrimination against a marginalized could be prevented through solidarity with the majority group; however when that group is numerically small and/or economically powerless, this solidarity imposes an economic cost on the majority makes it less likely to persist. By contrast, in Bramoulle and Goyal (2016), favoritism is easier to sustain in numerically small groups, provided they have sufficient economic power as an internally homogenous group can more easily resist retaliation by outgroups.

Another channel through which groups can provide material incentives to their members draws on the theory of identity-based clubs as developed to study religious organizations.[19] In this theory, clubs produce goods for their members and exclude non-members from consumption of those goods. However, such 'identity-based organizations' face two challenges: identity formation imposes externalities—both positive and negative—on others and individuals within the club have an incentive to free ride on the productive activity of other club members. To resolve these challenges, these organizations play three roles in attracting desirable members while repelling the wrong sort: screening, substitution, and sorting.

As identified in two papers by Iannaccone (1992, 1994), religious organizations typically impose strict behavioral rules on their members, requiring them to make sacrifices to remain members in good standing. The first role of these sacrifices is to identify and

[19]See e.g., surveys by Carvalho (2016, 2019) on clubs as applied to religion. Davis (2019) and Chelwa et al. (2022) suggest that the theory of clubs may be fruitfully applied to the study of race.

screen out those who are not pre-disposed to contribute to the club's collective well-being, i.e., in the religious sense, those who are not 'true believers.' Thus, those who choose to remain in the club and make these sacrifices are those who are sufficiently committed to the club's mission to support production of the club's restricted-access goods. Applied to settings of racial groups, behavior that often invites social opprobrium by the broader society helps to screen those who will not contribute to the racial group's collective interest from those who will. Through this lens, we provide one reinterpretation of racial prejudice and racist jokes/comments. Rather than an arbitrary preference, we view racial prejudice as a degree of commitment to one's own racial identity.[20] Then, racist comments and other similar actions which can signal one's innate prejudice help the group to screen out those insufficiently committed to the group's interests, so the prejudiced do not divert benefits to the non-prejudiced.[21] We note that this screening is not limited to advantaged groups; marginalized groups also benefit from identifying those who are more committed to the group's collective interest, requiring group members to signal this commitment, even at personal economic cost to avoid exclusion from group-produced resources.

These behavioral rules also induce substitution away from activity outside the group and into within-group production. In the religious context, behavioral restrictions make it harder for religious-group members to participate in market activities outside the club, such as requirements to wear clothes that clearly mark one as a club member, spend time in an isolated location, or avoid eating and drinking certain foods or with certain people. As these prohibitions raise the cost of non-club activity, it induces substitution towards club participation. McBride (2015) provides a dynamic extension of this framework, showing that clubs can accept initial free-riding because as individuals invest more time and energy into the club, this escalating commitment makes it costlier to cease participating and return to non-club activities. Thus, a period of initial free-riding can create lock-in, after which the club can demand stricter adherance to behavioral rules. Addi-

[20]Kranton and Sanders (2017) and Kranton et al. (2020) find that ingroup bias is a stable personality trait, heterogeneous across individuals, such that someone who displays strong ingroup bias towards one of their ingroups tends to display strong ingroup bias towards all of their ingroups.

[21]A frequently-told story involves White people 'testing the waters' with a new White acquiantance or colleague with 'edgy' jokes. If the new colleague responds well, they can be brought more fully into the club, whereas if they respond poorly, the speaker has deniability through the comment being 'just a joke.'

tionally, Carvalho and Sacks (2021) consider the possibility of increasing returns to club participation, which induces a trade-off between club size and strictness. They show that discrimination against a club tends to make a club stricter and more cohesive, consistent with the strong sense of shared identity among Black Americans and other marginalized groups. Furthermore, stigmatizing club participation decreases the share of the population who actively join the club but also increases the extremism of those who remain club members.[22]

Finally, clubs provide a sorting role by helping individuals to match with like-minded individuals (Carvalho, 2020). In this role, club participation does not explicitly affect club good production; however, by demanding sacrifices from club members, clubs can identify those with strong ingroup preferences, who are more desirable friends, business partners, and mates to others with similarly strong ingroup preferences. For example, Stackman et al. (2016) find that Black students who choose to attend HBCUs tend to have stronger preferences to exclusively date other Black people, indicating that choice of college plays a sorting role for intra and interracial dating.

The third channel for material rewards is the accumulation of identity capital. As Roediger (1991) and Harris (1993) argue, first during chattel slavery and again during Jim Crow Segregation, 'Whiteness' was an advantaged status which protected one from legal restrictions and offered greater access to public and private resources McGee (2025a). Today, 'Whiteness as property' is best understood as a form of identity-group capital; analogously to human capital or social capital, individuals can make investments in their identity capital and although it cannot be alienated, these investments provide extra returns to their holders.[23] These returns are closely connected to the 'club' model discussed above; by investing in identity capital, one displays one's commitment to the identity group, to which others respond by directing rewards to the holders of that capital. Similar to the analysis of social capital by Glaeser et al. (2002), we observe that investment in

[22]For example, the broad stigmatization of overtly racist comments has coincided with both a decline in the overall number of individuals expressing such beliefs and a rise in extremist 'alt-right' groups holding explicitly White supremacist beliefs.

[23]In the 'bundle of rights' definition of property, we observe that although individuals lack a right of disposition over their Whiteness, they do possess rights to possession, enjoyment, and exclusion. Furthermore, historically one had to prove their Whiteness, but not Blackness, as only White status conferred social and economic benefits.

identity-group capital should exhibit a life-cycle pattern, first rising and then falling with age, should be higher for those who primarily interact with their own identity group and increasing in others' identity-capital investments due to complementarities, and that identity-capital investments should positively co-vary with human and social capital investments. Additionally, as in Iannaccone (1990), parent and child identity capital should be strongly positively correlated.

Building on this reduced-form approach, Darity et al. (2006) provide a model in which agents can either be 'individualists' who behave identically, irrespective of the physical color of others, or 'racialists' who are altruistic towards their ingroup and hostile to the outgroup. In line with the informal discussion in the authors' paper, we can think of the latter as those who invest in identity capital and the former as those who do not. They show that under a standard replicator dynamic, the stable equilibria either involve all agents converging to individualist behavior or all agents adopting racialist behavior, with an unstable mixed equilibrium in between. Yet, from Bisin and Verdier (2001), we know that cultural heterogeneity persists (stably) in equilibrium whenever socialization is endogenous. Thus, by drawing on the large literature on cultural transmission, we identify other possible features of identity capital.[24] First, as highlighted in this literature and discussed in Darity et al. (2006), other group members, such as parents, peers, and local leaders, intentionally influence the traits, beliefs, and behaviors transmitted within their group. However, as each of these agents have their own preferences and beliefs, the behaviors they see as important to cultivate will differ. Thus, an individual's specific identity investment will respond to their precise influence group. Second, identity investments depend on the structure of strategic interactions between individuals. Della Lena and Dindo (2024) show that when intergroup interactions primary exhibit strategic complements, cultural transmission generates assimilation and integration, whereas strategic substitutes promote marginalization. As such, we observe a potential feedback loop between identity capital and strategic discrimination: since divert and exploit strategies are more prevalent under strategic substitutes, they promote investment in identity capital,

[24] As this literature is extremely large, we only highlight a few important issues here and refer the reader to a pair of surveys by Bisin and Verdier (2023, 2025).

which in turn supports discrimination by the advantaged group. Finally, as noted by Bisin and Verdier (2024), cultural transmission influences the evolution of political institutions, especially the relative political power of different social groups. In particular, they observe that cultural heterogeneity is more likely when political power and culture are substitutes, in the sense that an increase in the political power of a group leads to an relative increase in identity-capital investment by opposing group. Hence, we see that social stratification tends to coincide with high investment in identity-group ties by marginalized groups to compensate for their relative socio-political weakness.

2.2.2 Psychological Incentives

As shown in the previous section, there are several channels through which groups direct material rewards to their members to create incentives for pro-group behavior. However, in many cases, the incentives identity groups provide are psychological, not material. Thus, in this section we highlight three channels through which psychological incentives affect behavior and can lead to discrimination, either through additional components to utility or through distortions to beliefs.

Darity (2022) argues that positional preferences must play a central role in our understanding of racial group behavior. In particular, when groups are closer to parity, people become more concerned about their group's status vis-a-vis other groups, rather than their own status within the group (Darity et al., 2017). Since status is naturally zero-sum—or even negative-sum when accounting for wasteful effort in status competitions—these preferences induce individuals to see the world as one where strategic discrimination is valuable because it reduces the opposing group's payoff while increasing one's own. Shayo (2020) demonstrates that these preferences for group status lead individuals to both display ingroup altruism and outgroup hostility (i.e., prejudice); however, he also shows that it creates an incentive for members of the low-status group to pass as high-status group members and that both groups wish to prevent such passing.[25] The notion that advantaged-group individuals benefit psychologically from a superior position in soci-

[25]Specifically, group permeability and ingroup altruism are substitutes as when individuals can change their group identity, their utility is less tied to their group's welfare.

ety reflects the idea from Du Bois (1935) of a 'psychological wage' of Whiteness, wherein even poor Whites understood themselves as the social betters of any Black American.

Per Postlewaite (2011), we can distinguish 'reduced-form' positional preferences, where rank is valued instrumentally, from 'deep' positional preferences, where individuals instrinically care about rank, separate from any material rewards it may offer. Cole et al. (1992, 2001) provide theoretical models in which some goods are not provisioned through a regular market; instead, these goods display a 'pecking order' phenomenon, where those with greater status and/or wealth are able to choose their desired good before those with lower status. As such, even though agents only care about their material utility, they act as though they have preferences for status given its influence in these non-market interactions. They also note the possibility of an alternate social order in which rank is unconnected from wealth and instead linked to parental social status, which they term an 'aristocratic' order. We observe that the different racial orders in Brazil and the United States roughly correspond to these two systems. In Brazil, mixed-race individuals can achieve high social status if they are university educated and wealth, such that these SES markers are better predictors of social race than one's skin color (Telles, 2006). By contrast, in the United States, race is seen as a pure product of ancestry, so even the idea of a Black person becoming White merely by going to university and making money sounds bizarre.[26] However, individuals may have intrinsic 'deep' preferences for status. As stated by Postlewaite (2011), winning simply feels good, even against opponents that one is unlikely to ever encounter again. These preferences may even have evolutionary foundations as an intrinsic desire to ascend the social hierarchy may have been useful in our collective past.

Another source of utility from identity in the notion that identity production is itself a good that individuals value (Stewart, 1995). He models racial identity as a commodity that groups produce collectively and argues that as identity production will have positive externalities for individuals with similar identity profiles and negative externalities

[26] As additional empirical support for reduced-form preferences, Ball et al. (2001) and Imas and Madarasz (2024) conduct experiments where status and exclusion affect demand and market prices, Charles et al. (2009) show that expenditure on visible goods (e.g., clothing, cars) responds to a demand for social signaling, and Lavetti et al. (2025) document that health status is improving in workplace status, controlling for income.

for those with dissimilar identities, groups will attempt to disrupt the identity production of others to preserve their own utility. Furthermore, he argues that stocks of cultural knowledge serve as the production technology for identity, which suggests a link to the cultural transmission literature, in the sense that the set of knowledge and practices that have been retained from the previous generation may limit which traits and behaviors can be acculturated in the new generation. This also relates to the notion of identity distance as discussed in Shayo (2020), who argues that individuals experience a disutility from perceiving themselves as dissimilar from their ingroup. As individuals can reduce perceived distance either by changing their own behavior to more closely match the group or by changing the group's behavior to more closely resemble their own, this desire for similar creates a drive for conformity, including punishing other group members for deviation from the group norm.[27]

Additionally, Benabou and Henkel (2025) notes that beliefs are an important component to identity: one's identity is a set of beliefs about their traits, values, goals, and social ties. This cognitive approach implies that individuals and groups will exert significant effort to shape the information available about both themselves and the outgroups in society. In the context of racial inequality, one major form of collective beliefs are 'legitimizing myths:' ideological narratives which either legitimize or oppose extant inequalities by providing explanations of how those inequalities came to be (McGee, 2025b). When racial inequalities are explained as being the result of marginalized groups' own failings, such as through a lack of ability or effort, those inequalities come to be seen as morally legitimate, so that not only is it acceptable to leave them in place, but that it would be unfair not to give these marginalized groups their 'just deserts.'[28] Darity (2005) and Chelwa et al. (2022) highlight that even theories of 'induced' inequality, such as those that rely on underinvestment in human or social capital by marginalized groups, serve an ideological purpose beyond their neutral scientific content as legitimizers of inequality. McGee (2025b) constructs a model with political action to support or oppose inequality along-

[27]However, we note that the scope for changing one's own perceived identity is constrained by the possibility of cognitive dissonance when one too strongly identifies with the outgroup, even if there are material rewards for doing so (Chelwa et al., 2022).

[28]Blumer (1958) notes that one element of prejudice is a sense of a 'proprietary claim' and deservingness over resources in society amongst advantaged groups.

side narrative competition over ideological explanations of this inequality. He shows that because ideological beliefs motivate political action, disagreement about the sources of inequality appears alongside political conflict over how government resources should be spent. However, beliefs are not perfectly polarized between advantaged and disadvantaged groups; when redistribution to correct inequality is sufficiently unlikely, even marginalized groups benefit from the psychological salve of seeing their position as deserved (Davis, 2024). This also points at an ideological role for model minorities: by downplaying differences in circumstances between different minority groups, such as immigrants who are typically positively selected from their home population compared to the descendents of enslaved people in the United States, they serve as further legitimization for the social system and to obscure that members of the advantaged group can often achieve wealth and social status without the ability or effort necessary (but often not sufficient) for minorities.

A final psychological channel to support strategic discrimination is the production of stereotypes. Following McGee (2025c), we define a stereotype as a generalization about a group of people, which could be positive, negative, or neutral.[29] These beliefs matter as they shape how individuals expect others to behave, particularly when the stereotypes apply to payoff-relevant traits, such as work ethic, intelligence, or leadership ability. For example, if an advantaged-group agent holds negative stereotypes about the marginalized group, they will expect them to perform worse. Then, if the two are competing, this negative stereotype encourages the advantaged-group member to compete more aggressively and crowd out the marginalized group, a form of stereotype lift (Walton and Cohen, 2003). Thus, from the perspective of the individual advantaged-group member, they engage in inaccurate statistical discrimination (Bohren et al., 2025), as they discriminate on the basis of these distorted beliefs. However, in response to this discrimination, the marginalized group optimally reduces their effort as they anticipate the greater aggression from the dominant group, creating a pattern of divert and exploit. As such, this framework reframes 'stereotype threat' Steele and Aronson (1995) as an equilibrium re-

[29]Hence, this definition is more restrictive than the definition in Bordalo et al. (2016) which applies to any generalization, including about non-human objects.

sponse to the perceived beliefs and anticipated behaviors of advantaged groups, and so as a product of strategic discrimination. Consistent with this framework, Huguet and Régner (2009) find that stereotype threat is present even when the individual thinks the stereotype is inaccurate, provided they are aware that others hold these beliefs, highlighting the importance of this equilibrium feedback. Furthermore, Alston et al. (2022) find no evidence of stereotype threat among Black students attending HBCUs, suggesting that these effects are only present when outgroup interactions are likely.[30]

Stereotypes can also be relevant when actions have positive spillovers, so the strategy of divert and exploit involves free riding on marginalized groups' effort. Here, positive stereotypes claim that the marginalized group is better suited to the task, justifying less effort from the advantaged group and boosting effort from the marginalized (Shih et al., 2002). For example, White slave-owners used beliefs that Black Americans were stronger and more physically resilient to pain to justify their exploitation of enslaved people's labor. Crucially, under both positive and negative spillovers, when stereotypes arise from strategic discrimination, they need not be accurate, nor even contain a 'kernel of truth' as in the 'cognitive schema' approach to stereotyping (Schneider, 2004; Bordalo et al., 2016). Thus, we interpret stereotypes as a product of collective motivated reasoning on the part of advantaged groups in pursuit of material gains from strategic discrimination. This model also sheds light on why certain group identities become salient. In purely informational accounts of stereotyping, smaller and more isolated groups should be the most stereotyped and that stereotypes should dissipate as information becomes more readily available. However, these groups offer smaller strategic benefits from stereotyping as the scope for strategic discrimination is smaller. Therefore, stereotypes can be pervasive and persistent about marginalized groups which are neither small, nor isolated from the advantaged group as these interacts provide an incentive to transmit and maintain stereotypes, even if they have the information to debias themselves.

[30]More broadly, a large literature finds advantaged groups compete more aggressively and marginalized groups less aggressively in between-group competition than in ingroup competitions. See e.g., Gneezy et al. (2003), Niederle and Vesterlund (2007), Booth and Yamamura (2018), and Siddique and Vlassopoulos (2020).

3 Stratification in the Macro

Theoretical work on SE also contributes to two fundamental questions at the macro level. First, what are the aggregate consequences of inter-group discrimination in terms of inequalities in income and wealth, and also for societal welfare? Second, why does discrimination tend to persist and what policy solutions are there to break this cycle?

In what follows, we present a simple overlapping generations model that builds on Brundage and Tavani (2024) and formalizes Arthur Lewis's ideas about competing vs. non-competing groups (Lewis, 1985), as well as more recent work on stratification economics (Darity, 2005; Darity et al., 2017; Chelwa et al., 2022).

3.1 Economic Environment

Consider a society populated by two groups, M (for marginalized) and D (for dominant). Individuals live two periods. In the pre-market period they invest h in marketable skills; these skills translate into earned income in the market period. We abstract from modeling firm behavior, and model individuals as investing in skill acquisition in the non-market phase of their life to then run their own 'small enterprises' that generate market income in the second period. Below, we will also consider the market-period choice of consumption and bequests to leave to future generations.

Individuals in the Marginalized Group: An individual $j \in \{1, \ldots, Q\}$ in group M chooses how much to invest in acquiring skills in the pre-market phase of their life in order to generate income (become competitive) in the market phase. Market income, denoted by y_j^M, is a function of j's effort in acquiring a marketable skill, $h_{j,M}$. In addition, the M-individual's market income can be negatively affected by the total discriminatory effort $d \in [0, 1]$ exerted by group D. However, there is also a degree of societal anti-discrimination enforcement that blunts the impact of d, denoted by $\varepsilon \in [0, 1]$. Accordingly, the function $y_j^M(h_{j,M}; d, \varepsilon)$ describes the skill-acquisition technology for an individual in group M as a function of her own investment, the discriminatory effort by the other group, and the extent of anti-discriminatory measures in society. We make the following

assumptions:

1. $y_j^M(0; d, \varepsilon) = 0$ (No-free lunch).

2. $\partial y_j^M / \partial h_{j,M} > 0$ (Productive investment); $\partial^2 y_{j,M} / \partial h_{j,M}^2 < 0$ (diminishing returns to skill acquisition).

3. $\partial y_j^M / \partial d < 0$ (Economic harm from discrimination).

4. $\partial y_j^M / \partial \varepsilon > 0$ (Effectiveness of anti-discrimination enforcement).

To sharpen our conclusions, we assume the following Cobb-Douglas specification:

$$y_j^M(h_{j,M}; d, \varepsilon) = A\, h_{j,M}^{\alpha}\, [1 - d(1-\varepsilon)]^{1-\alpha}, \quad \alpha \in (0,1),\ A > 0, \tag{1}$$

Notice that $\varepsilon = 1$, that is full anti-discrimination enforcement, completely neutralizes the effect of discrimination on the M-individual's income; but we will show below that there are strong economic reasons for which anti-discriminatory measures will never be fully enforced.[31] Individual j in group M begins their life with an endowment w_j^M of inherited wealth (more on this later). The total material resources available to an M-individual is therefore $w_j^M - h_{j,M} + y_j^M(h_{j,M}; d)$ and what is basically equivalent to a participation constraint requires that the income generated through investment in skill acquisition makes the person at least indifferent between investing or not. Thus, it must be true that

$$w_j^M - h_{j,M} + y_j^M(h_{j,M}, d) \geq w_j^M \tag{2}$$

which reduces to $y_j^M \geq h_{j,M}$. The pre-market problem is then

$$\max_{h_{j,M} \geq 0} \quad y_j^M(h_{j,M}; d, \varepsilon) - h_{j,M}.$$

Under the Cobb–Douglas specification above, we find the following reaction function $h_M(d; \varepsilon)$ and corresponding market income $y^M(d; \varepsilon)$, both symmetric across all $j \in M$

[31] We will also assume all necessary restrictions on the parameter A for the model to deliver economically meaningful results.

individuals :

$$h_M(d;\varepsilon) = \left(\alpha A\right)^{\frac{1}{1-\alpha}}\left[1-d(1-\varepsilon)\right], \qquad y^M(d;\varepsilon) = \alpha^{\frac{\alpha}{1-\alpha}} A^{\frac{1}{1-\alpha}}\left[1-d(1-\varepsilon)\right]. \tag{3}$$

Resources spent on skill acquisition, and therefore market income, are linearly decreasing in d –and increasing in ε– given the assumption on technology: more discriminatory efforts by the dominant group reduce skill investments by the marginalized group members, capturing Lewis's point about the ability of dominant groups to limit access by subordinate groups and ultimately make them non-competitive at the market stage. Importantly, the result also implies that discrimination can play a role in reducing educational attainment by marginalized groups. The reverse is true about the extent of anti-discriminatory measures ε, which increase investment in skill acquisition, and therefore market income for M-individuals.

Individuals in the Dominant Group – Each $i \in \{1,\ldots,N\}$ in group D is not harmed by discrimination, and earns

$$y_i^D(h_{i,D}) = A\, h_{i,D}^{\alpha}.$$

Discrimination is a group-specific public bad with free-riding: For each i, total d is $d = \eta d_i + (1-\eta)d_{-i}$ with $\eta \in (0,1)$ and d_{-i} the average discriminatory effort exerted by all the non-i individuals. Individual discriminatory effort has convex cost $c(d_i) = \frac{1}{2}d_i^2$ for concreteness, to capture the fact that more blatant discriminatory effort will be either societally or legally sanctioned, but more subtle discrimination may not be. A D-agent chooses $(h_{i,D}, d_i)$ so as to maximize their position relative to a M-individual, that is the difference between her market income and the average M-person market income:

$$\max_{h_{i,D}\geq 0,\, d_i \geq 0} \Big[y_i^D(h_{i,D}) - y^M(h_M, d_i; d_{-i}; \varepsilon)\Big] - h_{i,D} - c(d_i), \tag{4}$$

which yields symmetric choices and incomes across all D-individuals:

$$h_{i,D} = h_D = (\alpha A)^{\frac{1}{1-\alpha}}, \qquad d_i = d = \eta(1-\alpha) A\, h_M^{\alpha}\left[1-d(1-\varepsilon)\right]^{-\alpha} \tag{5}$$

3.2 Symmetric Equilibrium

A symmetric equilibrium consists of choices $\{h_{j,M}\}_{j=1}^{Q}$ that maximize the net material resources for all $j \in M$ individuals given d and ε, and choices $\{h_{i,D}, d_i\}_{i=1}^{N}$ that maximize the difference in net material resources between each $i \in D$ individual and the average M-individual, given d_{-i} and ε defined above. Symmetry also requires that $d_i = d$ for all i. Market income for dominant-group members is:

$$y^{E,D} = \alpha^{\frac{\alpha}{1-\alpha}} A^{\frac{1}{1-\alpha}}, \tag{6}$$

and equilibrium discrimination is found as:

$$d^E = \eta \left(\frac{1-\alpha}{\alpha} \right) (\alpha A)^{\frac{1}{1-\alpha}}, \tag{7}$$

independent of ε because D internalizes only the *relative* payoff difference.[32] Plugging d^E into M's outcomes, we find the (average) market income for a marginalized group member as:

$$y^{E,M}(\varepsilon) = \alpha^{\frac{\alpha}{1-\alpha}} A^{\frac{1}{1-\alpha}} \left[1 - (1-\varepsilon)\eta \left(\frac{1-\alpha}{\alpha} \right) (\alpha A)^{\frac{1}{1-\alpha}} \right], \tag{8}$$

so that equilibrium income inequality, defined as the ratio of equilibrium average incomes of the two groups, is:

$$\frac{y^{E,D}}{y^{E,M}(\varepsilon)} = \frac{1}{1 - (1-\varepsilon)\eta \left(\frac{1-\alpha}{\alpha} \right) (\alpha A)^{\frac{1}{1-\alpha}}} > 1. \tag{9}$$

Perfect equality is achieved only under full enforcement of anti-discriminatory measures $\varepsilon = 1$, which is unlikely if enforcement is costly, as shown just below. Thus, average incomes of the D group will exceed those of the M group, as per equation (9).

[32] Intuitively, M's enforcement raises h_M and y^M one-for-one via the effective term $[1 - d(1-\varepsilon)]$, so D's best reply preserves the same d that maintains the status differential.

3.3 Anti-Discrimination with Costly Enforcement

Why are anti-discriminatory measures not fully enforced? Assume that it is costly to do so, since the group (or individual) that is discriminated against has to incur the legal or bureaucratic costs of proving that there in fact was discrimination against themselves or their group's members. If M-individuals bear the cost (for example, the burden of the proof in legal cases) of ensuring enforcement, and assuming that such cost is strictly convex ($c(\varepsilon) = \frac{1}{2}\varepsilon^2$), a group M-agent will now choose (h_M, ε) to maximize:

$$Ah_M^{\alpha}[1 - d(1-\varepsilon)]^{1-\alpha} - h_M - c(\varepsilon),$$

implying the FOCs:

$$h_M \text{ as in (3)}, \qquad \varepsilon = (1-\alpha)Ah_M^{\alpha}[1 - d(1-\varepsilon)]^{-\alpha}d.$$

Combining with the D best-response (5) yields:

$$\varepsilon = \eta(1-\alpha)^2\,\alpha^{\frac{2\alpha}{1-\alpha}}A^{\frac{2}{1-\alpha}} \propto (d^E)^2, \tag{10}$$

hence $0 \leq \varepsilon < d^E < 1$ whenever enforcement is costly. Discrimination therefore persists even when M actively enforces their rights.

3.4 Welfare

Let $\mu \equiv Q/(Q+N) \in (0,1)$ be the population share of the M-group, and assume $\varepsilon = 0$ for ease of exposition. A benevolent planner maximizes the society's average net income, taking into account that all individuals belonging to one group make the same choices, that is chooses $\{h_M, h_D, d\,\varepsilon\}$ to maximize:

$$y = \mu\,[Ah_M^{\alpha}[1 - d(1-\varepsilon)]^{1-\alpha} - h_M - c(\varepsilon)] + (1-\mu)\Big[Ah_D^{\alpha} - h_D - c(d)\Big].$$

For any (h_M, h_D), societal income decreases with discrimination, $\partial y/\partial d < 0$, so the planner sets $d = 0$ and correspondingly $\varepsilon = 0$. Under this allocation, group outcomes are equalized:

$$h_M^* = h_D^* = (\alpha A)^{\frac{1}{1-\alpha}}, \qquad y^{*,M} = y^{*,D}.$$

Thus, purposeful discrimination is purely wasteful if the societal goal is to maximize average net material resources. Moreover, the allocation that maximizes the society's average net income is also egalitarian.

3.5 Wealth Dynamics

We now turn to the problem, faced by individuals during the market phase of their lives, to consume and leave bequests to their descendants. Assume, following Galor and Zeira (1993), a simple utility function defined over consumption and bequests: $u^r(c_r, b_r) = \beta \ln c_r + (1-\beta) \ln b_r, r = \{M, D\}$. Assume further that each parent has one child so that population remains constant. Then, consumption and bequest are constant shares of net lifetime income Y^r:

$$c_r = \beta Y^r, \quad b_r = (1-\beta) Y^r.$$

If bequests earn gross return $(1+\rho_t)$, next-generation wealth satisfies:

$$w_{t+1}^r = (1-\beta) Y_t^r (1+\rho_t).$$

Brundage and Tavani (2024) have shown that under the assumptions made in this model, and under $\varepsilon = 0$ for simplicity, the lifetime income gap is constant and equal to:

$$Y^D - Y^M = \frac{[\eta(\frac{1-\alpha}{\alpha})(\alpha A)^{\frac{1}{1-\alpha}}]^2}{2} \left[\frac{2}{\eta} - 1\right]$$

However, even with stationary income differences and $\varepsilon = 0$, the racial wealth gap obeys

$$w_{t+1}^D - w_{t+1}^M = (1-\beta)(Y^D - Y^M)(1+\rho_t), \tag{11}$$

which grows over time given the rate of return to wealth ρ_t. The wealth gap will grow even if the gap in incomes are constant.

3.6 Political Economy

We now depart from the assumption of a planner maximizing net market incomes, and instead analyze the social choice of discriminatory effort under two social welfare aggregators. In both cases, the planner internalizes the symmetric within-group choices.

3.6.1 Bernoulli–Nash Aggregation

Suppose that the planner maximizes a *geometric* average of the two groups' objective functions. Even though it is not possible to find a closed-form solution for the planner's choice, we can still show that this problem will lead to an interior value for the amount of discrimination d. Under $\varepsilon = 0$ and taking logs to simplify the problem, the planner chooses $h_M \geq 0, h_D \geq 0, d \geq 0$ to maximize

$$\mu \ln \left[A h_M^{\alpha} (1-d)^{1-\alpha} - h_M \right] + (1-\mu) \ln \left[A h_D^{\alpha} - A h_M^{\alpha} (1-d)^{1-\alpha} - h_D - \frac{d^2}{2} \right] \tag{12}$$

The choices of educational investment for the dominant group is once again given by equation (5) above. The first-order condition on the choice of investment in skill acquisition for the M-group is:

$$\frac{\mu \left(\alpha A h_M^{\alpha-1} (1-d)^{1-\alpha} - 1 \right)}{A h_M^{\alpha} (1-d)^{1-\alpha} - h_M} = \frac{(1-\mu) \alpha A h_M^{\alpha-1} (1-d)^{1-\alpha}}{A h_D^{\alpha} - A h_M^{\alpha} (1-d)^{1-\alpha} - h_D - \frac{d^2}{2}}$$

Intuitively, in the decentralized problem presented in the text only the numerator in the LHS matters. The presence of an additional term in the RHS here implies that the planner will allocate *fewer* resources to skill investment for the marginalized group members, because of the adverse effect that such investment has on the dominant group's objective function. The first-order condition on the choice of discriminatory effort is:

$$\frac{\mu (1-\alpha) A h_M^{\alpha} (1-d)^{-\alpha}}{A h_M^{\alpha} (1-d)^{1-\alpha} - h_M} = \frac{(1-\mu) \left[(1-\alpha) A h_M^{\alpha} (1-d)^{-\alpha} - d \right]}{A h_D^{\alpha} - A h_M^{\alpha} (1-d)^{1-\alpha} - h_D - \frac{d^2}{2}}$$

In the decentralized problem with no free-riding ($\eta = 1$), only the numerator term in the RHS would matter: the presence of a positive term on the LHS of the equation implies a *lower* amount of discrimination relative to the corresponding decentralized choice. And yet, given the problem under consideration, we can show that discrimination will take a positive value in this case. The MRS between h_M and d is:

$$\frac{\alpha A h_M^{\alpha-1}(1-d)^{1-\alpha} - 1}{(1-\alpha)Ah_M^{\alpha}(1-d)^{-\alpha}} = \frac{\alpha A h_M^{\alpha-1}(1-d)^{1-\alpha}}{(1-\alpha)Ah_M^{\alpha}(1-d)^{-\alpha} - d}$$

Cross-multiplying, we obtain:

$$(\alpha A h_M^{\alpha-1}(1-d)^{1-\alpha} - 1)[(1-\alpha)Ah_M^{\alpha}(1-d)^{-\alpha} - d] \quad = (1-\alpha)\alpha A h_M^{2\alpha-1}(1-d)^{1-2\alpha}$$

Expanding the terms on the LHS, factoring and simplifying, we can find d from the solution of:

$$d = \frac{(1-\alpha)Ah_M^{\alpha}(1-d)^{-\alpha}}{1 - \alpha A h_M^{\alpha-1}(1-d)^{1-\alpha}}$$

which, under the restrictions required on the productivity parameter A, delivers a positive value for the planner's choice of discriminatory effort. Once again, the planner takes into account the benefit of discrimination for the dominant group, and therefore will select to exert some discriminatory effort.

3.6.2 Linear Aggregation

If we impose a linear aggregator, the choice of discriminatory effort by the planner can lead to even more extreme results, depending on the weight of the marginalized group, only two outcomes can emerge: either the planner will allocate no resources toward skill investment for the marginalized group—and therefore there will be no need for discriminatory effort given that M-group members will be non-competitive already—or M-group individuals will end up earning higher incomes than D-group individuals. Assume again that the non-benevolent planner internalizes the symmetric choices made by individuals in the same group, and that the weight of the marginalized group in the planner's linear objective function is given by μ; the planning authority will choose $h_M \geq 0, h_D \geq 0, d \geq 0$

to maximize:

$$\mu \left[A h_M^{\alpha} (1-d)^{1-\alpha} - h_M \right] + (1-\mu) \left[A h_D^{\alpha} - A h_M^{\alpha} (1-d)^{1-\alpha} - h_D - \frac{d^2}{2} \right] \tag{13}$$

Notice that the income of the marginalized group enters both positively, with wight μ, but also *negatively*, with weight $1-\mu$, in the planner's objective function: differently from the problem presented in the body of the paper, the planner does not consider discrimination as completely wasteful, because it now values its benefit for the dominant group.

As is usual with linear welfare aggregators, we will show that only corner solutions are possible in this problem: either maximal inequality or perfect equality between the two groups, depending on the weight of the marginalized group μ. The choice of skill investment for the dominant group is given by equation 5 above. The choice of educational investment for the marginalized group members is:

$$h_M = \max \left\{ 0, \left[\frac{2\mu - 1}{\mu} (\alpha A) \right]^{\frac{1}{1-\alpha}} (1-d) \right\} \tag{14}$$

which is positive only if the weight of the M-group welfare in the planner's objective function is greater than one-half. If $\mu \in [0, 1/2]$, the planner will choose to allocate zero resources to skill investment for the marginalized group: it only chooses to maximize the welfare of the dominant group. The corresponding allocation involves zero income for the M-group members.

The first-order condition on the choice of d is:

$$(1-2\mu)(1-\alpha) A h_M^{\alpha} (1-d)^{-\alpha} = (1-\mu) d$$

which, upon substition of h_M from equation (14), yields:

$$d = \max \left\{ \left(\frac{1-2\mu}{1-\mu} \right) \left(\frac{2\mu - 1}{\mu} \right)^{\frac{\alpha}{1-\alpha}} \left(\frac{1-\alpha}{\alpha} \right) (\alpha A)^{\frac{1}{1-\alpha}}, 0 \right\} \tag{15}$$

Notice however that only two cases are possible: either $\mu \in [0, 1/2]$ or $\mu \in (1/2, 1]$. We will show now that in both cases, discriminatory effort will be zero, but for very different

reasons.

- In the first case, we already know that $h_M = 0$: the planner only cares about the dominant group, and allocates zero resources to skill acquisition for marginalized group members. Therefore, it does not need to allocate any resources toward discrimination, given that only dominant group members will earn income in the market stage of their lives. Inequality is the highest possible given that M-group members earn no income in the market stage: $y^D/y^M = \infty$.

- In the second case, h_M is positive. Given that discriminatory effort cannot take a negative value, it must be zero: the planner will choose to allocate no resources to discrimination. Because the choice of skill investment by D-group members is still given by (5), we obtain that D-group members will earn a higher income that M-group members: in fact,

$$y^D/y^M = \left(\frac{\mu}{2\mu - 1}\right)^{\frac{\alpha}{1-\alpha}}$$

  which is larger than one provided that $1/2 < \mu < 1$, but tends to one (perfect equality) as μ tends to one.

Takeaway. Under a geometric social aggregator, discrimination persists at an interior $d > 0$ because suppression of M directly raises the planner's objective through D's term. Under linear aggregation, political under-representation of M ($\mu \leq \frac{1}{2}$) produces the starkest form of inequality without requiring active discrimination; when $\mu > \frac{1}{2}$ the planner chooses $d = 0$ but still leaves M behind, unless μ is very large.

3.7 Systemic Discrimination

As we have shown, direct discrimination has sizable consequences for the distribution of income and wealth in society. This inter-generational wealth inequality is important as it is a central channel through which individuals and groups transmit advantage and disadvantage across generations. For example, access to wealth in the presence of credit con-

straints allows one to absorb financial shocks and enables investment decisions that those without wealth cannot make. Yet, inequality is not limited to environments in which agents are subject to discrimination. As an economy is interconnected, so decisions in one area affect agents in other areas, agents can be negatively (or positively) affected by discriminatory spillovers. Thus, in this section, we discuss these broader channels through which discrimination becomes persistent.

Bohren et al. (2025) formalizes a distinction between 'direct' discrimination, which they define as a decision-maker conditioning their choice on another agent's group identity, controlling for other decision-relevant information, and 'systemic' discrimination. For example, a judge who makes different decisions for otherwise-identical defendants when one is Black and another is White is engaging in direct discrimination. By contrast, systemic discrimination occurs when discrimination elsewhere in the economy means that ex-ante identical individuals have become interim unequal by the time the decision in question is being made. Thus, even if that specific decision is (directly) race-neutral, one racial group will systematically perform better at that decision node. For example, if police tend to arrest and charge Black people for behavior that they do not arrest or charge White people, then on average when a Black defendant faces a judge, they are likely to have a longer criminal record than a White defendant with the same underlying activity. As such, even if the judge's decision is race-neutral, if it conditions on the past criminal record, it will produce disparate outcomes between Black and White defendants not based on an underlying difference in criminal behavior. This definition implies that systemic discrimination is not discrimination that is widespread or embedded into formal policies or institutions, but a process by which an inter-connected economic system amplifies, rather than dampens, direct discrimination (McMillon, 2025). In particular, this implies two analytically distinct (but not mutually exclusive) ways that a social environment could have persistent inequality: ongoing direct discrimination and systemic discrimination.

In their paper, Bohren et al. (2025) discuss two potential channels for systemic discrimination: discrimination which creates inequality in one's traits or productivity at the focal node (technological inequality) and discrimination which creates inequality in one's

signals given their underlying traits and productivity (informational inequality). For example, if a Black student attends a systematically under-resourced school or faces teachers who directly discriminate against him, so that given the same underlying intelligence and after making the same effort at their studies as a White student, he develops less academic ability than that White student, this Black student faces technological systemic discrimination.[33] Similarly, if given equal academic ability and performance, the Black student is evaluated more harshly and so receives worse grades than the White student, he faces informational systemic discrimination. Additionally, informational systemic discrimination can arise if the evaluation criteria are chosen to weight traits more common in the advantaged group, creating a wedge between the individual's true productivity and their evaluated productivity under these criteria. For example, if SAT questions rely in part on White cultural knowledge which does not indicate academic ability or preparedness for college, test-takers face systemic discrimination.

Beyond their formalization of systemic discrimination, they provide an experimental tool for identifying direct, systemic, and total discrimination using an iterated audit design. In this design, decision-makers are presented with three types of 'candidates' (e.g., potential employees, defendants) in paired decisions: an advantaged-group candidate, a marginalized-group candidate, and a counterfactual candidate. The first two are presented along with signals typical for their group, such as a recommendation letter written for the advantaged group versus a recommendation letter for the marginalized group, or the criminal record of an advantaged-group defendant versus the criminal record of a marginalized-group defendant. Thus, paired decisions between the advantaged and marginalized candidates should capture the total disparities arising between the two groups. However, the counterfactual candidate has a marginalized-group identity but the signals typical of the advantaged group. Hence, the paired decision between the advantaged and counterfactual candidates replicates the standard audit study design and identifies direct discrimination. Finally, the paired decision between the counterfactual and marginalized candidate identifies any systemic discrimination: inequalities

[33]We note a close connection to the idea of advantaged groups rendering marginalized groups non-competing in the pre-market stage as per Lewis (1985).

arising from differences in traits and signals outside the focal decision.[34]

Finally, McMillon (2025) provides a taxonomy of four potential amplification mechanisms which could contribute to systemic discrimination more broadly. The first two channels are analogous to systemic discrimination as described by Bohren et al. (2025). First, he highlights inter-sectoral spillovers, where later decisions condition on earlier outcomes, such as the hiring manager conditioning on grades from a potentially discriminatory teacher. Second, there can be reinforcement processes, such as learning-by-doing, which can be thought of as 'self-spillovers' within a firm, task, or individual. Here, the earlier and later decision nodes are within the same sector, firm, or task, but similarly, discrimination at the earlier node creates inequality in either productivity or signals of productivity at the later node. Third, inequality can amplify the impact of shocks on an individual. As noted above, wealth is important as it allows individuals to absorb financial shocks. Thus, if two agents both lose employment but one has a wealth cushion and the other does not, the latter may be forced to accumulate debt at high interest rates or even be forced out of their home if they cannot afford rent or mortgage payments. Conversely, when an opportunity appears, such as an idea for an entrepreneurial venture (a positive shock), the wealthy individual might be able to invest in their idea and voluntarily leave their job to pursue it full-time, while the wealth-poor individual could not.[35] Finally, social multipliers can arise when individuals interact whenever choices exhibit strategic complementarities or substitutability (Burke, 2017). For example, if the level of academic ability developed by students depends not only on their own innate intelligence, effort, and teacher quality, but on the quality of their peers, then the effect of discrimination on one student spills over to their peers. Thus, to the extent that students are disproportionately likely to have same-race peers, this magnifies any initial consequences of discrimination on the students.[36]

[34] As the authors note, this decomposition is in the spirit of Kitagawa (1955), Oaxaca (1973), and Blinder (1973) which decompose wage differentials in terms of the difference in traits between two groups and the difference in the returns to those traits for the groups.

[35] Krueger et al. (2016) and Violante (2022) provide surveys of how household heterogeneity amplifies macroeconomic shocks; some recent contributions include Bilbiie (2025), Bilbiie et al. (2023, 2025), and Bayer et al. (2024).

[36] Classic contributions to the theory of social spillovers include Loury (1977), Durlauf (1996), and Glaeser et al. (2003); Boucher et al. (2024) is a recent contribution; Durlauf and Ioannides (2010) and Jackson

4 Discussion and Conclusion

As we show in this survey, formal theory has much to contribute to Stratification Economics. At the micro level, explicit microfounding of the processes of intergroup discrimination and group cohesion help us understand the mechanics by which groups act. At the macro level, formal models illuminate the interconnections and feedback channels within the economy and provide the structure necessary for quantitative estimation and calibration. As such, we see significant scope for productive dialogue between researchers active in the SE tradition, who can adopt the insights and methods of formal theorists, and formal theorists, who will find many interesting questions to study on racial stratification.

Although an exhaustive list of open questions is beyond the scope of this paper, we conclude with some suggestions for future work. First, in contrast to the more extensive work on how groups resolve collective action problems, there has been relatively less effort studying forms of strategic discrimination, particularly in environments where strategic discrimination involves divert and exploit strategies. As there are many environments in which strategic inter-dependencies imply that distorting other agents' actions can create benefits for the advantaged group, there are many possible applications of this concept. Second, there is great scope to apply insights about group conflict to macroeconomic questions. For example, in a dynamic market where individuals can either produce and consume in closed clubs or in the open market, when will groups choose to 'crowd out' market production by diverting investment into closed clubs? As a broader example, could we formalize a notion of 'opportunity hoarding' (Tilly, 1999) and capture this mechanism of inequality transmission in our macroeconomic models? Third, can we construct richer models of labor market competition where productivity may be difficult to define for worker or employer? This could allow us to capture environments where qualifications and 'merit' may be endogenous to the choice of technology, enabling us to model situations where advantaged groups can define 'merit' in their own image to maintain advantage, despite facially neutral selection processes. Fourth, theoretical work has been limited in modeling the agency of marginalized group members in re-

(2020) provide surveys of the literature.

sponse to discrimination. Does discrimination motivate individuals to work harder than dominant group members, a phenomenon often referred to as John 'Henry-ism' (James, 1994)? Alternatively, does it lead to feelings of discouragement or internalized disparagement when facing structural obstacles? Or is it a combination of both? These contrasting behavioral responses may have significant implications for health outcomes and other socioeconomic indicators. Finally, much of the theoretical work on the four key areas we have highlighted has primarily taken place in isolation from each other. Thus, theoretical work can also contribute to identify how these forces interact: how do groups resolve collective action problems while engaging in strategic discrimination? Can we build proper microfoundations for discriminatory behavior into macroeconomic models? Can we identify channels for systemic discrimination alongside these other forces? As these many questions suggest, further application of formal theory to Stratification Economics is very likely to continue to be fruitful.

References

Alston, M., W. A. Darity, C. Eckel, L. McNeil, and R. Sharpe (2022). The effect of stereotypes on Black college test scores at a historically Black university. *Journal of Economic Behavior & Organization 194*, 408–424.

Arnesen, E. (1994). "Like Banquo's ghost, it will not down": The Race Question and the American Railroad Brotherhoods, 1880-1920. *American Historical Review 99*(5), 1601–1633.

Arrow, K. J. (1973). The theory of discrimination. In O. Ashenfelter (Ed.), *Discrimination in Labor Markets*, Chapter 1, pp. 3–33. Princeton: Princeton University Press.

Ashraf, N., O. Bandiera, V. Minni, and V. Quintas-Martinez (2023). Gender roles and the misallocation of labor across countries. *Mimeo*.

Ball, S., C. Eckel, P. J. Grossman, and W. Zame (2001). Status in markets. *Quarterly Journal of Economics 116*(1), 161–188.

Bayer, C., B. Born, and R. Luetticke (2024). Shocks, frictions, and inequality in US business cycles. *American Economic Review 114*(5), 1211–1247.

Benabou, R. and L. Henkel (2025). Identity as self-image. *Mimeo*.

Bhattacharya, S., J. Deb, and T. Kundu (2015). Mobility and conflict. *American Economic Journal: Microeconomics 7*(1), 281–319.

Bhattacharya, S. and A. Mukherjee (2023). Identity, economic mobility, and conflict. *Mimeo*.

Bilbiie, F. O. (2025). Monetary policy and heterogeneity: An analytical framework. *Review of Economic Studies 92*(4), 2398–2436.

Bilbiie, F. O., S. M. Galaasen, R. S. Gurkaynak, M. Maehlum, and K. Molnar (2025). HANKSSON. *Mimeo*.

Bilbiie, F. O., G. Primiceri, and A. Tambalotti (2023). Inequality and business cycles. *Mimeo*.

Bisin, A. and T. Verdier (2001). The economics of cultural transmission and the dynamics of preferences. *Journal of Economic Theory 97*(2), 298–319.

Bisin, A. and T. Verdier (2023). Advances in the economic theory of cultural transmission. *Annual Review of Economics 15*, 63–89.

Bisin, A. and T. Verdier (2024). On the joint evolution of culture and political institutions: Elites and civil society. *Journal of Political Economy 132*(5), 1485–1564.

Bisin, A. and T. Verdier (2025). Economic models of cultural transmission. *Mimeo*.

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources 8*(4), 436–455.

Blumer, H. (1958). Race prejudice as a sense of group position. *Pacific Sociological Review 1*(1), 3–7.

Bohren, J. A., K. Haggag, A. Imas, and D. G. Pope (2025). Inaccurate statistical discrimination: An identification problem. *Review of Economics and Statistics 107*(3), 605–620.

Bohren, J. A., P. Hull, and A. Imas (2025). Systemic discrimination: Theory and measurement. *Quarterly Journal of Economics 140*(3), 1743–1799.

Booth, A. and E. Yamamura (2018). Performance in mixed-sex and single-sex competitions: What can we learn from speedboat races in Japan. *Review of Economics and Statistics 100*(4), 581–593.

Bordalo, P., K. Coffman, N. Gennaioli, and A. Shleifer (2016). Stereotypes. *Quarterly Journal of Economics 131*(4), 1753–1794.

Boucher, V., M. Rendall, P. Ushchev, and Y. Zenou (2024). Toward a general theory of peer effects. *Econometrica 92*(2), 543–565.

Boulware, K. D. and K. N. Kuttner (2019). Labor market conditions and discrimination: Is there a link? *AEA Papers and Proceedings 109*, 166–170.

Bramoulle, Y. and S. Goyal (2016). Favoritism. *Journal of Development Economics 122*, 16–27.

Brundage, B. and D. Tavani (2024). Group conflict, racial inequality, and stratification. *Review of Black Political Economy*.

Buchanan, J. M. and G. Tullock (1962). *The Calculus of Consent*. Ann Arbor: University of Michigan Press.

Burke, M. A. (2017). Social multipliers. In S. N. Durlauf and L. E. Blume (Eds.), *The New Palgrave Dictionary of Economics*. London: Palgrave Macmillan.

Cajner, T., T. Radler, D. Ratner, and I. Vidangos (2017). Racial gaps in labor market outcomes in the last four decades and over the business cycle. *Federal Reserve Finance and Economics Discussion Series*.

Carvalho, J.-P. (2016). Identity-based organizations. *American Economic Review 106*(5), 410–414.

Carvalho, J.-P. (2019). Religious clubs: The strategic role of religious identity. In J.-P. Carvalho, S. Iyer, and J. Rubin (Eds.), *Advances in the Economics of Religion*, Chapter 2, pp. 25–41. Berlin: Springer Nature.

Carvalho, J.-P. (2020). Sacrifice and sorting in clubs. *Forum for Social Economics 49*(4), 357–369.

Carvalho, J.-P. and M. Sacks (2021). The economics of religious communities. *Journal of Public Economics 201*.

Caselli, F. and W. J. Coleman (2013). On the theory of ethnic conflict. *Journal of the European Economic Association 11*(1), 161–192.

Challe, L., Y. L'Horty, P. Petit, and F.-C. Wolff (2024). Cyclical behavior of hiring discrimination: Evidence from repeated experiments in France. *Annals of Regional Science 72*(3), 711–733.

Charles, K. K., E. Hurst, and N. Roussanov (2009). Conspicuous consumption and race. *Quarterly Journal of Economics 124*(2), 425–467.

Chelwa, G., D. Hamilton, and J. Stewart (2022). Stratification economics: Core constructs and policy implications. *Journal of Economic Literature 60*(2), 377–399.

Choy, J. P. (2018). Social division with endogenous hierarchy. *The Economic Journal 128*(615), 2711–2742.

Choy, J. P. (2024). A theory of discriminatory institutions, with applications to apartheid and to the political economy of migration. *Mimeo*.

Coate, S. and G. C. Loury (1993). Will affirmative-action policies eliminate negative stereotypes. *American Economic Review 83*(5), 1220–1240.

Cole, H. L., G. J. Mailath, and A. Postlewaite (1992). Social norms, savings behavior, and growth. *Journal of Political Economy 100*(6), 1092–1125.

Cole, H. L., G. J. Mailath, and A. Postlewaite (2001). Investment and concern for relative position. In T. Ichiishi and T. Marschak (Eds.), *Markets, Games, and Organizations*, Chapter 7, pp. 115–135. New York: Springer.

Cox, O. C. (1948). *Caste, Class, and Race: A Study in Social Dynamics*. New York: Monthly Review Press.

Darity, W. A. (2001). The functionality of market-based discrimination. *International Journal of Social Economics 28*(10), 980–986.

Darity, W. A. (2005). Stratification economics: The role of intergroup inequality. *Journal of Economics and Finance 29*(2), 144–153.

Darity, W. A. (2022). Position and possessions: Stratification Economics and intergroup inequality. *Journal of Economic Literature 60*(2), 400–426.

Darity, W. A., D. Hamilton, P. L. Mason, G. N. Price, A. Davila, M. T. Mora, and S. K. Stockly (2017). Stratification Economics: A general theory of intergroup inequality.

In A. Flynn, D. T. Warren, F. J. Wong, and S. R. Holmberg (Eds.), *The Hidden Rules of Race: Barriers to an Inclusive Economy*, Chapter 2, pp. 35–51. Cambridge: Cambridge University Press.

Darity, W. A., D. Hamilton, and J. B. Stewart (2015). A tour de force in understanding intergroup inequality: An introduction to Stratification Economics. *Review of Black Political Economy 42*(1), 1–6.

Darity, W. A., P. Mason, and J. B. Stewart (2006). The economics of identity: The origin and persistence of racial identity norms. *Journal of Economic Behavior & Organization 60*(3), 283–305.

Darity, W. A. and R. M. Williams (1985). Peddlers forever?: Culture, competition, and discrimination. *American Economic Review 75*(2), 256–261.

Davis, J. B. (2019). Stratification Economics as an economics of exclusion. *Journal of Economics, Race, and Policy 2*(3), 163–172.

Davis, J. B. (2024). Stratification Economics: Historical origins and theoretical foundations. *Oxford Review of Economic Policy 40*(3), 475–485.

Debortoli, D. and J. Gali (2025). Heterogeneity and aggregate fluctuations: Insights from TANK models. *NBER Macroeconomics Annual 39*(1), 307–357.

Della Lena, S. and P. Dindo (2024). An economic model of acculturation under strategic complements and substitutes. *Mimeo*.

Dewan, T. and S. Wolton (2022). A political economy of social discrimination. *Mimeo*.

Drugov, M. and D. Ryvkin (2017). Biased contests for symmetric players. *Games and Economic Behavior 103*, 116–144.

Du Bois, W. E. B. (1935). *Black Reconstruction in America, 1860-1880*. New York: The Free Press.

Durlauf, S. N. (1996). A theory of persistent income inequality. *Journal of Economic Growth 1*(1), 75–93.

Durlauf, S. N. and Y. M. Ioannides (2010). Social interactions. *Annual Review of Economics 2*, 451–478.

Eeckhout, J. (2006). Minorities and endogenous segregation. *Review of Economic Studies 73*(1), 31–53.

Esteban, J. and D. Ray (2008). On the salience of ethnic conflict. *American Economic Review 98*(5), 2185–2202.

Ferejohn, J. (1974). *Pork Barrel Politics*. Palo Alto: Stanford University Press.

Ferejohn, J. (1986). Incumbent performance and electoral control. *Public Choice 50*(1), 5–25.

Fields, B. J. (1990). Slavery, race and ideology in the United States of America. *New Left Review 181*(3), 95–118.

Fields, B. J. and K. E. Fields (2022). *Racecraft: The Soul of Inequality in American Life*. New York: Penguin Random House.

Frohlich, N. (1975). The instability of minimum winning coalitions. *American Political Science Review 69*(3), 943–946.

Galor, O. and J. Zeira (1993). Income distribution and macroeconomics. *Review of Economic Studies 60*(1), 35–52.

Glaeser, E. L., D. Laibson, and B. Sacerdote (2002). An economic approach to social capital. *The Economic Journal 112*(483), 437–458.

Glaeser, E. L., B. I. Sacerdote, and J. A. Scheinkman (2003). The social multiplier. *Journal of the European Economic Association 1*(2), 345–353.

Gneezy, U., M. Niederle, and A. Rustichini (2003). Performance in competitive environments: gender differences. *Quarterly Journal of Economics 118*(3), 1049–1074.

Harbaugh, R. and T. To (2014). Opportunistic discrimination. *European Economic Review 66*(3), 192–204.

Harris, C. I. (1993). Whiteness as property. *Harvard Law Review 106*(8), 1707–1793.

Hoynes, H., D. L. Miller, and J. Schaller (2012). Who suffers during recessions? *Journal of Economic Perspectives 26*(3), 27–48.

Huguet, P. and I. Régner (2009). Counter-stereotypic beliefs in math do not protect school girls from stereotype threat. *Journal of Experimental Social Psychology 45*(4), 1024–1027.

Iannaccone, L. R. (1990). Religious practice: A human capital approach. *Journal for the Scientific Study of Religion 29*(3), 297–314.

Iannaccone, L. R. (1992). Sacrifice and stigma: Reducing free-riding in cults, communes, and other collectives. *Journal of Political Economy 100*(2), 271–291.

Iannaccone, L. R. (1994). Why strict churches are strong. *American Journal of Sociology 99*(5), 1180–1211.

Imas, A. and K. Madarasz (2024). Superiority-seeking and the preference for exclusion. *Review of Economic Studies 91*(4), 2347–2386.

Jackson, M. O. (2020). A typology of social capital and associated network measures. *Social Choice and Welfare 54*(2), 311–336.

James, S. A. (1994). John Henryism and the health of African-Americans. *Culture, Medicine, and Psychiatry 18*, 163–182.

Jones, M. E. C., T. D. Logan, D. Rose, and L. D. Cook (2025). Competition and consumer discrimination. *Mimeo*.

Kandori, M. (1992). Social norms and community enforcement. *Review of Economic Studies 59*(1), 63–80.

Kawamura, K. and I. Moreno de Barreda (2014). Biasing selection contests with ex-ante identical agents. *Economics Letters 123*(2), 240–243.

Kitagawa, E. M. (1955). Components of a difference between two rates. *Journal of the American Statistical Association 50*(272), 1168–1194.

Kranton, R. E., M. Pease, S. G. Sanders, and S. Huettel (2020). Deconstructing bias in social preferences reveals groupy and not-groupy behavior. *Proceedings of the National Academy of the Sciences 117*(35), 21185–21193.

Kranton, R. E. and S. G. Sanders (2017). Groupy versus non-groupy social preferences: Personality, region, and political party. *American Economic Review 107*(5), 65–69.

Krueger, D., K. Mitman, and F. Perri (2016). Macroeconomics and household heterogeneity. In J. B. Taylor and H. Uhlig (Eds.), *Handbook of Macroeconomics*, Chapter 11, pp. 843–921. Amsterdam: Elsevier.

Lagerlof, J. N. M. (2020). Strategic gains from discrimination. *European Economic Review 122*.

Lavetti, K. J., L. Hong, J. A. Holmes, and T. D. Logan (2025). Workplace stratification and racial health disparities. *Mimeo*.

Lefebvre, S. (2025). A review essay of The Economics of Structural Racism by Patrick Mason. *Journal of Economic Literature 63*(3), 1011–1037.

Lewis, W. A. (1985). *Racial Conflict and Economic Development*. Cambridge: Harvard University Press.

Logan, T. D. (2022). American enslavement and the rediscovery of Black economic history. *Journal of Economic Perspectives 36*(2), 81–98.

Lopez, I. F. H. (1994). The social construction of race: Some observations on illusion, fabrication, and choice. *Harvard Civil Rights-Civil Liberties Law Review 29*, 1–62.

Loury, G. C. (1977). A dynamic theory of racial income differences. In P. A. Wallace and A. A. LaMond (Eds.), *Women, Minorities, and Employment Discrimination*, pp. 153–186. Lexington: Lexington Books.

Mason, P. L. (2023). *The Economics of Structural Racism*. Cambridge: Cambridge University Press.

Mason, P. L., J. B. Stewart, and W. A. Darity (2022). Collective wealth and group identity: Insights from Stratification Economics. *Review of Evolutionary Political Economy 3*(3), 463–491.

McBride, M. (2015). Why churches need free-riders: Religious capital formation and religious group survival. *Journal of Behavioral and Experimental Economics 58*, 77–87.

McGee, D. (2025a). Exploitation through racialization. *Quarterly Journal of Economics 140*(2), 1581–1631.

McGee, D. (2025b). Legitimizing myths. *Mimeo*.

McGee, D. (2025c). Stereotypes and strategic discrimination. *Mimeo*.

McMillon, D. (2025). What makes systemic discrimination 'systemic?' Exposing the amplifiers of inequity'. *Mimeo*.

Mills, C. W. (1997). *The Racial Contract*. Ithaca: Cornell University Press.

Moro, A. and P. Norman (2004). A general equilibrium model of statistical discrimination. *Journal of Economic Theory 114*(1), 1–30.

Morrock, R. (1973). Heritage of strife: The effects of colonialist 'divide and rule' strategy upon the colonized peoples. *Science & Society 37*(2), 129–151.

Muller, D. and P. Fabian (2021). Markets reduce discrimination. *Mimeo*.

Niederle, M. and L. Vesterlund (2007). Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics 122*(3), 1067–1101.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review 14*(3), 693–709.

Onuchic, P. (2024). Recent contribution to theories of discrimination. *Conditionally Accepted at Journal of Political Economy: Micro*.

Peski, M. and B. Szentes (2013). Spontaneous discrimination. *American Economic Review 103*(6), 2412–2436.

Postlewaite, A. (2011). Social norms and social assets. *Annual Review of Economics 3*, 239–259.

Roediger, D. R. (1991). *The Wages of Whiteness*. New York: Verso Books.

Schneider, D. (2004). *The Psychology of Stereotyping*. New York: Guilford Press.

Shayo, M. (2020). Social identity and economic policy. *Annual Review of Economics 12*, 355–389.

Shih, M., N. Ambady, J. A. Richeson, K. Fujita, and H. M. Gray (2002). Stereotype performance boosts: The impact of self-relevance and the manner of stereotype activation. *Journal of Personality and Social Psychology 83*(3), 638–647.

Siddique, A. and Vlassopoulos (2020). Competitive preferences and ethnicity: Experimental evidence from Bangladesh. *The Economic Journal 130*(627), 793–821.

Siddique, A., M. Vlassopoulos, and Y. Zenou (2023). Market competition and discrimination. *European Economic Review 152*.

Stackman, V. R., R. Reviere, and B. C. Medley (2016). Attitudes toward marriage, partner availability, and interracial dating among Black college students from Historically Black and Predominantly White Institutions. *Journal of Black Studies 47*(2), 169–192.

Steele, C. M. and J. Aronson (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology 69*(5), 797–811.

Stewart, J. B. (1995). NEA Presidential Address, 1994: Toward broader involvement of Black economists in discussions of race and public policy: A plea for a reconceptualization of race and power in economic theory. *Review of Black Political Economy 23*(3), 13–36.

Sundstrom, W. A. (1990). Half a career: Discrimination & railroad internal labor markets. *Industrial Relations 29*(3), 423–440.

Telles, E. E. (2006). *Race in Another America: The Significance of Skin Color in Brazil*. Princeton: Princeton University Press.

Tilly, C. (1999). *Durable Inequality*. Oakland: University of California Press.

Violante, G. L. (2022). What have we learned from hank models, thus far? *Proceedings of the ECB Forum on Central Banking 2021*.

Walton, G. M. and G. L. Cohen (2003). Stereotype lift. *Journal of Experimental Social Psychology 39*(5), 456–467.