Discrimination as Retaliation*

Till Wicker

October 9, 2025

[Link to latest version]

Abstract

Discrimination remains pervasive, but little is known about how past personal experiences of discrimination shape our future discriminatory behavior. This paper introduces and empirically documents retaliatory discrimination, whereby individuals are more likely to discriminate against a group after perceiving that they were personally discriminated against by members of that group. Guided by a theoretical framework that models retaliatory discrimination alongside taste-based and statistical discrimination, I conduct experiments with refugees in Uganda and men in the United States. In a two-stage experiment, I first randomly expose participants to (un)fair task allocations from managers of varying identities: coethnic, non-coethnic, or computer. I then observe whether they discriminate against non-coethnic workers when later placed in a managerial role. Negative past experiences with a non-coethnic manager increase subsequent discrimination by 78%, driven by an increase in the number of discriminators rather than in the intensity of discrimination. I distinguish between four pre-registered micro-foundations and find empirical support for motivated beliefs, while ruling out alternative mechanisms. Retaliatory discrimination provides a behavioral foundation for anticipated discrimination and the evolution of discriminatory tastes. Its policy relevance is underscored by an experiment showing that removing affirmative action policies triggers a backlash that amplifies discrimination, consistent with the model's predictions.

*t.n.wicker@tilburguniversity.edu. AEARCTR-0015358 and AEARCTR-0016047. IRB approval was obtained from Tilburg University (IRB FUL 2024-015, TiSEM_RP2233) and Mildmay Institute of Health Sciences (MUREC-2025-790). I am grateful to Patricio Dalton and Daan van Soest for excellent supervision. I further thank Quamrul Ashraf, Aditi Bhowmick, Luisa Cefala, Alex Chan, Elena Cettolin, Rema Hanna, Sylvan Herskowitz, Jonas Hjort, Yuen Ho, John Horton, Alex Imas, Kelsey Jack, Pamela Jakiela, Supreet Kaur, Kevin Lang, Louis-Pierre Lepage, Ulrike Malmendier, Jeremy Magruder, Benjamin Marx, Edward Miguel, Francesca Miserocchi, Owen Ozier, Gautam Rao, Gerard Roland, Frank Schilbach, Emma Smith, Sigrid Suetens, Denni Tommasi, Dominik Wehr, and Duncan Webb for insightful comments. Special thanks to Alexander Negassi and Noah Sumile for excellent research assistance.

1 Introduction

> *Minority groups are often tempted to "retaliate" against discrimination from others by returning the discrimination* (Becker, 1957)

Discrimination has been documented across a variety of domains, including the labor market (Neumark, 2018), real estate (Bartoš et al., 2016), credit markets (Fisman et al., 2020; Fiorin et al., 2025), bail decisions (Arnold et al., 2022), and medical referrals (Sarsons, 2022). Individuals of both minority and majority groups also perceive widespread discrimination against their own in-group: 24% of Black and Hispanic workers, and 13% of White workers, in the US reported experiencing discrimination at work (NPR, 2017; Gallup, 2021).[1] However, our understanding of how perceived discrimination shapes current and future discriminatory preferences, and subsequently behavior, is limited.

This paper shows that individuals discriminate more against a group after perceiving discrimination from that group. The new documented source of discrimination — retaliatory discrimination — can rationalize patterns in the data that the two existing workhorse models of discrimination, taste-based (Becker, 1957) and statistical discrimination (Arrow, 1972a,b; Phelps, 1972), struggle to explain, such as the immediate and persistent increases in discrimination against minorities after ethnic conflicts (Hjort, 2014; Shayo and Zussman, 2017; Fisman et al., 2020), terrorist attacks (Kaushal et al., 2007; Glover, 2019), and pandemics (Lu and Sheng, 2022; Luca et al., 2024; Lanzara et al., 2025).

Incorporating retaliatory discrimination into a theoretical framework generates novel predictions about how past experiences shape current discriminatory preferences and actions, which I test using experiments in Uganda and the USA. In the first stage of the experiments, participants are paired with a non-coethnic worker and complete tasks assigned to them by a manager. The manager's ethnicity and allocation of tasks across the two workers vary exogenously, such that participants are randomly assigned to a coethnic or non-coethnic manager, who either divides the tasks fairly or favors one of the two workers. In a subsequent, second stage, participants become the manager, and allocate tasks between a coethnic

[1] Over 60% of Black people and Muslims regularly experience discrimination in Germany (DeZIM, 2023; Welle, 2025). Of those who perceive discrimination in the US, 49% and 61% of Black and White Americans respectively say the larger problem is discrimination based on the prejudice of individual people, rather than due to laws and government policies (NPR, 2017).

and non-coethnic worker. Workers are paid a piece rate per task and hence the manager's allocation of tasks across workers in the second stage is a measure of discrimination.[2]

After being randomly assigned in the first stage to a non-coethnic manager who gives them less than half the tasks, participants retaliate against the non-coethnic worker in the second stage: they allocate them fewer tasks, despite the worker being unrelated to their previous manager. Discrimination in the allocation of tasks to the non-coethnic worker increases by 78%, compared to treatment arms where the non-coethnic manager in the first stage allocates a fair (or favorable) number of tasks, reducing the non-coethnic worker's earnings by 15%. This increase reflects an expansion at the extensive margin — a 41% rise in the number of discriminators — rather than at the intensive margin among those already discriminating. Retaliatory discrimination is less severe than direct retaliation against the manager from the first stage, and is not documented when the manager in the first stage is coethnic. The findings are in line with theoretical predictions of retaliatory discrimination, but contrast predictions of other models of discrimination. Sub-treatments rule out alternative explanations — including inaccurate statistical discrimination, tit-for-tat, in-group favoritism, social preferences, inequality aversion, and norm violations — while supporting motivated beliefs as a micro-foundation of retaliatory discrimination.

Two applications of retaliatory discrimination highlight its importance. Firstly, a subtreatment of the experiment simulates the removal of affirmative action policies: participants were informed that managerial allocations in the first stage of the experiment were subject to an affirmative action policy that was removed before the second stage. The policy reversal induced stronger retaliatory discrimination compared to a treatment arm where participants received no justification for the stage 1 manager's allocation of tasks, lowering the non-coethnic worker's earnings by a further 4%. This stands in contrast to models of taste-based and statistical discrimination, which do not predict any increase in discrimination.[3] Secondly, a real effort task embedded within the experiment illustrates that retaliatory discrimination can be a micro-foundation for anticipated discrimination. After the first two

[2]This definition is in line Bohren et al. (2025b) — who define discrimination as "disparities arising from the direct effects of group identity" — and Lang and Kahn-Lang Spitzer (2020) — who define discrimination as "treating someone differently based on characteristics such as gender, race, or religion."

[3]Findings are in line with laboratory and field evidence of backlash after the introduction of (affirmative action) quotas (Leibbrandt et al., 2018; Petters and Schröder, 2020; Belmonte and Di Lillo, 2021; Neschen and Hügelschäfer, 2021).

stages, participants signaled their productivity to a non-coethnic manager who was hiring, by completing as many tasks as possible in 60 seconds. Having a non-coethnic manager who favored the other worker in the first stage reduced the number of tasks completed by 12%. The negative past experience with a non-coethnic manager increased expected discrimination, which lowered the effort the participant put into the productivity signal to the future manager.

Identifying a new source of discrimination also presents new mitigation measures. Given the dynamic nature of retaliatory discrimination, increasing the salience of future interactions (and hence the consequences of current discriminatory actions), could reduce discrimination. To investigate this, I conduct a sub-treatment that exogenously varies the salience of future interactions. This treatment increases the number of tasks allocated to the non-coethnic worker in the second stage, providing suggestive evidence that increasing the salience of future interactions can reduce retaliatory discrimination.

Related literature This paper contributes to three strands of literature. Firstly, it contributes to the theoretical literature on discrimination. The workhorse models of discrimination — taste-based (Becker, 1957) and statistical discrimination (Arrow, 1972a,b; Phelps, 1972) — have been extended in recent years to allow for inaccurate beliefs (Bohren et al., 2025a), experience-based learning (LePage, 2024), and paternalistic discrimination (Buchmann et al., 2024). I define and identify a novel source of discrimination, retaliatory discrimination, which evolves in response to past experiences. Unlike taste-based discrimination (Becker, 1957), which models discriminatory preferences as exogenous, I illustrate that these preferences depend on past experiences and can thus be endogenous. It differs from statistical discrimination as retaliatory discrimination does not arise due to imperfect information. Nevertheless, the channel through which past experiences with individuals shape endogenous, group-level discriminatory tastes mirrors how experience-based discrimination can micro-found statistical discrimination (LePage, 2024; Benson and Lepage, 2024).

Retaliatory discrimination, and the endogenous updating of group-level discriminatory preferences following interactions with individuals, builds on a literature in social psychology on vicarious retribution and group generalization (Lickel et al., 2006; Paolini et al., 2010; Barlow et al., 2012; Paolini et al., 2024). This paper is the first to formalize an economic model of retaliatory discrimination, providing a micro-foundation for the emergence and

persistence of discriminatory tastes, something which economists have thus far overlooked (Cain, 1986). I furthermore propose and empirically test several pre-registered functional forms of retaliatory discrimination, building on recent work linking motivated beliefs and memories to discriminatory behavior (Eyting, 2022; Miserocchi, 2023).

Secondly, this paper contributes to the existing literature using lab and field settings to document discrimination and its determinants (Bertrand and Duflo, 2017; Neumark, 2018). I develop a novel experimental setting that differs from existing studies by consisting of multiple stages, hence allowing for past experiences to affect future discriminatory actions. Results of the two experiments are in contrast to predictions of taste-based and statistical discrimination, but are in line with predictions of retaliatory discrimination. I contribute to discussions surrounding the importance of correctly identifying the source and nature of discrimination for policy recommendations (Bohren et al., 2025a), by experimentally showing that the removal of affirmative action policies can greater subsequent discrimination against non-coethnic workers.[4] Lastly, I contribute to the empirical literature documenting anticipated discrimination (Charness et al., 2020; Agüero et al., 2023; Aksoy et al., 2023; Angeli et al., 2025; Gagnon et al., 2025), by establishing a causal link between negative past experiences, endogenous discriminatory tastes, and anticipated discrimination.

Thirdly, this paper contributes to the literature on the role of past experiences on economic decisions (Malmendier, 2021; Giuliano and Spilimbergo, 2025). While empirical support primarily comes from financial decisions, Fisman et al. (2020) document how exposure to early-life religious riots affect the lending behaviors of bank managers through increased discrimination. While these studies look at the consequences of past macro-level experiences (such as financial crises, or riots) on present-day decisions, this paper looks at the channels through which individual, micro-level past experiences can affect future discriminatory preferences and behaviors. As such, retaliatory discrimination can offer a micro-foundation for the emergence, persistence, and consequences of inter-group tensions on economic transactions, with respect to both microeconomic interactions (Hjort, 2014; Ghosh, 2025), and macro-level relationships between ethnic divisions, conflict, and economic development (Alesina and Ferrara, 2005; Arbatlı et al., 2020).[5]

[4] In Section 3, I experimentally show that academic economists misidentify retaliatory discrimination as taste-based when they are unaware of the existence of previous rounds.

[5] Related literatures are those regarding positive inter-group contact hypothesis (Lowe, 2025) and the persistence of attitudes against (minority) groups (Schindler and Westcott, 2020; Bursztyn et al., 2024).

This paper proceeds as follows: Section 2 develops a theoretical model that incorporates taste-based, statistical, and retaliatory discrimination, generating two novel predictions. These predictions are subsequently taken to the data in Sections 3 and 4, which present results from experiments in Uganda and the USA. Section 5 distinguishes between four pre-registered micro-foundations, and hence functional forms, of retaliatory discrimination, before Section 6 discusses two applications: the removal of affirmative action policies, and anticipated discrimination. Section 7 presents a measure to reduce retaliatory discrimination, and Section 8 concludes.

2 Theoretical Model

I develop a theoretical model that incorporates taste-based, statistical, and retaliatory discrimination. While discrimination is pervasive across a variety of domains, most theoretical models and empirical applications — including the experiments in Sections 3 and 4 — focus on the labor market. As such, the theoretical model discussed in this section is specific to the labor market. A more general model of discrimination is presented in Appendix A, reflecting its generalizability to other discriminatory settings, such as teachers grading students (Carlana, 2019; Miserocchi, 2023), or loan officers awarding loans to applicants (Fisman et al., 2020).

2.1 Basic Model of Labor Market Discrimination

An employer decides how many workers to hire from groups A and B at time t to maximize their expected utility. Their expected utility is linear and additively separable along two dimensions:

1. The expected firm profit from hiring L_A and L_B workers from groups A and B, respectively: $\pi_t = Y_t(L_{A,t}, \theta_A, L_{B,t}, \theta_B) - w_A L_{A,t} - w_B L_{B,t}$. Profits depend on the number of workers hired from groups A and B at time t ($L_{A,t}, L_{B,t}$), their productivity (θ_A, θ_B, unknown to the employer), and their wages (w_A, w_B). This generic specification can capture the setting where workers of both groups are perfect substitutes in production (Becker, 1957), as well as the case where output is a function of the group-specific productivity (Bohren et al., 2025a).

2. The non-pecuniary costs of hiring workers from groups A and B: $f\Big(d_A, F(\chi_{A,t})\Big)L_{A,t} + f\Big(d_B, F(\chi_{B,t})\Big)L_{B,t}$. This group-specific cost captures both a fixed, time-invariant "taste" parameter, d_g, as well as a dynamic component that is a function of cumulative past experiences (χ) with individuals of group g at time t, $F(\chi_{g,t})$. Both components are group-specific. The function f is weakly increasing and concave in both the static and dynamic variables.[6,7] The non-pecuniary cost term enters the employer's maximization problem in the same way as an effective increase in the wage of group g. A higher value of $f(d_g, F(\chi_{g,t}))$ makes hiring workers from that group more "costly", even though this cost is psychological rather than monetary. The employer therefore behaves as if the wage w_g were higher for workers in groups associated with greater non-pecuniary costs. Consequently, shifts in $F(\chi_{g,t})$ directly alter the cost of hiring workers from different groups, providing a channel through which prior experiences can translate into current discriminatory behavior.

In particular, the employer's utility function is:

$$\max_{L_{A,t},L_{B,t}} \underbrace{Y(L_{A,t},\theta_A,L_{B,t},\theta_B) - \sum_{g\in\{A,B\}} L_{g,t}w_g}_{\text{Firm Profit}} - \underbrace{\sum_{g\in\{A,B\}} L_{g,t} f\Big(d_g, F(\chi_{g,t})\Big)}_{\text{Non-Pecuniary Costs}} \tag{1}$$

The employer's utility function in equation 1 has two conceptually distinct components: firm profits and non-pecuniary costs. The first term, $Y(L_{A,t},\theta_A,L_{B,t},\theta_B) - \sum_g L_{g,t}w_g$, captures firm output and wage payments. The second term, $\sum_g L_{g,t} f(d_g, F(\chi_{g,t}))$, introduces a psychological cost associated with employing individuals from group g. This cost reflects both a static preference component d_g, which represents the employer's underlying taste for or against a group (as in Becker (1957)), and a dynamic component $F(\chi_{g,t})$, which depends on the employer's accumulated past experiences with that group. Intuitively, if previous interactions with workers from group g were perceived as negative or discriminatory, $F(\chi_{g,t})$

[6]Mathematically, this means that $\frac{\partial f}{\partial d_g} \geq 0, \frac{\partial f}{\partial F(\chi_{g,t})} \geq 0, \frac{\partial^2 f}{\partial d_g^2} \leq 0, \frac{\partial^2 f}{\partial F(\chi_{g,t})^2} \leq 0, \quad G \in \{A,B\}$. d_g and $F(\chi_{g,t})$ can be either substitutes, or complements.

[7]Prior to starting the online experiment in the USA, I pre-registered four micro-foundations of $f(d_g, F(\chi_{g,t}))$ — Retaliatory Tit-for-Tat, Bayesian Updating, Motivated Beliefs, and Memory Recall. The pre-registration can be found at the AEA RCT Registry under AEARCTR-0016047. Empirical support for these micro-foundations will be discussed in Section 5.

increases, thereby raising the disutility of hiring workers from group g in subsequent periods. In this way, past interactions shape current discriminatory behavior by endogenously adjusting the perceived cost of hiring workers from each group.

The true productivity of a worker, θ, is not observable to the employer at the time of hiring. The productivity is drawn from a group-specific normal distribution $\theta_g \sim N(\mu_g, 1/\tau_g)$. Workers send a signal of their productivity to the employer equal to $s = \theta + \epsilon$, where $\epsilon \sim N(0, 1/\eta_g)$. Employers have priors about the the productivity distribution of group g ($\hat{\theta}_g \sim N(\hat{\mu}_g, \hat{\tau}_g)$), as well as the precision of the signal from group g ($\hat{\eta}_g$). Following Bohren et al. (2025a), we denote an employer's subjective group-specific beliefs by $\psi_g \equiv (\hat{\mu}_g, \hat{\tau}_g, \hat{\eta}_g)$. After observing the worker's group identity g and signal s, the employer forms a posterior belief about the worker's productivity using Bayes' Rule.[8]

Therefore, there are two conceptually separate channels through which group membership affects hiring decisions. The first is through imperfect information: employers may hold different priors $\psi_g = (\hat{\mu}_g, \hat{\tau}_g, \hat{\eta}_g)$ about the expected productivity or signal precision of workers from different groups, leading to statistical discrimination (Arrow, 1972a; Phelps, 1972). Past experiences with workers of group g can micro-founding statistical discrimination, by providing information about group level productivity, as discussed by LePage (2024). The second channel is through preferences: in addition to exogenous tastes d_g (Becker, 1957), past interactions with members of a group can affect the employer's perceived cost of hiring workers from that group through $F(\chi_{g,t})$. Unlike statistical discrimination, this retaliatory channel does not stem from updating beliefs about worker productivity but from the updating of preferences themselves. The coexistence of these two mechanisms means that discrimination can persist and evolve even when employers have accurate information about productivity distributions.

Definition 1: Discrimination is defined as follows:

$$D_t(s, \psi_A, \psi_B) \equiv L_{A,t}|s - L_{B,t}|s$$

namely the differential hiring of a worker of groups A and B at time t by an employer with subjective beliefs ψ_A and ψ_B, conditional on workers sending the same productivity

[8] I am being agnostic as to whether the employers' priors are accurate or inaccurate. For more discussion on this, see Bohren et al. (2025a).

signal s. Discrimination occurs when $D_t(s, \psi_A, \psi_B) \neq 0$; the employer discriminates against individuals of group A if $D_t(s, \psi_A, \psi_B) < 0$, and discriminates against individuals of group B if $D_t(s, \psi_A, \psi_B) > 0$.

2.2 Incorporating Other Models of Discrimination

Equation 1 presents an employer's maximization problem when deciding to hire workers from groups A and B and provides a general framework that encompasses several well-known theories of discrimination as special cases. By turning on or off specific terms, we can recover classic results. When only the static taste component d_g matters, the model collapses to Becker's (1957) taste-based discrimination. When the non-pecuniary cost is absent but employers hold group-differentiated beliefs about productivity, we obtain statistical discrimination (Arrow, 1972a; Phelps, 1972). The model thus generalizes these frameworks and extends them by allowing the disutility of hiring a group to evolve with experience, giving rise to retaliatory discrimination.

2.2.1 Taste-Based Discrimination

Setting $f(d_g, F(\chi_{g,t})) = d_g$ simplifies equation (1) to the following:

$$\max_{L_{A,t}, L_{B,t}} Y(L_{A,t}, \theta_A, L_{B,t}, \theta_B) - \sum_{g \in \{A,B\}} L_{g,t} w_g - \underbrace{\sum_{g \in \{A,B\}} L_{g,t} d_g}_{\text{Distaste}}$$

If workers are perfect substitutes and no productivity signal s is sent, the model simplifies to the taste-based discrimination model of Becker (1957). In this setting, past experiences do not influence the employer's decision in the present period, nor the level of discrimination, D_t. The non-pecuniary costs associated with hiring a worker from group g are due to the employer's static discriminatory taste d_g. Discriminatory tastes and behaviors are time-invariant and hence $D_t = D \ \forall t$, ceteris paribus.

2.2.2 Statistical Discrimination

If managers do not display a non-pecuniary distaste towards workers, equation (1) simplifies to:

$$\max_{L_A, L_B} Y(L_{A,t}, \theta_A, L_{B,t}, \theta_B) - \sum_{g \in \{A,B\}} L_{g,t} w_g$$

Managers do not observe the true productivity of the worker, but have priors about the productivity distribution and signal precision of workers from group g ($\psi_g \equiv (\hat{\mu}_g, \hat{\tau}_g, \hat{\eta}_g)$). Differences in both the true and believed moments of the productivity distribution and signal precision of both groups can give rise to (accurate and inaccurate) statistical discrimination (Arrow, 1972a,b; Phelps, 1972; Bohren et al., 2025a).

Equation (1) therefore incorporates both taste-based discrimination à la Becker (1957), and accurate and inaccurate statistical discrimination, while allowing for non-pecuniary costs to evolve based on past experiences (retaliatory discrimination). Appendix B extends the theoretical model to allow for paternalistic discrimination (Buchmann et al., 2024) and experience-based discrimination (LePage, 2024).

2.3 Predictions

The dynamic term $F(\chi_{g,t})$ implies that employers' discriminatory preferences are not fixed, but evolve in response to past interactions. Suppose an employer has a negative experience with an individual from group B. This negative experience increases $F(\chi_{B,t})$, thereby raising the non-pecuniary cost of hiring workers from group B in the future. In turn, the employer is less willing to hire workers from group B in subsequent periods, even when new applicants from A and B send identical productivity signals. This self-reinforcing dynamic generates retaliatory discrimination: discrimination that arises endogenously from past experiences and their effect on discriminatory preferences. Because $F(\chi_{g,t})$ is defined separately for each group, these effects are group-specific.

Without loss of generality, I assume that the employer discriminates against workers from group B, and hence $D_t(s, \psi_A, \psi_B) > 0$. This model generates two novel predictions about the role of past experiences on present discriminatory preferences, compared to the models of taste-based and statistical discrimination:

Prediction 1: (*Retaliatory Discrimination*) Ceteris paribus, more negative past ex-

periences with workers from group B at time t ($\chi^{\text{mod}}_{B,t} < \chi^{\text{neg}}_{B,t}$) have a non-negative effect on discrimination against workers of group B at time t:

$$\chi^{\text{mod}}_{B,t} < \chi^{\text{neg}}_{B,t} \quad \Rightarrow \quad D_t(s, \psi_A, \psi_B | \chi^{\text{mod}}_{B,t}) \leq D_t(s, \psi_A, \psi_B | \chi^{\text{neg}}_{B,t})$$

See the proof in Appendix C. If $\frac{\partial f}{\partial F(\chi_{g,t})} > 0$ (strict inequality), then more negative past experiences with workers from group B will *increase* the employer's discrimination against workers of group B.

Prediction 2: (*Group-Specific Retaliatory Discrimination*) Ceteris paribus, more negative past experiences with individuals from group $g' \neq B$, $g' \neq A$ at time t ($\chi^{\text{mod}}_{g',t} < \chi^{\text{neg}}_{g',t}$) have no effect on discrimination against workers of group B :

$$\chi^{\text{mod}}_{g',t} < \chi^{\text{neg}}_{g',t} \quad \Rightarrow \quad D_t(s, \psi_A, \psi_B | \chi^{\text{mod}}_{g',t}) = D_t(s, \psi_A, \psi_B | \chi^{\text{neg}}_{g',t})$$

See the proof in Appendix C. These two predictions capture that past interactions can affect current discriminatory attitudes, however these past interactions, and hence their consequences, are group-specific.

3 Main Experiment: Uganda

The theoretical model thus predicts that past negative interactions with individuals of a certain group will result in greater future discrimination towards other members of the same group (Prediction 1). However, this retaliatory discrimination is group-specific (Prediction 2). To test these predictions, a lab-in-the-field experiment was conducted among 224 Eritrean refugees in Kampala, Uganda, in Spring 2025. Uganda was home to 58,720 Eritrean refugees in April 2025, of which 98% lived in Kampala, the country's capital (UNHCR, 2025). Due to Uganda's progressive policies, refugees have the freedom of movement and right to work in Uganda. Despite their refugee status, Eritreans have similar economic opportunities to Ugandans in Kampala, due to their comparable levels of education and network-based hiring. Furthermore, living situations and housing quality are comparable to Ugandans. This makes Uganda an ideal context for this topic, as refugees have more autonomy and opportunities than in many other settings.

Despite ample opportunities, economic integration between Eritreans and Ugandans is

limited. Many Kampala-based Eritrean refugees live in the same neighborhoods and form a tight-knit community. They therefore rarely engage with Ugandans, in part driven by the language barrier, and as a consequence, Eritreans tend to work and hire among themselves. Similarly, Ugandans rarely hire Eritreans, limiting labor market integration. This is reflected in the study's sample, who have an average of only 2.52 Ugandan friends (see Appendix Table A3), and 23.21% of participants felt that Ugandan firms discriminated against them.[9]

Participants in the lab-in-the-field were male, with an age range from 18 to 51 years (mean: 30.75 years). The earliest year of arrival in Uganda was 1990 and the latest arrival year was 2024 (mean: 2016). Participants were recruited for a short work task, and completed the experiment independently in a private environment.

3.1 Experimental Design

Figure 1 depicts the experimental design, which consisted of two stages. In both stages, a manager delegated 8 tasks between two workers. The manager was paid a fixed wage, however workers were paid a piece rate of 500 UGX per completed task.[10] Workers and managers were given alias names that revealed their nationality, but preserved their anonymity.[11]

The task consisted of making an envelope (used for a cash transfer, as in Wicker et al. (2025) and inspired by Bulte et al. (2020)) out of a sheet of A4 paper. This was a novel task that participants had never completed before, hence reducing the likelihood that participants had strong priors regarding differential abilities of Ugandans and Eritreans at completing the task. This reduces the role of statistical discrimination.

[*Figure 1 here*]

Before the first stage of the game, participants were informed both verbally and in writing that (i) the other participants were based in different regions of Uganda, (ii) no communication or interaction would take place between the manager and the workers, (iii)

[9] This is in line with insights of Loiacono and Vargas (2019) and Loiacono and Silva Vargas (2025).

[10] 500 UGX ≈ \$0.14. Additionally, participants received 2000 UGX as a show-up fee. Average compensation was 3500 UGX, and the study lasted 20 minutes on average. As such, average compensation equaled half a day's worth of work, and each envelope (which took less than 2 minutes per envelope) corresponded to 37 minutes of work at the minimum wage. The monthly minimum wage in Uganda is 130,000 UGX.

[11] Appendix D illustrates that during pilot work, both Eritreans and Ugandans were able to correctly identify the nationality of an individual based on their name in 97% of the cases.

there would be no future interactions, and that (iv) none of the workers had completed this task before. Additionally, participants were shown data from the pilot study, illustrating that Ugandans and Eritreans were on average equally good at making the envelopes (see Appendix D.2). Through these design choices, I am able to minimize the role of strategic concerns, future interactions, and (inaccurate) statistical discrimination. Participants were further shown how to complete the task by the enumerator, and made a practice envelope before commencing with the two stages of the game.

In the first stage of the experiment, the Eritrean participant ($\mathbf{E_1}$ in Figure 1) was assigned the role of one of the two workers. They were informed that they were paired with a Ugandan male worker U_1 (signaled by their name), and that a manager had decided the allocation of the eight tasks across the two workers. Experimental variation came in the nature of the manager in stage 1 (M_0 in Figure 1): the manager was (i) either a Computer or a Ugandan, and (ii) either divided the eight tasks evenly across the two workers (4, 4); or assigned more tasks to the Ugandan worker (6, 2). Allocations are pre-determined, and based on actual decisions made by Ugandans during the pilot study. As such, there are four treatment arms, as depicted in Figure 1.

Once participants learned how many tasks they had been assigned by their manager, they made the envelopes. The enumerator recorded how long it took the participants to make the envelopes, and after data collection was completed, the enumerators evaluated the quality of the envelopes based on five dimensions.[12]

The first stage of the experiment finished once the participant completed making the envelopes, after which the second stage of the experiment commenced. Importantly, the participant did not receive any feedback regarding the quality of the envelopes they, or their paired Ugandan worker (U_1), made. As such, the information set available to the participant regarding the relative productivity of Ugandans and Eritreans did not change throughout stage 1, nor across the four treatment arms.

The set-up of the second stage was identical to the first stage, except that this time, the Eritrean participant ($\mathbf{E_1}$) was the manager who had to delegate 8 tasks between two male workers: one Ugandan (U_2) and one Eritrean refugee (E_2), neither of whom the participant,

[12]The five dimensions are: sides of envelope have a finger width; triangle fold is in the middle; creases are tight and straight; glue still sticks; top fold is sharp. For each envelope, these categories received a binary score that were subsequently averaged across envelopes.

nor their previous manager (M_0) had interacted with before. In this stage, the Eritrean participant ($\mathbf{E_1}$) was paid a flat wage, while workers were paid a piece-rate for every produced envelope.

3.2 Outcome Variables

The primary pre-registered outcome variable is the allocation of tasks across the two workers in the second stage of the experiment, as a measure of discrimination based on Definition 1: any deviation from an equal split of the 8 tasks indicates discrimination. Further pre-registered outcome variables are the time taken to make the envelopes in the first stage of the experiment, and quality of the envelopes.

3.3 Predictions from Models of Discrimination

Taste-based and statistical discrimination do not predict differential discrimination (and hence allocation of tasks) across the four treatment arms. This is because discriminatory tastes are exogenous, and participants do not differentially learn about individual- or group-level productivity across the four treatment arms. Retaliatory discrimination, on the other hand, and Prediction 1 of Section 2, predicts that participants randomly assigned to a Ugandan stage 1 manager who allocates few tasks to them (T3) will retaliate against the Ugandan worker in the second stage, resulting in more discrimination compared to the case when a Ugandan stage 1 manager allocates tasks evenly (T4). Prediction 2 argues that a negative experience with the Computer manager in stage 1 (T1) will not affect stage 2 allocations.

Appendix E presents detailed theoretical predictions of taste-based, statistical, and retaliatory discrimination, as well as other explanations (including paternalistic discrimination, systemic discrimination, social norms, fairness concerns, and experimenter preferences), with differing theoretical predictions. However none of these models generate the same empirical predictions as retaliatory discrimination.

3.4 Results

Allocation of Tasks as a Manager

Figure 2 presents the Eritrean participant's allocation of tasks to the Ugandan worker (U_2) when they were the manager in the second stage of the game. The participant had to divide 8 tasks, and hence allocating 4 tasks to the Ugandan worker would have been an equal division of tasks, and hence no discrimination ($D_t = 0$). This is represented by the dashed horizontal gray line at $y = 4$. Any allocation of tasks that is not an even split between the two workers is categorized as discrimination, following the definition from Section 2: $D_t = L_{A,t}|s - L_{B,t}|s$.

Eritrean participants allocate fewer tasks to the Ugandan worker (and hence more to the Eritrean worker E_2) when they are the manager in the second stage of the experiment, on average 3.49 tasks ($p < 0.001$). This suggests some degree of discrimination against the Ugandan. By providing group-level statistics of the productivity of Eritrean and Ugandan workers during the pilot of this low-skill task, I minimize the role of statistical discrimination, following the approach of Bohren et al. (2025a), Chan (2025), and Montoya et al. (2025). However, I cannot distinguish whether the differential allocation is due to taste-based discrimination, statistical discrimination, or alternative explanations (e.g. fairness considerations, see Appendix E).

When the computer is the manager in the first stage of the experiment (referring to T1 and T2, the two bars on the left-hand-side in Figure 2), the allocation of tasks in the second stage do not differ depending on whether the participant was allocated two or four tasks in the first stage (T1 vs. T2, $p = 0.389$). This is in line with *Prediction 2*, as unrelated past experiences do not affect current discriminatory actions.

[*Figure 2 here*]

When the manager in the first stage is a Ugandan who evenly splits the tasks between the two workers (T4, the furthest right bar in in Figure 2), Eritrean managers allocate slightly more tasks to the Ugandan worker than when the Computer is the manager, but this difference is not statistically significant (T1 & T2 vs. T4, $p = 0.199$).[13] However,

[13]Eritreans had different prior expectations of how many envelopes they would receive in the first stage when the manager was a Computer vs. a Ugandan (3.87 vs. 4.21, $p = 0.015$). Eritreans thought the computer was not biased ($p = 0.206$).

when the Ugandan manager in the first stage allocates more tasks to the Ugandan than the Eritrean worker, the Eritrean participant retaliates in the second stage, and gives only 3.09 tasks to the Ugandan worker — despite the Ugandan worker (U_2) not being related to the previous Ugandan manager (M_0). Compared to when the Ugandan manager in the first stage evenly splits the tasks, this difference is highly statistically significant (T3 vs. T4, $p < 0.001$). This allocation is also statistically significantly different compared to when the Computer manager allocated fewer tasks to the participant in the first stage (T1 vs. T3, $p = 0.038$).[14] This provides support for Prediction 1.

Documenting increased discrimination in response to previous perceived discrimination raises the question of whether the average increase in discrimination is due to more people discriminating, or the same number of people discriminating more aggressively? There is no difference in the number of discriminators, or the magnitude of discrimination, when the manager in the first stage is a computer compared to the setting where the Ugandan manager treats both workers evenly in stage 1 (T1 & T2 vs. T4, $p = 0.388$ and $p = 0.528$, respectively).

There are statistically significantly more discriminators when the Ugandan manager favors the Ugandan worker in stage 1: 40.35% of participants discriminate (defined as assigning fewer than four tasks to the Ugandan worker in stage 2, following Definition 1) when their previous Ugandan manager treated them fairly. When their previous Ugandan manager treated them unfairly, this number jumps to 57.14%, a 17pp increase (41%, $p = 0.075$). However, conditional on discriminating, individuals discriminate less aggressively on average in T3, allocating 2.91 tasks to the Ugandan worker in stage 2, compared to 2.41 tasks in T4 ($p < 0.001$).[15] This indicates that retaliatory discrimination creates new discriminators, rather than heightening discriminatory attitudes among those who already discriminate. Without knowledge of the first round of the game and retaliatory discrimination, these individuals would have been mis-categorized as taste-based discriminators, as I show next.

[14] The regression tables underlying Figure 2 are presented in Appendix Tables A8 and A9.

[15] Appendix H.2 presents histograms of the allocations to the Ugandan worker in stage 2, across Treatments 1-4.

Misidentifying Retaliatory as Taste-Based Discrimination

Through a survey of academics, I illustrate that only considering behavior in round t, without considering rounds $t-i$, $i>0$ can lead to the mis-classification of retaliatory discrimination as taste-based.

51 academics were randomly shown either the allocation of tasks to the Ugandan worker of Treatments 3 or 4 of Figure 2.[16] First, participants received an overview of the two workhorse models of discrimination and subsequently the experimental set-up of stage 2 of the experiment among refugees in Uganda. Then, half of the participants were randomized to see the division of tasks across the Ugandan and refugee worker in T3, while the other half were seen the division of tasks in T4. Not only did the magnitude of discrimination differ significantly across the two treatments (following Definition 1, $p = 0 < 001$), but also the source of discrimination, and extensive margin of discriminators ($p = 0.075$). While the discrimination in T4 can be attributed to taste-based or statistical discrimination (although I aim to minimize the prevalence of statistical discrimination), the additional discrimination observed in T3 compared to T4 can be causally attributed to the allocation of tasks by the manager in the first stage, and hence retaliatory discrimination.

However, experts overwhelming identify the source of the discrimination as being taste-based in both treatment arms. 65.4% and 64.0% of respondents in the two treatments identified the source of discrimination as taste-based ($p = 0.920$).[17] This illustrates that, if researchers are unaware of retaliatory discrimination, the source of discrimination can be mis-identified as other forms of discrimination, in particular taste-based discrimination. As Section 6 will illustrate, mis-identifying the source of discrimination can have policy consequences. Another insight is that retaliatory discrimination can provide a micro-foundation for prejudice, which is often assumed to be exogenous and fixed - while this paper illustrates is can evolve as a result of past interactions and is hence dynamic.

[16] The study was not hosted on the Social Science Prediction Platform as the primary outcome variable was not a quantity, but instead the source of discrimination.

[17] 67% of the respondents were graduate students, 31% were faculty, and 2% were working in the private sector post-PhD. 22%, 22%, and 14% had worked on topics related discrimination, refugees, and Uganda, respectively.

Time Taken and Envelope Quality as a Worker

In addition to retaliating against future individuals of the same ethnic group as the manager in stage 1, participants could also "retaliate" against the manager directly by producing lower-quality envelopes — despite this having no effect on the manager's payoff. This form of futile retaliation has been documented in impunity games (Bolton et al., 1998; Yamagishi et al., 2012), and can also indicate reduced effort in response to perceived discrimination (Gagnon et al., 2025; Ruebeck, 2025).

Tables A11 and A12 illustrates that the quality of the envelopes — defined as the average of five pre-registered quality measures — decreases by $\sim 20\%$ as a result of having a discriminatory Ugandan manager, compared with when the Ugandan manager divides tasks evenly (T3 vs. T4, $p = 0.159$). This provides suggestive evidence that workers engage in tit-for-tat retaliation against the manager directly, where possible, but subsequently also retaliate against other individuals of the same background as the manager when they are subsequently placed in a consequential decision-making role. Gagnon et al. (2025) find that workers put in less effort into a work taste after perceiving discrimination. Appendix Table A10 documents no statistically treatment effects on the worker's effort, defined as the time taken to complete the envelopes.[18,19]

3.5 Discussion

The lab-in-the-field experiment in Uganda provides causal evidence of retaliatory discrimination, as the documented patterns across the four treatment arms cannot be rationalized by taste-based or statistical discrimination, or other explanations (see Appendix E). Instead, results from the experiment align with Predictions 1 and 2 of Section 2.

Through eliciting participant's priors about how many tasks they expected to receive, we can learn about the role of expectations and beliefs in retaliatory discrimination. Table A14 regresses the discrepancy between a participant's expected number of tasks in stage 1, and the actual number of tasks they received in stage 1, on the number of tasks assigned

[18] Appendix Table A13 finds no treatment effect on the expected number of tasks received in the first stage — a placebo test that was elicited prior to the experiment's first round.

[19] Appendix Tables A15-A19 present heterogeneous heterogeneous treatment effects by their number of Ugandan friends, empathy, retaliation, attitudes towards Ugandans, or years spent in Uganda. No consistent patterns are documented, however this could also be due to limited statistical power.

to a Ugandan worker in the second stage of the experiment. While coefficients cannot be interpreted causally (as expectations are endogenous), the magnitude and sign of the coefficients in columns (1) and (2) indicate that when individuals received fewer tasks than they expected from a Ugandan manager in the first stage, they retaliated in the second stage by assigning fewer tasks to the unrelated Ugandan worker. The same pattern is not observed when the individual received fewer tasks than expected from the Computer manager. This, combined with qualitative evidence from the pilot study that receiving fewer than half the tasks was attributed to discrimination (see Appendix D.3), suggests motivated beliefs are an important micro-foundation of retaliatory discrimination. This will be discussed more in Section 5.

This experiment also illustrates how the dynamic retaliatory discrimination may be misclassified as taste-based discrimination à la Becker (1957). Academics that were not informed about stage 1 classified the discrimination observed in stage 2 as taste-based discrimination. While the discrimination arising in Treatments 1, 2, and 4 can plausibly be attributed to taste-based discrimination, attributing the differential treatment effect between $T3$ and $\{T1, T2, T4\}$ would misclassify it as such. Furthermore, retaliatory discrimination generates *new* discriminators, rather than amplifying existing discriminators.

4 Mechanisms Experiment: USA

A subsequent online experiment with 639 American men was conducted on Prolific.[20] The experimental set-up mirrors that of the experiment in Uganda, except for four main deviations. Firstly, the task differs: following Gagnon et al. (2025), participants had to copy a randomly generated sequence of letters and numbers. Secondly, the nature of the discrimination differed: workers and managers either had distinctively White or Black names.[21] Thirdly, participants were both White and Black American men, and thus participants belonged to both the majority and minority group. Fourth, the allocation of the 8 tasks in

[20]Prolific has been used for several discrimination-related studies (Eyting, 2022; Miserocchi, 2023; Gagnon et al., 2025; Ruebeck, 2025), and the sample pool performs well compared to other samples (e.g. a lab setting Gupta et al. (2021)). The screening criteria used include: US nationals aged between 20 and 60 whose primary language is English and were born in the USA. Their gender and sex is man and male, respectively, and they had to have completed at least 20 previous studies, with an approval rate of at least 95%.

[21]Names were taken from Bertrand and Mullainathan (2004) and Kline et al. (2022).

stage 1 of the experiment were either favoring the participant, equally splitting the tasks, or favoring the other worker. Appendix F outlines the motivation for each of these design choices.

Otherwise, the experimental design mirrored that of the experiment in Uganda, with participants first being a worker before becoming a manager. As such, the experiment consisted of 6 treatment arms across which participants are randomized, as depicted in Figure 3. In treatment arms 1-3, participants have a coethnic manager in the first stage, while in treatment arms 4-6 the stage 1 manager is non-coethnic. Managerial allocations in the first stage of the experiment are again pre-determined based on pilot data, and either favor the other worker (T1, T4), split the tasks evenly between the two workers (T2, T5), or favor the participant (T3, T6). 49% of the participants are African American, while the rest are White, with an average age of 40 years. Characteristics of the participants are balanced across treatment arms (see Appendix Table A4).

[*Figure 3 here*]

4.1 Results: Racial Retaliatory Discrimination

Figure 4 presents the allocations of tasks to a non-coethnic worker in the second stage of the experiment. As in Figure 2, allocating four out of the eight tasks to the non-coethnic worker indicates no discrimination, following Definition 1. For five out of the six experimental arms, there was on average no discrimination in the allocations across the two workers ($p = 0.239 - 1.000$). However, in T4, where the participants were previously exposed to a non-coethnic manager that only allocated two out of the eight tasks to them, participants retaliated against another individual of the same ethnicity as their previous manager, by allocating statistically significantly fewer tasks (3.79, $p = 0.003$). This allocation differs statistically significantly from the treatment arm where their previous non-coethnic manager evenly allocated the tasks across the two workers (T4 vs. T5, $p = 0.008$), and the treatment arm where their previous manager was coethnic, but only gave them two of the eight tasks (T1 vs. T4, $p = 0.019$).

[*Figure 4 here*]

Interestingly, the retaliatory nature of discrimination is not symmetric for positive past interactions. In treatments T3 and T6, the stage 1 manager — who was either coethnic (T3) or non-coethnic (T6) — allocated six of the eight tasks to the participant in stage 1 of the experiment. Nevertheless, the subsequent allocations across the two workers do not differ compared to the treatment arms with no discrimination in the first stage ($p = 0.259$ and $p = 0.367$, respectively). As such, retaliatory discrimination is asymmetric in the negative and positive domain.

Given 49% of the sample were African American, we can look at how treatment effects differed between members of minority and majority groups. Decomposing Figure 4 indicates that retaliatory discrimination is particularly pronounced among African American men, compared to White men. While White men on average assign 3.95 tasks to a worker with a Black-sounding name in T4 ($p = 0.370$), African American workers allocate 3.62 tasks, substantially less than the setting of no discrimination ($p = 0.005$). The different allocations between White and Black participants in T4 is statistically significant ($p = 0.019$).

In line with the results in Section 3, we again observe that retaliatory discrimination is driven by an increase in the extensive margin of discrimination. The number of individuals discriminating (following Definition 1) increases from 5.0% to 15.7% in T4 ($p < 0.001$).; however, conditional on discriminating, individuals do not differ in their intensity of discrimination across T4 and the other treatments (2.35 vs. 2.45, $p = 0.650$). Further support comes from exploring heterogeneity by discriminatory attitudes: treatment effects are particularly pronounced for households with below-median discriminatory attitudes ($p = 0.023$, see Appendix Table A26). These two results suggest that, rather than increasing the level of discrimination among individuals with stronger discriminatory tastes, retaliatory discrimination increases non-pecuniary costs for individuals with lower discriminatory tastes, and thus has a stronger effect on the marginal discriminator.

4.2 Persistence of Retaliatory Discrimination

Thus far, I have documented an immediate retaliatory nature of ethnic discrimination through a lab-in-the-field experiment in Uganda, and an online experiment among American men. However, Hjort (2014) and Fisman et al. (2020) document persistent effects of past experiences on discriminatory behaviors. Therefore, I next look at the persistence of retaliatory

discrimination through two design choices embedded within the online experiment.

Firstly, in between the first stage (when the participant was a worker) and the second stage of the experiment (when the participant was a manager), half of the respondents were randomized to complete a real-effort task, while the other half completed the real-effort task after the second round.[22] As such, there is variation in the time between the two stages of the experiment. Controlling for whether participants first completed a real-effort task (which lasted $\sim$ 3 minutes) does not affect the magnitude or statistical significance of the treatment effect estimates, nor is the corresponding coefficient statistically significant (see Appendix Table A29). Furthermore, allocations across stage 2 workers in T4 do not differ depending on whether participants first complete a real-effort task ($p = 0.199$).

Secondly, a follow-up study took place one week after the initial study. In the first part of the follow-up study, participants were assigned the role of the manager, identical to stage 2. Participants again had to allocate eight tasks between a non-coethnic and coethnic worker. As such, this allows me to test whether the managerial allocation decisions in stage 1 a week ago still had an effect on the participant's own discriminatory behavior a week later. One week later, initial managerial allocations do not have a persistent effect on the participant's allocations across workers.[23] Participants who had previously been randomly assigned to T4 do not discriminate one week later ($p = 0.718$), and do not allocate tasks differentially compared to the other five treatment arms ($p = 0.536$). One explanation for this is the limited importance and salience of discrimination: discrimination was never made explicit (unlike Gagnon et al. (2025)) and discrimination-related income losses were a mere \$0.20. As such, the stakes may have been too low in order for an initial discriminatory act to have effects a week later. This is illustrated by the fact that no participant could correctly recall the name of their stage 1 manager and task allocation of the previous week, despite monetary incentives to do so.

As such, the persistence of retaliatory discrimination is an interesting question for future research. While empirical papers have documented the persistent effects of major events, such as riots, on discriminatory behaviors, this study's exogenously induced (perceived) discrimination — subtle and of limited monetary significance — does not result in

[22] The real-effort task is discussed more in Section 6.

[23] Attrition across the two weeks was 27.7%, but did not differ systematically across treatment arms ($F = 1.215$, $p = 0.300$).

persistent retaliation.

4.3 Alternative Explanations

In this subsection, I briefly rule out alternative mechanisms, including (inaccurate) statistical discrimination, norm violation, and reciprocity. The underlying Tables and Figures, as well as more detailed discussion that rules out further mechanisms (anger, in-group favoritism, preference for equality, and experimenter demand effects), are reserved for Appendix J.

In a separate online experiment, participants are randomly assigned across the six treatment arms of Figure 3. However, instead of a manager delegating tasks between two workers, the manager allocates money between the two players — a form of dictator game. As worker productivity does not matter, (inaccurate) statistical discrimination does not play a role (List, 2006). Appendix Figure A6 illustrates that the pattern documented in Figure 4 is replicated in the dictator game version of the experiment, ruling out accurate and inaccurate statistical discrimination as a mechanism.

I can rule out that the treatment effects are driven by norm violations, as a result of the asymmetry of treatments effects between T4 and {T1,T3,T6} in Figure 4. All four treatment arms have a first stage manager who violates the social norm of equal division of tasks - however differential treatment effects are nevertheless observed for T4. Furthermore, detailed beliefs were elicited from half of the participants; when asked to justify why they allocated their tasks, none of them mentioned that a social norm had previously been violated.

In-group favoritism can firstly be ruled out by looking at the *Computer* Manager treatment arms of the lab-in-the-field experiment with Eritrean refugees in Uganda: while the discrimination observed in T1 and T2 (following Definition 1) could be attributed to in-group favoritism, participants randomized into T3 still discriminate statistically significantly more than those randomized into T1 and T2. Secondly, 95.25% of American men believed that an even division of tasks was fair, in contrast to what one would expect if participants wanted coethnic workers to earn more money, and hence receive more tasks. Thirdly, no discrimination is documents in all treatment arms except for T4 ($p = 0.239 - 1.000$), in contrast to predictions of in-group favoritism.

Lastly, I can rule out reciprocity through a minimal group paradigm experiment (Tajfel, 1970). The study is identical to the online experiment of Figure 3, expect that participant's

group affiliation is arbitrarily determined (a Red, and Blue team). Participants are giving pseudo-names that are ethnicity-neutral, and hence this experimental design thus does not make the participant's ethnicity salient. No discrimination, or retaliatory discrimination is documented across the six treatment arms (see Appendix Figure A7). This is in contrast to predictions of theoretical models of reciprocity (Rabin, 1993). Finally, tit-for-tat reciprocity is further ruled out as a potential mechanism by illustrating that participants retaliate more strongly ($p = 0.080$) against their manager from stage 1 (direct retaliation), rather than an unrelated worker of the same ethnicity (retaliatory discrimination), see Appendix J.

5 Microfounding Retaliatory Discrimination: Memory, Preferences, or Beliefs?

Prior to starting the online experiment, I pre-registered four theoretical micro-foundations, and hence functional forms, for retaliatory discrimination: memory recall, social preferences, Bayesian updating, and motivated beliefs.[24] Extensions to the basic experiment outlined in Section 4 were designed to disentangle the underlying mechanisms. Support is found in favor of identity-specific social preferences and motivated beliefs, but neither alone can rationalize retaliatory discrimination.

5.1 Memory

Unlike Miserocchi (2023), who documents the role of biased memory recall on discriminatory behavior, I find no empirical support that the recall of memories affects retaliatory discrimination. While participants who had a non-coethnic stage 1 manager in the previous week were statistically significantly more likely to recall that their manager was non-coethnic during the follow-up survey a week later ($p = 0.004$), their recall of allocated tasks was statistically indistinguishable compared to participants whose previous manager was coethnic ($p = 0.468$). Furthermore, the recall of allocated tasks in the previous week had no impact on their subsequent allocation of tasks between a coethnic and non-coethnic worker, and hence discriminatory behavior, when they were the manager (see Appendix Table A21). This sug-

[24] The document can be accessed on the AEA RCT Registry (AEARCTR-0016047).

gests participants did not have distorted memories, and these memories did not impact their retaliatory discriminatory behavior.

Further evidence of the limited role of memories on retaliatory discrimination comes from the follow-up study one week later. After participants made allocation decisions between a coethnic and non-coethnic worker (akin to stage 2 of the previous week and Figure 3, they were shown ten rounds of managers allocating eight tasks across two workers. One worker always had a White-sounding name, while the other worker had a Black-sounding name. In five of the ten rounds, the manager had a White-sounding name, while in the other five rounds the manager had a Black-sounding name. Allocations of the managers — based on pilot data — were such that, on average, there was no discrimination by White or Black managers.[25] All participants were shown the same ten rounds, in a randomized order. After recalling the managerial allocations (with financial incentives to do so), participants were assigned the role of the manager to divide eight tasks between two workers. Mirroring stage 2 of the earlier experiments, one of the workers had a White-sounding name, and the other had a Black-sounding name. As such, this experimental design tests (i) for the participants' ability to recall past rounds, and (ii) whether this (biased) recall affects their discriminatory behavior when they are in a decision-making position and can thus discriminate.

Firstly, I find that participants more accurately recall allocations of tasks for rounds with a coethnic manager ($p = 0.068$, see Appendix Table A24 columns (1)-(2)), however this does not differ depending on whether the coethnic manager favored coethnic workers, or not.[26] On the intensive margin, participants do not differentially recall the number of tasks allocated to coethnic workers based on (their recall of) the manager's ethnicity ($p = 0.751$, see Appendix Table A24 Column (3)). Participants overstate allocations to a coethnic worker when (i) a coethnic manager favors a non-coethnic worker, and (ii) when a non-coethnic manager prefers the coethnic worker. Similarly, they understate allocations to a coethnic worker when (i) a coethnic manager favors a coethnic worker, and (i) a non-coethnic manager favors the non-coethnic worker (see Appendix Table A24 Column (4)).[27]

[25] See Appendix M for an overview of the allocations.

[26] On average, participants correctly recalled 40.32% of past rounds, with no statistically significant difference between White and Black participants ($p = 0.321$).

[27] There is a statistically significant correlation between participants' discrimination index and the intensive margin of task recall for White participants ($\rho = -0.110$, $p = 0.078$), but not for African American participants ($\rho = 0.058$, $p = 0.428$). This suggests that among White participants, those that had stronger discriminatory tendencies thought managers with a Black-sounding name discriminated more against White

Secondly, a participant's (biased) recall of the allocation of managers in previous rounds has no effect on their allocation of tasks (and hence discriminatory behavior) across the two workers when they become a manager ($p = 0.927$, see Appendix Table A25).

These two findings — (i) the absence of a systematically biased recall of past decisions by managers, and (ii) the null effect of past recall on current discriminatory behaviors — suggest that memories are not shaping retaliatory discrimination. One mitigating circumstance is that the task, copying a randomly generated sequence of letters and numbers, does not have a stereotype associated to it. It would be interesting to explore whether memory recall is biased in stereotype-related tasks (in line with Miserocchi (2023)), can trigger collective memory (Fouka and Voth, 2023), and whether this subsequently affects retaliatory discrimination.

5.2 Social Preferences

Social preferences, including distributional and belief-dependent preferences, are unable to rationalize the findings of Sections 3 and 4 that past experiences with one individual can affect future preferences about other, similar individuals. Identity-dependent social preferences can provide a micro-foundation for some of the documented results related to retaliatory discrimination, however have not yet been formalized. Furthermore, social preferences struggle to rationalize other findings, for example the effects of retaliatory discrimination on anticipated discrimination, discussed in Section 6.

Models of distributional social preferences represent individual's utility functions as being concerned with inequality aversion (Fehr and Schmidt, 1999), the individual's relative payoff standing (Bolton and Ockenfels, 2000), increasing social welfare (Charness and Rabin, 2002), and the trade-off between equity and efficiency (Andreoni and Miller, 2002; Fisman et al., 2007). However, in the experiments outlined in Sections 3 and 4, the managerial allocations across the two workers do not affect social welfare, efficiency, or the participant's relative payoff standing. Furthermore, non-equal allocations across workers in all treatment arms, particularly in the lab-in-the-field experiment in Uganda (Figure 2), are in contrast to inequality aversion predictions of Fehr and Schmidt (1999). Most importantly, the retaliatory discrimination documented in T3 of Figure 2 and T4 of Figure 4 *increase* inequality and

workers. This provides further support for the role of motivated beliefs, discussed below.

reduce efficiency.

We document increased retaliation if participants can retaliate against their original manager, compared to when they can retaliate against a different worker (see Appendix Table A28), which can be rationalized using traditional models of reciprocity (Rabin, 1993). However, in Figures 2 and 4, participants cannot retaliate against their initial manager, but against a worker of the same ethnicity as their initial manager. In order for this form of retaliation to be rationalized using distributional social preferences, individuals would need to have other regarding preferences (such as inequality aversion) that are group- or identity-specific (Akerlof and Kranton, 2000).

Chen and Li (2009) document that induced group identity affect social preferences, with participants being more altruistic with in-group players. The differential degree of retaliatory discrimination between T1 and T4 of Figure 4 ($p = 0.019$) — when the stage 1 manager was a non-coethnic vs. a coethnic manager — is in line with these findings. However, models of social preferences and group identity have not been extended such that individual actions are extrapolated to affect group-level social preferences, which is what this paper, and other studies on scapegoating (Bursztyn et al., 2022; Bauer et al., 2023), find.[28] As such, distributional social preferences that incorporate an individual's identity (Akerlof and Kranton, 2000) and hence link other-regarding preferences to their identity, and the actions of others with a shared identity, could help rationalize retaliatory discrimination. However, such a theoretical foundation does not yet exist.

A second strand of social preferences focuses on belief-dependent preferences, where beliefs about other player's intentions and kindness affect player's utility and subsequent behavior. Intentions-based reciprocity (Rabin, 1993), building on psychological game theory (Geanakoplos et al., 1989), assumes that the perceived fairness of another player's behavior affects the individual's desire to increase or decrease their payoffs, captured through a non-pecuniary fairness payoff. However, similar to distributional preferences, belief-dependent preferences do not extrapolate beyond the participant, towards other individuals with the same background or group identity. As such, while participants may perceive the initial managerial allocation (in T1 and T4 of Figure 3) as unfair, they do not have the opportunity to lower the manager's payoff. Instead, and in contrast to predictions of independent

[28]We find no retaliatory discrimination using a minimum group paradigm (see Appendix Table A7), suggesting one's real identity, rather than an exogenously imposed one, plays an important role.

intentions-based reciprocity, participants retaliate against an unrelated worker who shares the same identity as the initial manager.[29]

Identity-specific, belief-dependent preferences, where the (un)fairness of others' behaviors affect the individual's desire to increase or decrease payoffs of unrelated individuals of the same identity, has promise to micro-found retaliatory discrimination — however these theoretical models have not yet been formalized. For example, Section 7 illustrates that making the existence of future rounds more salient reduces retaliatory discrimination, which can be rationalized by participants having other regarding preferences that are linked to identity. However, the documented negative treatment effects of negative past experiences on future labor supply (discussed in Section 6) cannot be rationalized through social preferences as participants are not interacting with others.

5.3 Beliefs

Beliefs play an important role underlying retaliatory discrimination, particularly motivated beliefs.

To gain relevant insights, half of the participants in the online experiment were asked to state their beliefs throughout the experiment.[30] On average, participants wanted more than half of the tasks (5.34 tasks), and after learning the name (and hence the ethnicity) of their manager in the first stage of the experiment, participants who had a non-coethnic manager did not expect to receive fewer tasks compared to participants who had a coethnic manager ($p = 0.421$). Therefore, there was no ex-ante anticipated discrimination. However, participants wanted to receive slightly more tasks from a coethnic manager ($p = 0.159$), particularly among Black men ($p = 0.068$). Furthermore, participants thought it was fair to receive more tasks from a coethnic manager ($p = 0.090$), which is again driven by Black men ($p = 0.079$).

Prior to dividing eight tasks between two workers as a manager in second stage of the experiment, participants overwhelmingly believed that an even division of tasks is both fair (95.25%) and efficient (89.24%). Furthermore, 80.70% of participants believed other partic-

[29] Models of guilt aversion, self-image, and social image concerns similarly cannot rationalize the documented patterns in Sections 3 and 4.

[30] There are no differences between participants from whom beliefs were elicited versus not, see Appendix Table A5.

ipants would split the tasks evenly. We do not observe any difference between participants randomized into T4 and the other treatments in terms of beliefs about fair or efficient allocations, nor second-order beliefs. However we observe that, among African American men, those randomized into T4 on average believe that 3.88 tasks allocated to the non-coethnic worker is fair. This is significantly less than 4 tasks ($p = 0.090$) and differs from what African American men perceived as a fair allocation in the other five treatment arms ($p = 0.059$). Perceiving more tasks allocated to the non-coethnic worker as fair is positively correlated with actual allocation decisions, both for the whole sample ($\rho = 0.192$, $p < 0.001$), and for individuals randomized to T4 ($\rho = 0.368$, $p = 0.006$). This provides suggestive evidence that perceiving discrimination from a non-coethnic manager increases the belief that discrimination against a non-coethnic worker is fair, which consequentially increases actual subsequent discrimination.

The stage 1 manager's allocation can be directly connected to the participant's beliefs when they are the manager. The correlation between the number of tasks allocated to a non-coethnic worker that is deemed a fair allocation in the second stage, and the discrepancy between the number of tasks the participant expected and actually received in the first stage, is not significant across all treatment arms ($\rho = -0.0153$, $p = 0.786$). However in T4, this correlation is negative and statistically significant ($\rho = -0.312$, $p = 0.021$), indicating that the individuals in T4 are more likely to think it is fair to assign fewer tasks to the non-coethnic worker if they received fewer tasks than expected from their non-coethnic manager in the first stage.[31] This provides further support for the importance of beliefs in retaliatory discrimination.

An additional online experiment with two treatment arms conducted parallel to the main online experiment provides further insights into the role of beliefs. The experiment was conducted among a separate sample of White men. While participants in the *Status Quo* treatment arm were exposed to T4 of Figure 3, participants in the *Uncertain Manager* treatment arm were also allocated two out of the eight tasks in the first stage of the experiment, however prior to the allocation of tasks were told that with 50% probability their manager was White, and with 50% probability their manager was Black (signaled by the name). Subsequently, participants completed the two assigned tasks and proceeded onto the

[31] In T1, where the manager is a co-ethnic that assigns two tasks to the participant, the correlation is $\rho = -0.126$ ($p = 0.411$).

second stage of the experiment as a manager, dividing eight tasks between two workers: one White worker and one Black worker.

Compared to the *Status Quo* treatment, the *Uncertain Manager* treatment arm gives participants some moral wiggle room regarding the ethnicity of the manager in the first stage. As such, participants can selectively process the managerial allocations in stage 1 in line with their prior beliefs, giving rise to motivated beliefs that induce discrimination (Eyting, 2022). In between tasks being allocated in stage 1 and completed, participants in the *Uncertain Manager* treatment arm were asked with what probability they now thought their stage 1 manager was White or Black. While on average participants still believed there was a 50.89% probability that the stage 1 manager was Black based on the task allocation, there is substantial variation: only 57% of respondents' posterior beliefs equaled the prior probability of 50%.

There is no differential allocation of tasks to the worker with a Black-sounding name in the second stage across the *Status Quo* and *Uncertain Manager* treatment arms (3.91 vs. 3.86, $p = 0.510$). For both treatment arms, allocations are statistically significantly different from an even split of tasks ($p = 0.072$ and $p = 0.004$, respectively), indicating discrimination following Definition 1.

In the *Uncertain Manager* treatment arm, subjective posterior beliefs about the ethnicity of the manager in stage 1 strongly influenced their division of tasks in the second round. The greater the subjective belief that the stage 1 manager was Black, the fewer tasks they assigned to the Black worker in stage 2. The correlation is $\rho = -0.360$ and is highly significant ($p < 0.001$). Appendix Figure A4 plots the individuals' subjective posterior beliefs of the probability of the stage 1 manager being Black, and allocations to the Black worker in stage 2, documenting a negative relationship. This negative relationship is asymmetrically driven by retaliation against the Black worker in stage 1 when participants had a posterior probability greater than 50% that the stage 1 manager was Black.[32]

The insights from this experiment highlight the role of motivated beliefs. Bayesian updating would predict that participants do not update their beliefs about the background of the manager in stage 1 as a result of the allocation of tasks in the *Uncertain Manager* treatment arm. Subsequently, this should have no effect on their discriminatory behavior as

[32]This is in line with the documented asymmetry of retaliatory discrimination based on whether past experiences were positive or negative (see T4 and T6 of Figure 4.

a manager in stage 2. Motivated beliefs on the other hand predict that the moral wiggle room in the *Uncertain Manager* treatment arm allows participants to interpret the ambiguous data in line with their priors. The biased interpretation subsequently affects the participants' discriminatory behaviors in line with their motivated beliefs. This is precisely what I find.

Two more pieces of evidence are found in favor of motivated beliefs, rather than Bayesian updating of beliefs, from the initial online experiment (Figure 3). Firstly, we would expect symmetric updating of beliefs (and hence behaviors) as a result of being exposed to treatments where the manager in the first stage favors or discriminates against the participant in the case of Bayesian updating. However, we only observe significant effects of past experiences on future discriminatory beliefs and behavior in the negative domain (see T4 in Figure 4). Secondly, when we ask participants why they thought the stage 1 manager made their decision, we document a pattern in line with the fundamental attribution error theory of social psychology (Jones and Harris, 1967). Prior to the managerial allocation, 66% of participants expect the manager to allocate the task evenly, which is balanced across treatment arms (F-statistic= 0.279, $p = 0.924$). However, once the stage 1 manager has divided the tasks, justifications for the allocations differ. When the participant only receives two tasks from a non-coethnic manager, individuals cite the manager's ethnicity as a reason in 25.58% of cases. This drops to 16.22% when the manager is a coethnic. Conversely, when the participant receives six tasks from a non-coethnic manager, individuals cite efficiency gains as a reason in 14.58% of cases. This jumps to 27.27% when the manager is a coethnic. Hence individuals are more likely to cite ethnic discrimination when they receive fewer tasks from a non-coethnic manager, however attribute the reverse situation to efficiency gains when they stand to benefit from a coethnic manager. This is again in line with motivated beliefs, and in contrast to Bayesian updating.

Nevertheless, not all of the results can be rationalized using motivated beliefs. For example, Section 7 discusses how increasing the salience of future rounds of the game reduces retaliatory discrimination. This cannot be rationalized using motivated beliefs. Furthermore, motivated beliefs would predict that those with the strongest discriminatory tastes would retaliate the most, as past perceived discrimination would be in line with motivated priors; however Appendix Table A26 illustrates that treatment effects are larger among participants with below-median discriminatory tastes.

6 Applications of Retaliatory Discrimination

In this Section, I illustrate the importance of retaliatory discrimination through two applications. Firstly, I discuss how retaliatory discrimination can generate different policy implications than taste-based and statistical discrimination, with an experiment simulating the reversal of affirmative action policies. Secondly, using the experimental design of Section 4, I experimentally show that negative past experiences are a micro-foundation of anticipated discrimination, and discuss the equilibrium consequences of this for the labor market.

6.1 Retaliation and the Removal of Affirmative Action Policies

To illustrate how retaliatory discrimination can affect policy implications, I consider the removal of affirmative action policies. Affirmative action (AA) policies aim to increase minority representation (e.g. across universities, the workforce, or company boards), and are typically successful (Bagde et al., 2016; Bertrand et al., 2018; Ellison and Pathak, 2021). However, recently several countries and organizations have been removing affirmative action policies and Diversity, Equity, and Inclusion (DEI) programs.

Neither taste-based nor statistical discrimination predict that the introduction and removal of affirmative action policies would increase subsequent discrimination against minority workers, compared to a case where affirmative action policies never existed. Taste-based discrimination predicts that the hiring of minority workers will return to pre-AA levels after affirmative action policies are removed, as employer's tastes are exogenous and thus unaffected by the introduction and removal of the policy. The policy introduces an additional constraint to the utility maximization problem of the manager (in the form of a minimum quota of the number of minority workers hired), however once the policy is removed, this constraint is too. Statistical discrimination argues that the information asymmetry between majority and minority workers will not get worse as a result of affirmative action policies: compared with a scenario where affirmative action policies were not introduced (and subsequently removed), employers have employed weakly more minority workers as a result of the affirmative action policy, and hence the information asymmetry about group-level productivity has weakly decreased, reducing discrimination.

Contrary to taste-based and statistical discrimination, retaliatory discrimination predicts that the introduction and removal of affirmative action policies can amplify discrimina-

tion, by amplifying endogenous discriminatory tastes against minority workers. For example, 55% of White Americans in believed that discrimination exists against them (NPR, 2017), and 36% of White men state that DEI policies hurt them (Rachel Minkin, 2024).[33] As such, AA and DEI policies can increase the number of negative past experiences with minorities and in turn prejudice against minority workers can increase as a result of pro-minority policies.[34] When affirmative action policies get removed, discrimination from majority individuals against minorities may subsequently actually increase as a result of the introduction and removal of affirmative action policies.

To causally test the effects of the removal of affirmative action policies on discriminatory preferences, I conduct a separate experiment among American White men on Prolific. In particular, T4 of Figure 3 is repeated. As such, participants have a manager in the first stage of the game who has a Black-sounding name, and allocates six tasks to the worker with a Black-sounding name, and two tasks to the participants. The participant subsequently becomes the manager and allocates eight tasks between two workers: one with a White-sounding name, and one with a Black-sounding name.

Experimental variation is introduced in the description of the manager's decision in the first stage. In the *Status Quo* condition, participants receive the same instructions as in the experiment outlined in Section 4, where the motivation of the manager in the first stage is unknown. In the *Affirmative Action Removed* condition, participants were informed in the first round of the game that the manager's allocation of tasks across workers were influenced by affirmative action policies, which had been removed before the second round.[35] Within this experimental set-up, taste-based and statistical discrimination would not predict differences between the two treatment arms, while retaliatory discrimination would predict stronger

[33] In another example, the United States Supreme Court ruled 9-0 in favor of a heterosexual woman claimed she was denied a promotion as a result of her sexual orientation, a form of "reverse discrimination" (Guardian, 2025b). The White House also ordered all federal agencies to end any DEI programs as of January 2025 (House, 2025), and companies including Meta, Google, Amazon, and Disney have rolled back DEI policies (Guardian, 2025a).

[34] For example, NPR (2017) quotes a 68-year-old White man from Akron, Ohio; "If you apply for a job, they seem to give the blacks the first crack at it ... and, basically, you know, if you want any help from the government, if you're white, you don't get it. If you're black, you get it."

[35] The exact wording was: "Please note that the manager's allocation decisions are guided by an affirmative action policy, which aims to provide additional opportunities to ethnic minority workers," and "The affirmative action policy has been abolished and no longer applies to your allocation decisions. You are free to distribute the tasks as you see fit."

retaliation, and hence greater discrimination, as a result of the presence of affirmative action policies in the past.[36]

Table 1 presents the number of tasks allocated to the worker with a Black-sounding name in the second stage of the experiment. The *Status Quo* group replicates the treatment effects documented in the original experiment ($p = 0.622$), and is statistically significantly different from case where tasks are split equally and hence there is no discrimination ($p = 0.072$).

When the unfavorable allocation of tasks in the first stage can be attributed to affirmative action policies that are subsequently abolished, the White manager retaliates against the worker with the Black-sounding name in the second stage of the experiment. In particular, they allocate 0.17 fewer tasks, equaling 0.32 standard deviations of the number of tasks allocated in the *Status Quo* treatment arm, equivalent to a 72% increase in discrimination following Definition 1. The difference between the two treatment groups is statistically significant ($p = 0.085$), while the one-sided t-test — effectively testing the predictions of retaliatory vs. taste-based and statistical discrimination — is highly significant ($p = 0.042$).

[*Table 1 here*]

These findings suggest that retaliatory discrimination generates policy implications that are distinct from both taste-based and statistical discrimination models, as the removal of affirmative action policies can induce greater subsequent discrimination. While the experiment provides causal evidence in a controlled setting, the external validity of these results remains limited (Levitt and List, 2007). Further empirical work in field and quasi-experimental contexts is therefore required to assess whether similar dynamics emerge in real labor markets.

6.2 Micro-foundation for Anticipated Discrimination

A second application of retaliatory discrimination relates to anticipated discrimination, which occurs when individuals expect to be treated unfairly by others in the future as a result of their observable characteristics (Charness et al., 2020; Agüero et al., 2023; Aksoy et al., 2023;

[36] As no learning takes place between rounds 1 and 2, statistical discrimination does not predict any change in discrimination between the two treatments. The distaste parameter of taste-based discrimination is exogenous and hence is not affected by the past existence and removal of affirmative action policies or not.

Angeli et al., 2025). This can have consequences in the labor market, for example by reducing the effort exerted by job-seekers, hence turning labor market discrimination into a self-fulfilling prophecy. Nevertheless, little is understood about the formation of the underlying beliefs and expectations, and subsequently the micro-foundations of anticipated/expected discrimination.

Past experiences could not only inform own discriminatory preferences, as modeled in Section 2 and empirically shown in Sections 3 and 4, but it could also affect expectations of future discriminatory attitudes of others: negative past experiences with individuals of a certain group could also affect expectations about the degree of discrimination from other individuals of that group. This in turn can affect the desire to interact and work with members of that group, making discrimination self-fulfilling.

To test this, participants in the online experiment of Section 4 complete a real-effort task after stage 1: they are informed that they will have one minute to correctly enter as many sequences of randomly generated letters and numbers as possible. Their pseudo-name and number of correctly completed tasks will be shared with a manager who must then choose ten workers to engage in a work task where both the workers and the manager receive a piece-rate wage for every sequence correctly completed.[37] As such, the number of completed tasks is a measure of the effort the participant put into the "job application", and hence of their search effort.

The manager was always a non-coethnic. If participants think that the manager is a taste-based discriminator, no difference in the number of completed tasks (a proxy for effort) is expected across the six treatment arms of Figure 3. This is because tastes are exogenous, and hence, with rational expectations, ones expectations of other people's tastes are also exogenous, and unaffected by their previous experiences with managers of the same ethnicity. Statistical discrimination is based on the decision-maker (in this case, the manager) having imperfect information about the worker's productivity, and thus relying on group-level information. The manager's information asymmetry is the same across treatment arms, and hence the participant's beliefs about the degree of statistical discrimination by the manager is not expected to differ across treatment arms.[38] As such, neither taste-based nor

[37] This is akin to the "non-blind" treatment of Boring et al. (2025) and the "manager" arm of Ruebeck (2025), as the participant's ethnicity is revealed through their pseudo-name.

[38] Statistical discrimination in the reliability of the signal could be present, however this would not differ across treatment arms, and hence not result in differential treatment effects.

statistically discrimination would expect there to be a difference in the level of anticipated discrimination — and hence effort put into the "job application" — as a result of exogenously induced variation in past experiences with managers of the same and different ethnicity as the potential future manager.

Retaliatory discrimination on the other hand would predict that expectations of discriminatory tastes can be affected as a result of past experiences: negative past experiences with individuals of a certain group can increase an individual's expectations of prejudice among managers of the same group, and hence increase anticipated discrimination.[39]

Figure 5 presents the number of tasks participants completed in 60 seconds during the real-effort task. Participants who were randomly exposed to a non-coethnic, discriminatory manager — who is of the same ethnicity as the hiring manager — complete statistically significantly fewer tasks (T4 vs. rest, $p = 0.030$). The number of tasks completed by participants randomly assigned to T4 is statistically significantly less compared to individuals who had a coethnic manager that only gave them two tasks, as well as individuals who had a non-coethnic manager that divided the tasks evenly ($p = 0.025$ and $p = 0.098$, respectively).

This presents experimental evidence that past experiences with individuals of a certain group can affect ones desire to interact with individuals of the same group in the future, proxied through a real effort task.[40] We do not observe that this affects the participants' expectations of the likelihood of being hired before completing the task, as this does not differ between T4 and the other treatments ($p = 0.607$).[41] Nevertheless, participants update their expectations after completing the 60-second task: the correlation between the number of completed tasks and the updating of expectations of being chosen is $\rho = 0.358$ ($p < 0.001$), indicating that individuals who performed well update their beliefs upwards.[42]

[*Figure 5 here*]

[39] Appendix N develops a theoretical foundation for the effect of retaliatory discrimination on (expectations of) anticipated discrimination, and subsequent job search effort.

[40] I can rule out that treatment effects are driven by anger, as half of the participants complete the effort task before stage 2, while the other half complete it after stage 2. This exogenous variation in timing does not affect the number of completed tasks ($p = 0.442$).

[41] Expectations of being selected for the work task are very optimistic: on average, participants believe there is a 62% chance they will be chosen for the additional work task, while in reality the chances were less than 3%.

[42] This correlation is weaker in T4 than the other treatments ($\rho = 0.274$ vs. $\rho = 0.377$), suggesting that participants are more skeptical about their likelihood of being hired — even when they perform well in the "job application" — when they perceived discrimination from a previous non-coethnic manager.

Equilibrium Effects of Retaliatory Discrimination

The equilibrium effects of retaliatory discrimination being a micro-foundation for anticipated discrimination goes beyond the scope of this paper. Nevertheless, I outline the intuition: a scenario could arise where the manager may inaccurately statistically discriminate due to retaliatory discrimination affecting the reliability of signals sent by job-seekers: if the manager is exposed to a sufficiently large number of non-coethnic workers who had negative experiences with managers of the same background as the decision-maker themselves, they will observe that, on average, their productivity signal is lower. Based on these signals (referring to signal s from Section 2), the manager can form inaccurate beliefs about the true productivity of the two groups, resulting in inaccurate statistical discrimination (Bohren et al., 2025a). Different to LePage (2024), statistical discrimination does not arise as a result of learning-through-hiring, but rather due to differential representativeness of the productivity signal of true productivity across different groups of applicants due to their past experiences.

7 Mitigating Retaliatory Discrimination

Different sources of discrimination have different remedies, with Bohren et al. (2025a) highlighting the importance of accurately identifying the driver of discrimination to effectively design policy interventions to reduce discrimination. As Section 6 illustrated, policies can have different effects from the perspective of retaliatory discrimination, compared to taste-based and statistical discrimination. As such, retaliatory discrimination also presents new opportunities aimed at mitigating discrimination.

This Section presents one mitigating action for which I provide causal empirical support that it can reduce the degree of retaliatory discrimination, and hence overall discrimination: increasing the salience of future interactions.[43]

[43] I pre-registered two other mitigating measures: costly mistakes, and inefficiencies due to non-even allocation of tasks. Appendix Tables A22 and A23 illustrate that neither helped mitigate retaliatory discrimination.

Future Interactions

In contrast to taste-based discrimination (Becker, 1957), retaliatory discrimination argues that current discriminatory preferences and hence behaviors are affected by past interactions. By taking into account the repeated nature of interactions and hence the evolution of discriminatory preferences and behaviors, individual's discriminatory decisions can be modeled as a repeated prisoners dilemma: while discriminating may be privately beneficial in round t, doing so could punish the individual (or other individuals with the same identity) in the future, if the player that is discriminated against in time t retaliates in future time periods.[44] As such, cooperation — in the form of no discrimination — is more likely to emerge if interactions take place over multiple rounds, or players are made aware of the existence of future rounds (Fudenberg and Maskin, 1986; Bó, 2005).

Modeling discrimination as a repeated prisoners dilemma where players adopt a grim trigger strategy of "always discriminating in rounds $t+i$, $i > 0$" when they are discriminated against in round t can sustain an equilibrium of "no discrimination".[45] This is in contrast to predictions of taste-based and statistical/experience-based discrimination: both of the workhorse models of discrimination's predictions are unaffected by whether the game is a one-period game or played over multiple periods.[46]

To experimentally investigate whether making participants aware of the existence of future rounds affects the extent to which they engage in discrimination, I run an additional set of stages in the online experiment. After the memory recall exercise (discussed in Section 5), all participants of the online experiment are assigned the role of one of the two workers. Participants have a non-coethnic manager, and participants are given two of the eight tasks. As such, all participants are exposed to T4 of Figure 3. Afterwards, participants become the manager and allocate tasks between two workers.

I induce experimental variation by randomizing participants across different treatment arms: the *Status Quo*, and the *Future Rounds* treatment arm. The only difference between the two treatment arms is that after participants are told that "This is the final round for

[44] For this to be important in an individual's decision to discriminate or not, individuals need to derive utility from their identity (Akerlof and Kranton, 2000), payoff of coethnic workers (Hjort, 2014), or group-specific altruistic preferences (Fehr and Schmidt, 1999; Chen and Li, 2009).

[45] See Appendix O for the mathematical foundations of this model.

[46] An exception is if participants learn about worker productivity in between rounds (LePage, 2024), which is not the case here.

you", participants in the *Future Rounds* treatment arm are informed that "there will be future rounds for the other two players, where the two workers you allocate the tasks across will become managers (and hence make similar decisions to you)". As such, this treatment arm makes salient the fact that the participant's allocation decisions can have an effect on future (discriminatory) decisions of the affected workers.

If the participant does not care about the payoff of other players, or decisions beyond the present round of the experiment, the *Future Rounds* treatment will have no effect on discriminatory behavior. Similarly, models of taste-based and statistical discrimination do not predict that the *Future Rounds* treatment will affect the delegation of tasks between the two workers.

[*Table 2 here*]

Table 2 illustrates that increasing the salience of future rounds increases the number of tasks allocated to a non-coethnic worker by 0.17 tasks, equal to 0.24 standard deviations of the division of tasks in the *Status Quo* treatment arm.[47] Allocations across workers in the *Future Rounds* treatment arm are no longer discriminatory ($p = 0.300$), illustrating how highlighting future interactions can affect the discriminatory actions of individuals in the current period, contrary to predictions of taste-based and statistical discrimination models.

8 Conclusion

Discrimination is widespread, and individuals perceive this too: in 2021, 8.86 million people in the EU reported that they felt discriminated against at work (Eurostat, 2024). However, little is understood about how (perceived) discriminatory experiences affect future discriminatory preferences and behaviors.

Contrary to predictions of the two workhorse models of discrimination — taste-based and statistical discrimination — experiments in Uganda and the USA document heightened discrimination in response to (perceived) past discrimination. This new form of discrimination, retaliatory discrimination, is not driven by the recall of past memories, but rather is driven by motivated beliefs. When academics are unaware of the existence of previous

[47]No statistically significant differences are documented between White and Black participants, see Appendix Table A27.

rounds, this form of discrimination is frequently mis-interpreted as taste-based. This can have implications for policies, such as the removal of affirmative action policies, which I find leads to heightened discrimination in an online experimental setting — in contrast to predictions of taste-based and statistical discrimination. Given the importance of individual instances of discrimination on future paths of systemic discrimination (Bohren et al., 2025b), correctly identifying the root cause of discrimination is important to design more effective policies (Bohren et al., 2025a).

Retaliatory discrimination can also be a micro-foundation for anticipated discrimination, as I experimentally show in Section 6.2: individuals who perceive past discrimination from a manager of the same ethnicity as a potential future manager exert less effort in their job application. As such, repeated past discriminatory interactions can result in discrimination becoming self-fulfilling and managers statistically discriminating based on inaccurate signals received that do not correctly reflect the true ability of the non-coethnic workers that were previously discriminated against.

Retaliatory discrimination — for which Appendix K presents anecdotal evidence — combines the literatures on the role of past experiences on economic decisions (Giuliano and Spilimbergo, 2025; Malmendier and Wachter, 2024), identity economics (Akerlof and Kranton, 2000), and social preferences (Rabin, 1993; Fehr and Schmidt, 1999; Charness and Rabin, 2002). This presents an interesting ground for future research, offering a theoretical foundation akin to the one developed in Section 2, for the empirically documented microeconomic and macroeconomic relationship between inter-group tensions and economic performance, as well as the role of ethnic divisions on conflict and economic development.

9 Tables and Figures

M_0 Allocations (U_1 ; E_1)			
Computer Manager		Ugandan Manager	
(6 ; 2)	(4 ; 4)	(6 ; 2)	(4 ; 4)
T1	T2	T3	T4

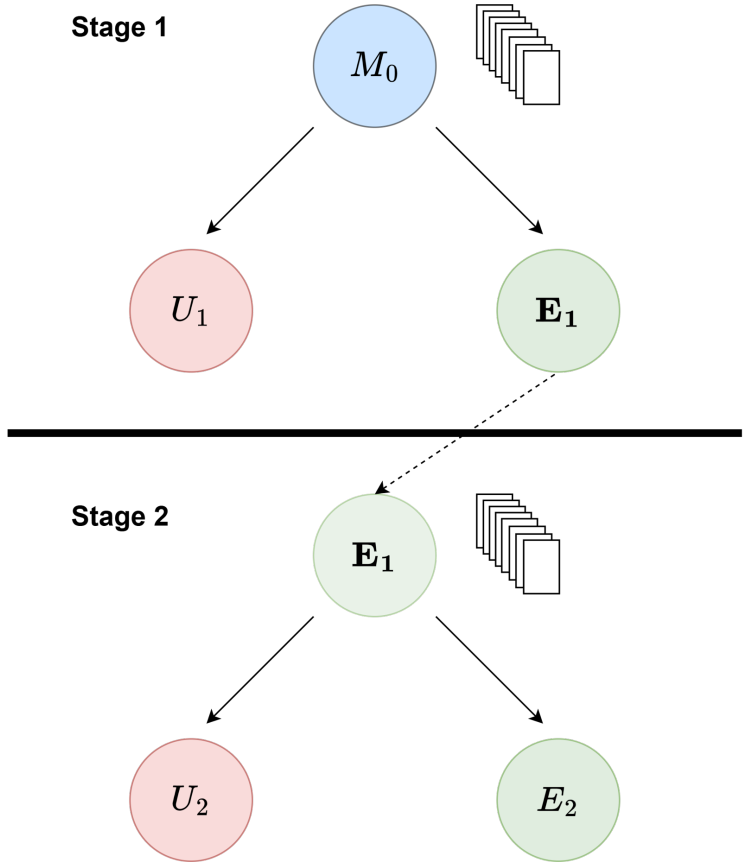


Figure 1. Experimental Design: Lab-in-the-Field in Uganda

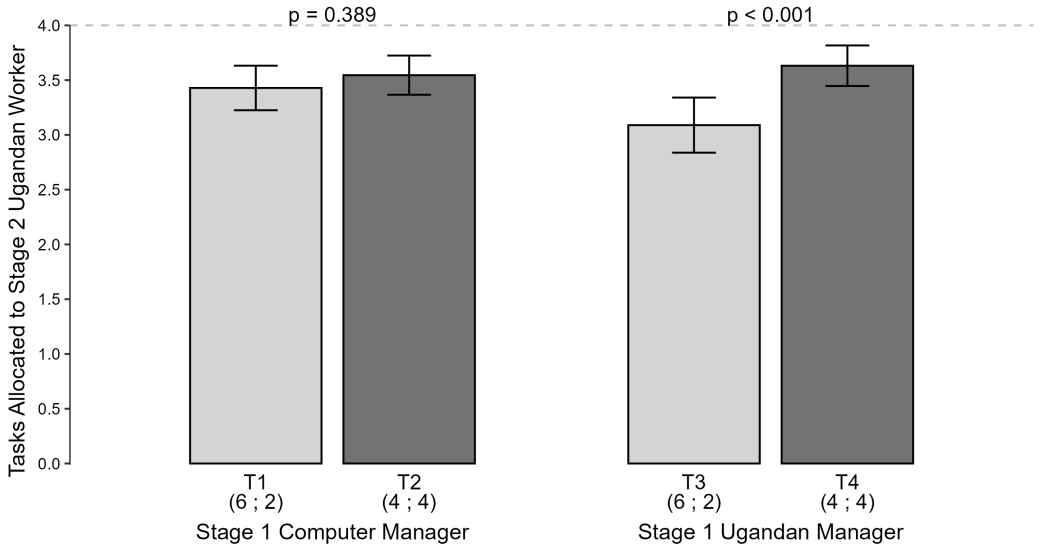


Figure 2. Task Allocation to Ugandan Worker in Stage 2 (U_2)

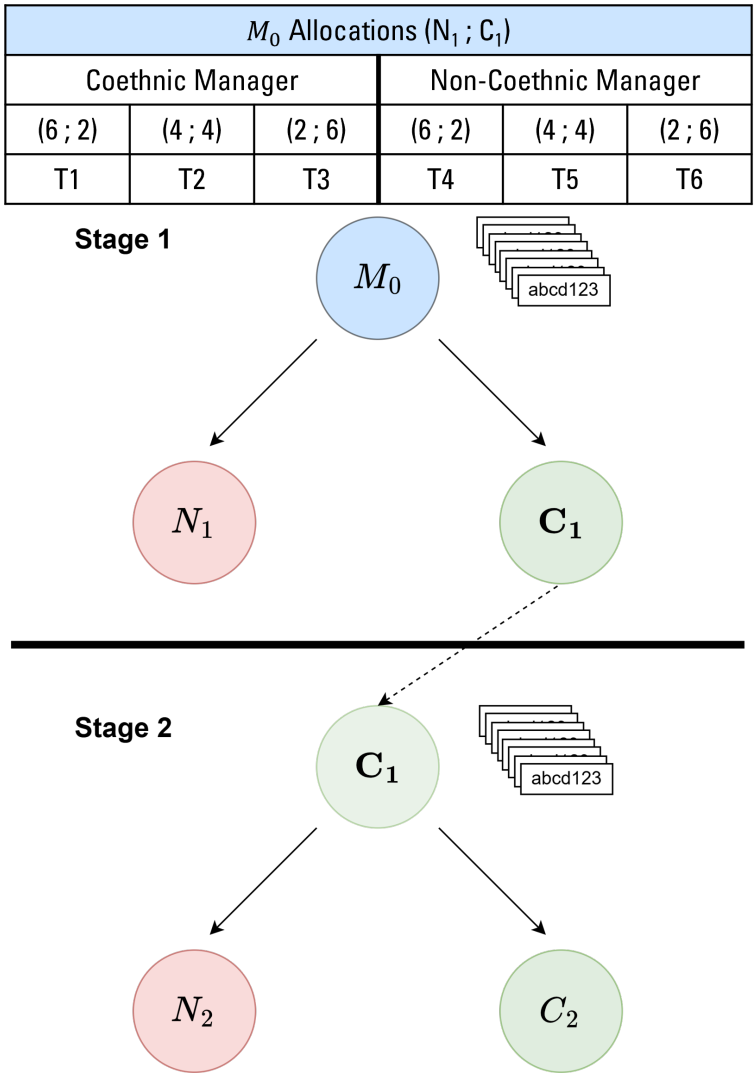


Figure 3. Experimental Design: Online Experiment in the USA

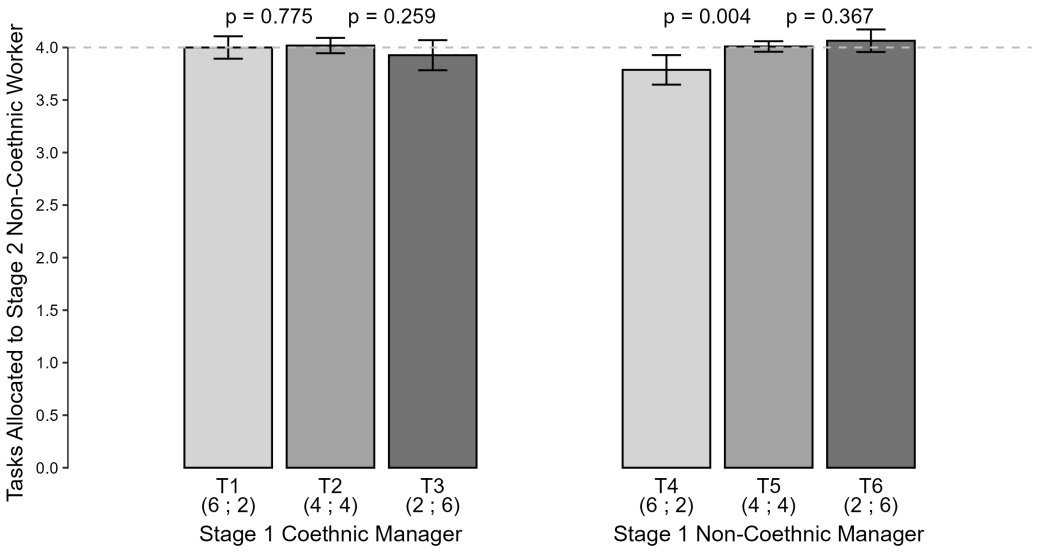


Figure 4. Task Allocation to Non-Coethnic Worker in Stage 2

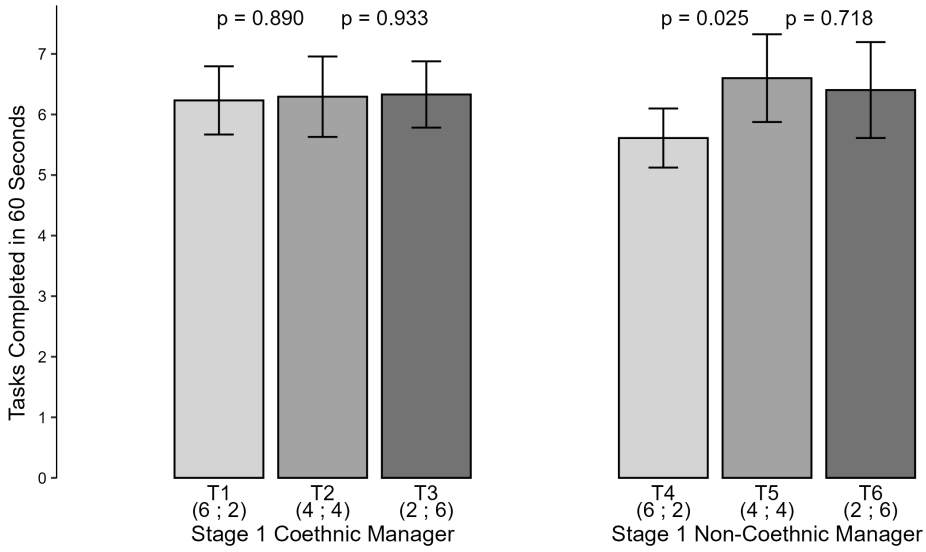


Figure 5. Tasks Completed in 60 Seconds as a Productivity Signal to a Non-Coethnic Manager

Table 1: Affirmative Action (AA) Removal and Discriminatory Allocations

	(1) Allocation of Tasks to Non-Coethnic Worker
Treatment: *AA Removal*	-0.17*
	(0.09)
Status Quo Mean	3.91
Status Quo S.D.	0.50
N	194

Notes: Intention to Treat estimates. The outcome variable is the number of tasks allocated to the Non-Coethnic worker by the participant in the second stage of the experiment, and ranges from 0 to 8. Control variables are selected using the post double LASSO machine learning algorithm outlined in Belloni et al. (2014). *AA Removal* refers to the treatment arm where the experimental instructions mentioned that Round 1 allocations were made under an affirmative action policy that was removed before stage 2. *Status Quo* mean and standard deviation refer to the mean value and standard deviation of the outcome in the treatment arm where the salience of future rounds was not made salient (and hence equivalent to T4 of Figure 3). Robust standard errors are in parentheses. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Table 2: Future Rounds and Discriminatory Allocations

	(1) Allocation of Tasks to Non-Coethnic Worker
Treatment: *Future Rounds*	0.17*
	(0.10)
Status Quo Mean	3.90
Status Quo S.D.	0.72
N	149

Notes: Intention to Treat estimates. The outcome variable is the number of tasks allocated to the Non-Coethnic worker by the participant in the second stage of the experiment, and ranges from 0 to 8. Control variables are selected using the post double LASSO machine learning algorithm outlined in Belloni et al. (2014). *Future Rounds* refers to the treatment arm where the experimental instructions heightened the salience of future rounds. *Status Quo* mean and standard deviation refer to the mean value and standard deviation of the outcome in the treatment arm where the salience of future rounds was not made salient (and hence equivalent to T4 of Figure 3). Robust standard errors are in parentheses. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

References

Agüero, J. M., Galarza, F., and Yamada, G. (2023). (Incorrect) Perceived Returns and Strategic Behavior among Talented Low-Income College Graduates. *AEA Papers and Proceedings*, 113:423–26.

Akerlof, G. A. and Kranton, R. E. (2000). Economics and Identity. *The Quarterly Journal of Economics*, 115(3):715–753.

Aksoy, B., Chadd, I., and Koh, B. H. (2023). Sexual identity, gender, and anticipated discrimination in prosocial behavior. *European Economic Review*, 154.

Alesina, A. and Ferrara, E. L. (2005). Ethnic Diversity and Economic Performance. *Journal of Economic Literature*, 43(3):762–800.

Andreoni, J. and Miller, J. (2002). Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism. *Econometrica*, 70(2):737–753.

Angeli, D., Matavelli, I., and Secco, F. (2025). Expected Discrimination and Job Search. Working paper.

Arbatlı, C. E., Ashraf, Q. H., Galor, O., and Klemp, M. (2020). Diversity and Conflict. *Econometrica*, 88(2):727–797.

Arnold, D., Dobbie, W., and Hull, P. (2022). Measuring Racial Discrimination in Bail Decisions. *American Economic Review*, 112(9):2992–3038.

Arrow, K. J. (1972a). Models of Job Discrimination. In Pascal, A. H., editor, *Racial Discrimination in Economic Life*, pages 83–102. D.C. Heath, Lexington, MA.

Arrow, K. J. (1972b). Some Mathematical Models of Race Discrimination in the Labor Market. In Pascal, A. H., editor, *Racial Discrimination in Economic Life*, pages 187–204. D.C. Heath, Lexington, MA.

Bagde, S., Epple, D., and Taylor, L. (2016). Does Affirmative Action Work? Caste, Gender, College Quality, and Academic Success in India. *American Economic Review*, 106(6):1495–1521.

Barlow, F. K., Paolini, S., Pedersen, A., Hornsey, M. J., Radke, H. R., Harwood, J., Rubin, M., and Sibley, C. G. (2012). The contact caveat: Negative contact predicts increased prejudice more

than positive contact predicts reduced prejudice. *Personality and Social Psychology Bulletin*, 38(12):1629–1643.

Bartoš, V., Bauer, M., Chytilová, J., and Matějka, F. (2016). Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition. *American Economic Review*, 106(6):1437–75.

Bauer, M., Cahlíková, J., Chytilová, J., Roland, G., and Želinský, T. (2023). Shifting Punishment onto Minorities: Experimental Evidence of Scapegoating. *The Economic Journal*, 133(652):1626–1640.

Becker, G. S. (1957). *The Economics of Discrimination*. University of Chicago Press.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). High-Dimensional Methods and Inference on Structural and Treatment Effects. *Journal of Economic Perspectives*, 28(2):29–50.

Belmonte, A. and Di Lillo, A. (2021). Backlash against affirmative action: Evidence from the South Tyrolean package. *European Economic Review*, 137:103802.

Benson, A. and Lepage, L.-P. (2024). Learning to Discriminate on the Job. Working Paper.

Bertrand, M., Black, S. E., Jensen, S., and Lleras-Muney, A. (2018). Breaking the Glass Ceiling? The Effect of Board Quotas on Female Labour Market Outcomes in Norway. *The Review of Economic Studies*, 86(1):191–239.

Bertrand, M. and Duflo, E. (2017). Chapter 8 - Field Experiments on Discrimination. In Banerjee, A. and Duflo, E., editors, *Handbook of Field Experiments*, volume 1 of *Handbook of Economic Field Experiments*, pages 309–393. North-Holland.

Bertrand, M. and Mullainathan, S. (2004). Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review*, 94(4):991–1013.

Bohren, J. A., Haggag, K., Imas, A., and Pope, D. G. (2025a). Inaccurate Statistical Discrimination: An Identification Problem. *The Review of Economics and Statistics*, pages 1–16.

Bohren, J. A., Hull, P., and Imas, A. (2025b). Systemic Discrimination: Theory and Measurement. *The Quarterly Journal of Economics*.

Bolton, G., Katok, E., and Zwick, R. (1998). Dictator Game Giving: Rules of Fairness Versus Acts of Kindness. *International Journal of Game Theory*, 27:269–299.

Bolton, G. E. and Ockenfels, A. (2000). ERC: A Theory of Equity, Reciprocity, and Competition. *American Economic Review*, 90(1):166–193.

Bordalo, P., Burro, G., Coffman, K., Gennaioli, N., and Shleifer, A. (2024). Imagining the Future: Memory, Simulation, and Beliefs. *The Review of Economic Studies*.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2020). Memory, Attention, and Choice. *The Quarterly Journal of Economics*, 135(3):1399–1442.

Boring, A., Coffman, K., Glover, D., and Gonzalez-Fuentes, M. J. (2025). Discrimination, Rejection, and Job Search. Working Paper.

Buchmann, N., Meyer, C., and Sullivan, C. D. (2024). Paternalistic Discrimination. Working paper.

Bulte, E., List, J. A., and van Soest, D. (2020). Toward an Understanding of the Welfare Effects of Nudges: Evidence from a Field Experiment in the Workplace. *The Economic Journal*, 130(632):2329–2353.

Bursztyn, L., Chaney, T., Hassan, T. A., and Rao, A. (2024). The Immigrant Next Door. *American Economic Review*, 114(2):348–84.

Bursztyn, L., Egorov, G., Haaland, I., Rao, A., and Roth, C. (2022). Scapegoating during Crises. *AEA Papers and Proceedings*, 112:151–55.

Bó, P. D. (2005). Cooperation under the Shadow of the Future: Experimental Evidence from Infinitely Repeated Games. *American Economic Review*, 95(5):1591–1604.

Cain, G. G. (1986). Chapter 13 - The economic analysis of labor market discrimination: A survey. In *Handbook of Labor Economics*, volume 1, pages 693–785. Elsevier.

Carlana, M. (2019). Implicit Stereotypes: Evidence from Teachers' Gender Bias. *The Quarterly Journal of Economics*, 134(3):1163–1224.

Chan, A. (2025). Discrimination Against Doctors: A Field Experiment. Working paper.

Charness, G., Cobo-Reyes, R., Meraglia, S., and Ángela Sánchez (2020). Anticipated Discrimination, Choices, and Performance: Experimental Evidence. *European Economic Review*, 127:103473.

Charness, G. and Rabin, M. (2002). Understanding Social Preferences with Simple Tests. *The Quarterly Journal of Economics*, 117(3):817–869.

Chen, Y. and Li, S. X. (2009). Group Identity and Social Preferences. *American Economic Review*, 99(1):431–57.

Coate, S. and Loury, G. (1993). Will Affirmative-Action Policies Eliminate Negative Stereotypes? *American Economic Review*, 83(5):1220–40.

de Quidt, J., Haushofer, J., and Roth, C. (2018). Measuring and Bounding Experimenter Demand. *American Economic Review*, 108(11):3266–3302.

DeZIM (2023). Racist Realities: How does Germany deal with racism? Preliminary study for the german national monitoring of discrimination and racism (nadira), German Centre for Integration and Migration Research (DeZIM), Berlin.

Ellison, G. and Pathak, P. A. (2021). The Efficiency of Race-Neutral Alternatives to Race-Based Affirmative Action: Evidence from Chicago's Exam Schools. *American Economic Review*, 111(3):943–75.

Eurostat (2024). Self-perceived discrimination at work - statistics.

Eyting, M. (2022). Why Do We Discriminate? The Role of Motivated Reasoning. Working Paper —, JGU Mainz & Stanford University. Working Paper.

Fehr, E. and Schmidt, K. M. (1999). A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics*, 114(3):817–868.

Fiorin, S., Hall, J., and Kanz, M. (2025). Discrimination Expectations in the Credit Market: Survey Evidence from India. *AEA Papers and Proceedings*, 115. forthcoming.

Fisman, R., Kariv, S., and Markovits, D. (2007). Individual Preferences for Giving. *American Economic Review*, 97(5):1858–1876.

Fisman, R., Sarkar, A., Skrastins, J., and Vig, V. (2020). Experience of Communal Conflicts and Intergroup Lending. *Journal of Political Economy*, 128(9):3346–3375.

Fouka, V. and Voth, H.-J. (2023). Collective remembrance and private choice: German–greek conflict and behavior in times of crisis. *American Political Science Review*, 117(3):851–870.

Fudenberg, D. and Maskin, E. (1986). The Folk Theorem in Repeated Games with Discounting or with Incomplete Information. *Econometrica*, 54(3):533–554.

Gagnon, N., Bosmans, K., and Riedl, A. (2025). The Effect of Gender Discrimination on Labor Supply. *Journal of Political Economy*, 133(3):1047–1081.

Gallup (2021). One in Four Black Workers Report Discrimination at Work.

Geanakoplos, J., Pearce, D., and Stacchetti, E. (1989). Psychological games and sequential rationality. *Games and Economic Behavior*, 1(1):60–79.

Ghosh, A. (2025). Religious Divisions and Production Technology: Experimental Evidence from India. *Journal of Political Economy*, 133(10):3249–3304.

Giuliano, L., Levine, D. I., and Leonard, J. (2009). Manager Race and the Race of New Hires. *Journal of Labor Economics*, 27(4):589–631.

Giuliano, P. and Spilimbergo, A. (2025). Aggregate Shocks and the Formation of Preferences and Beliefs. *Journal of Economic Literature*, 63(2):542–97.

Glover, D. (2019). Job Search under Discrimination: Evidence from Terrorist Attacks in France. Working paper.

Glover, D., Pallais, A., and Pariente, W. (2017). Discrimination as a Self-Fulfilling Prophecy: Evidence from French Grocery Stores. *The Quarterly Journal of Economics*, 132(3):1219–1260.

Guardian, T. (2025a). Rollback on diversity policies 'risks undoing decades of progress', says Co-op. *The Guardian*.

Guardian, T. (2025b). US supreme court sides with heterosexual woman in 'reverse discrimination' case. *The Guardian*.

Gupta, N., Rigotti, L., and Wilson, A. (2021). The experimenters' dilemma: Inferential preferences over populations.

Hellerstein, J. K. and Neumark, D. (2008). Workplace Segregation in the United States: Race, Ethnicity, and Skill. *The Review of Economics and Statistics*, 90(3):459–477.

Hjort, J. (2014). Ethnic Divisions and Production in Firms. *The Quarterly Journal of Economics*, 129(4):1899–1946.

House, T. W. (2025). Ending Radical and Wasteful Government DEI Programs and Preferencing. https://www.whitehouse.gov/presidential-actions/2025/01/ending-radical-and-wasteful-government-dei-programs-and-preferencing/. Presidential Action.

Jones, E. E. and Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology*, 3(1):1–24.

Kahana, M. (2012). *Foundations of Human Memory*. OUP USA.

Kahneman, D. (2011). Thinking, fast and slow. *Farrar, Straus and Giroux*.

Kaushal, N., Kaestner, R., and Reimers, C. (2007). Labor Market Effects of September 11th on Arab and Muslim Residents of the United States. *Journal of Human Resources*, XLII(2):275–308.

Kline, P., Rose, E. K., and Walters, C. R. (2022). Systemic Discrimination Among Large U.S. Employers. *The Quarterly Journal of Economics*, 137(4):1963–2036.

Kőszegi, B. and Rabin, M. (2006). A Model of Reference-Dependent Preferences. *The Quarterly Journal of Economics*, 121(4):1133–1165.

Lang, K. and Kahn-Lang Spitzer, A. (2020). Race Discrimination: An Economic Perspective. *Journal of Economic Perspectives*, 34(2):68–89.

Lang, K. and Lehmann, J.-Y. K. (2012). Racial Discrimination in the Labor Market: Theory and Empirics. *Journal of Economic Literature*, 50(4):959–1006.

Lanzara, G., Lazzaroni, S., and Masella, P. (2025). The Dynamics of Discrimination and Assimilation: Theory and Evidence. Working paper.

Leibbrandt, A., Wang, L. C., and Foo, C. (2018). Gender Quotas, Competitions, and Peer Review: Experimental Evidence on the Backlash Against Women. *Management Science*, 64(8):3501–3516.

LePage, L.-P. (2024). Experience-Based Discrimination. *American Economic Journal: Applied Economics*, 16(4):288–321.

Levitt, S. D. and List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, 21(2):153–174.

Lickel, B., Miller, N., Stenstrom, D. M., Denson, T. F., and Schmader, T. (2006). Vicarious Retribution: The Role of Collective Blame in Intergroup Aggression. *Personality and Social Psychology Review*, 10(4):372–390.

List, J. A. (2006). The Behavioralist Meets the Market: Measuring Social Preferences and Reputation Effects in Actual Transactions. *Journal of Political Economy*, 114(1):1–37.

Loiacono, F. and Silva Vargas, M. (2025). Matching with the right attitude: How job-seekers' beliefs shape search and hiring outcomes. Working Paper.

Loiacono, F. and Vargas, M. S. (2019). Improving access to labor markets for refugees: Evidence from uganda. Working paper, International Growth Centre.

Lowe, M. (2025). Has intergroup contact delivered? *Annual Review of Economics*, 17(Volume 17, 2025):321–344.

Lu, R. and Sheng, S. Y. (2022). How racial animus forms and spreads: Evidence from the coronavirus pandemic. *Journal of Economic Behavior & Organization*, 200:82–98.

Luca, M., Pronkina, E., and Rossi, M. (2024). The evolution of discrimination in online markets: How the rise in anti-asian bias affected airbnb during the pandemic. *Marketing Science*.

Malmendier, U. (2021). FBBVA Lecture 2020 Exposure, Experience, and Expertise: Why Personal Histories Matter in Economics. *Journal of the European Economic Association*, 19(6):2857–2894.

Malmendier, U. and Wachter, J. A. (2024). Memory of Past Experiences and Economic Decisions. In *The Oxford Handbook of Human Memory, Two Volume Pack: Foundations and Applications*. Oxford University Press.

Miserocchi, F. (2023). Discrimination through Biased Memory. Working paper.

Montoya, A. M., Parrado, E., Solis, A., and Undurraga, R. (2025). Bad Taste: Gender Discrimination in Consumer Lending. *Journal of Political Economy Microeconomics*.

Neschen, A. and Hügelschäfer, S. (2021). Gender bias in performance evaluations: The impact of gender quotas. *Journal of Economic Psychology*, 85:102383.

Neumark, D. (2018). Experimental Research on Labor Market Discrimination. *Journal of Economic Literature*, 56(3):799–866.

NPR (2017). Majority Of White Americans Say They Believe Whites Face Discrimination.

Paolini, S., Gibbs, M., Sales, B., Anderson, D., and McIntyre, K. (2024). Negativity bias in intergroup contact: Meta-analytical evidence that bad is stronger than good, especially when people have the opportunity and motivation to opt out of contact. *Psychological Bulletin*.

Paolini, S., Harwood, J., and Rubin, M. (2010). Negative intergroup contact makes group memberships salient: Explaining why intergroup conflict endures. *Personality and social Psychology bulletin*, 36(12):1723–1738.

Petters, L. M. and Schröder, M. (2020). Negative side effects of affirmative action: How quotas lead to distortions in performance evaluation. *European Economic Review*, 130:103500.

Phelps, E. S. (1972). The Statistical Theory of Racism and Sexism. *American Economic Review*, 62(4):659–661.

Rabin, M. (1993). Incorporating Fairness into Game Theory and Economics. *The American Economic Review*, 83(5):1281–1302.

Rabin, M. and Schrag, J. L. (1999). First Impressions Matter: A Model of Confirmatory Bias. *The Quarterly Journal of Economics*, 114(1):37–82.

Rachel Minkin (2024). Views of DEI have become slightly more negative among U.S. workers.

Ruebeck, H. (2025). Causes and Consequences of Perceived Workplace Discrimination. Working Paper.

Sarsons, H. (2022). Interpreting Signals in the Labor Market: Evidence from Medical Referrals. Revise and Resubmit, *The Review of Economic Studies*.

Schindler, D. and Westcott, M. (2020). Shocking Racial Attitudes: Black G.I.s in Europe. *The Review of Economic Studies*, 88(1):489–520.

Shayo, M. and Zussman, A. (2017). Conflict and the Persistence of Ethnic Bias. *American Economic Journal: Applied Economics*, 9(4):137–65.

Tajfel, H. (1970). Experiments in Intergroup Discrimination. *Scientific American*, 223(5):96–103.

Thaler, M. (2024). The Fake News Effect: Experimentally Identifying Motivated Reasoning Using Trust in News. *American Economic Journal: Microeconomics*, 16(2):1–38.

Welle, D. (2025). Racism in Germany is the norm, not the exception.

Wicker, T. (2025). Winsorizing and Trimming with Subgroups. Revise and Resubmit, *Journal of Development Economics*.

Wicker, T., Dalton, P., and van Soest, D. (2025). Mental Accounting and Cash Transfers: Experimental Evidence from a Humanitarian Setting. Working paper, CentER, Center for Economic Research. CentER Discussion Paper Nr. 2025-006.

Yamagishi, T., Horita, Y., Mifune, N., Hashimoto, H., Li, Y., Shinada, M., Miura, A., Inukai, K., Takagishi, H., and Simunovic, D. (2012). Rejection of unfair offers in the ultimatum game is no evidence of strong reciprocity. *Proceedings of the National Academy of Sciences of the United States of America*, 109.

Appendix to

Discrimination as Retaliation: Past Experiences and the Dynamics of Discrimination

by Till Wicker

A General Theoretical Model of Discrimination

This section presents a general model of discrimination that abstracts from the labor market model of Section 2 and applies to a broad set of decision-making contexts, including lending, tenant selection, grading, police search/enforcement intensity, or allocation decisions. The model incorporates taste-based, statistical, and retaliatory discrimination within a single framework.

A decision-maker (DM) repeatedly interacts with individuals indexed by $i \in I$ who belong to observable groups $g \in \{A, B\}$. At each time t, the DM observes an individual's group identity g and a noisy signal $s_{i,t}$ of the individual's latent quality $\theta_{i,t}$ (e.g., productivity, creditworthiness, intelligence). The DM chooses an action $a_{i,t} \in \mathcal{A}$ (e.g., hire, admit, lend, grade) to maximize expected utility, which consists of two components:

1. Expected material payoff $\Pi_t(a_{i,t}, \theta_{i,t}, g)$ from action $a_{i,t}$, and
2. Non-pecuniary costs $f(d_g, F(\chi_{g,t}))$ associated with interacting with members of group g.

Hence, the DM's problem at time t is:

$$\max_{a_{i,t} \in \mathcal{A}} \mathbb{E}\big[\Pi_t(a_{i,t}, \theta_{i,t}, g) \,\big|\, s_{i,t}, g\big] - f(d_g, F(\chi_{g,t}))\,.$$

The latent trait $\theta_{i,t}$ is drawn from a group-specific distribution: $\theta_{i,t} \sim N(\mu_g, 1/\tau_g)$, and the DM observes a signal $s_{i,t} = \theta_{i,t} + \varepsilon_{i,t}, \quad \varepsilon_{i,t} \sim N(0, 1/\eta_g)$. The DM holds subjective beliefs about each group's latent quality distribution and signal precision, summarized by $\psi_g \equiv (\hat{\mu}_g, \hat{\tau}_g, \hat{\eta}_g)$. After observing $(s_{i,t}, g)$, the DM forms posterior beliefs about $\theta_{i,t}$ following Bayes' rule. These beliefs determine the expected material payoff in the DM's problem at time t. Differences in ψ_g across groups generate statistical discrimination.

The term $f(d_g, F(\chi_{g,t}))$ captures group-specific non-pecuniary (psychological or social) costs of interacting with individuals of group g. It has two components:

1. A static "taste" parameter d_g representing time-invariant preferences or distastes toward group g (taste-based discrimination).

2. A time-varying component $F(\chi_{g,t})$ that depends on the DM's cumulative past experiences with members of group g at time t (retaliatory discrimination).

The function $f(\cdot)$ is weakly increasing and concave in both arguments:

$$\frac{\partial f}{\partial d_g} \geq 0, \quad \frac{\partial f}{\partial F(\chi_{g,t})} \geq 0, \quad \frac{\partial^2 f}{\partial d_g^2} \leq 0, \quad \frac{\partial^2 f}{\partial F(\chi_{g,t})^2} \leq 0.$$

Hence, more negative past experiences with group g raise the cost of engaging favorably with that group, but at a decreasing rate. Similarly, stronger discriminatory tastes (d_g) increase the cost of engaging with individuals of group g.

Definition of Discrimination Next, we define the DM's expected allocation or treatment toward group g conditional on a given signal s as $\Gamma_t(s \mid g, \psi_g)$. This can for example be the value of a loan given (Fisman et al., 2020), or a teacher's recommendation for the future school track of a student (Miserocchi, 2023). Then, discrimination at time t is:

$$D_t(s, \psi_A, \psi_B) \equiv \Gamma_t(s \mid A, \psi_A) - \Gamma_t(s \mid B, \psi_B).$$

Discrimination occurs when $D_t(s, \psi_A, \psi_B) \neq 0$. The DM discriminates against individuals of group B if $D_t(s, \psi_A, \psi_B) > 0$ and against individuals of group A if $D_t(s, \psi_A, \psi_B) < 0$.

Incorporating Other Models of Discrimination The DM's problem nests the canonical models of discrimination:

1. **Taste-based discrimination (Becker, 1957):** setting $f(d_g, F(\chi_{g,t})) = d_g$ yields an exogenous preference for or against group g.

2. **Statistical discrimination (Arrow, 1972a; Phelps, 1972):** setting $f(d_g, F(\chi_{g,t})) = 0$ but allowing $\psi_A \neq \psi_B$ yields group-dependent beliefs about θ, producing differential actions for identical signals. This also nests inaccurate statistical discrimination (Bohren et al., 2025a). Experience-based discrimination (LePage, 2024) can provide a micro-foundation for the emergence of statistical discrimination.

3. **Retaliatory discrimination:** allowing $f(d_g, F(\chi_{g,t}))$ to evolve with $F(\chi_{g,t})$ introduces endogenous preferences that amplify or attenuate discrimination over time, as a result of past experiences and interactions.

Predictions Without loss of generality, suppose the DM discriminates against group B such that $D_t(s, \psi_A, \psi_B) > 0$. The model yields two general predictions:

1. (*Retaliatory Discrimination*) More negative past experiences with individuals of group B increase discrimination against group B, ceteris paribus:

$$\chi^{\text{mod}}_{B,t} < \chi^{\text{neg}}_{B,t} \quad \Rightarrow \quad D_t(s, \psi_A, \psi_B \,|\, \chi^{\text{mod}}_{B,t}) \leq D_t(s, \psi_A, \psi_B \,|\, \chi^{\text{neg}}_{B,t}).$$

2. (**Group-Specific Retaliatory Discrimination**) Ceteris paribus, experiences with unrelated groups $g' \neq B$ do not affect discrimination toward B:

$$\chi^{\text{mod}}_{g',t} < \chi^{\text{neg}}_{g',t} \quad \Rightarrow \quad D_t(s, \psi_A, \psi_B \,|\, \chi^{\text{mod}}_{g',t}) = D_t(s, \psi_A, \psi_B \,|\, \chi^{\text{neg}}_{g',t}).$$

The proofs underlying the theoretical predictions follow directly from Appendix C.

This general model captures the evolution of discriminatory behavior as a function of both beliefs and preferences, providing a unified foundation for studying discrimination across diverse decision-making contexts.

B Extensions to Theory in Section 2

B.1 Paternalistic Discrimination (Buchmann et al., 2024)

To allow for paternalistic discrimination, the utility function of the employer also contains a fraction α_{eg} of the expected on-the-job welfare of the worker, where employers differ in whether they internalize their perception of the worker's perception of welfare, or internalize their own perception of workers' welfare. In particular, the utility function of the manager becomes:

$$\max_{L_{A,t},L_{B,t}} \underbrace{Y(L_{A,t},\theta_A,L_{B,t},\theta_B) - \sum_{g\in\{A,B\}} L_{g,t}w_g}_{\text{Firm Profit}} - \underbrace{\sum_{g\in\{A,B\}} L_{g,t}f\Big(d_g,F(e_{g,t})\Big)}_{\text{Non-Pecuniary Costs}} - \underbrace{\sum_{g\in\{A,B\}} L_{g,t}\alpha_{g,t}\mathcal{W}_{g,t}}_{\text{Other-regarding utility}}$$

where $\mathcal{W}_{g,t}$ is the manager's perception of the workers' perception of welfare, which is defined with respect to the outside option. The worker's welfare consists of their wage, and disutility of working ($\mathcal{W} = \mathbb{E}_i[w_g - u_g(c)]$). Paternalistic employers internalize their own perceptions of the worker's welfare.

B.2 Experience-based Discrimination (Lepage, 2024)

The employer's utility function remains as specified:

$$\max_{L_{A,t},L_{B,t}} Y(L_{A,t},\theta_A,L_{B,t},\theta_B) - \sum_{g\in\{A,B\}} L_{g,t}w_g - \sum_{g\in\{A,B\}} L_{g,t}f(d_g,F(e_{g,t}))$$

However, we now incorporate dynamic belief updating based on accumulated experiences with each group. In particular, at $t=0$, employers have prior beliefs about group g's productivity: $\hat{\theta}_{g,0} \sim N(\hat{\mu}_{g,0}, 1/\hat{\tau}_{g,0})$.

After each hiring decision, employers observe realized productivity $\theta_{g,i}$ for each hired worker i from group g. Following Bayesian updating combined with experience-based learning:

$$\begin{aligned}
\hat{\mu}_{g,t+1} &= \alpha_\mu \hat{\mu}_{g,t} + (1-\alpha_\mu)\left[\beta_g(e_{g,t})\cdot\overline{\theta}_{g,obs,t} + (1-\beta_g(e_{g,t}))\cdot\hat{\mu}_{g,t}\right] \\
\hat{\tau}_{g,t+1} &= \alpha_\tau \hat{\tau}_{g,t} + (1-\alpha_\tau)\left[\frac{H_{g,t}}{\text{var}(\theta_{g,obs,t})}\right]
\end{aligned}$$

where $\alpha_\mu, \alpha_\tau \in [0,1]$ are experience weights (higher values place more weight on past beliefs), $\beta_g(e_{g,t}) \in [0,1]$ is the experience-dependent learning rate from new observations, $H_{g,t}$ is the cu-

mulative number of workers hired from group g up to time t, and $\overline{\theta}_{g,obs,t} = \frac{1}{H_{g,t}} \sum_{i=1}^{H_{g,t}} \theta_{g,i}$ is the sample mean of observed productivity.

LePage (2024) illustrates that the learning rate itself depends on past experiences:

$$\beta_g(e_{g,t}) = \beta_0 \cdot \exp(-\gamma \cdot F(e_{g,t})) \quad \text{where } \gamma > 0, \beta_0 \in (0, 1]$$

This specification captures the psychological mechanism whereby negative experiences with respect to the productivity of hired workers make employers less receptive to contradictory information. As such, past experiences can have an effect on current discrimination through two channels:

1. Learning about group-level productivity, as a result of past experiences (Experience-based discrimination, LePage (2024))

2. Endogenously updating non-pecuniary costs of hiring workers from a specific group (Retaliatory discrimination)

B.3 Inaccurate Statistical Discrimination

The employer's utility function remains as specified in equation (1):

$$\max_{L_{A,t}, L_{B,t}} Y(L_{A,t}, \theta_A, L_{B,t}, \theta_B) - \sum_{g \in \{A,B\}} L_{g,t} w_g - \sum_{g \in \{A,B\}} L_{g,t} f(d_g, F(e_{g,t}))$$

Following Bohren et al. (2025a), we now explicitly distinguish between true and subjective productivity distributions:

True Productivity Distribution: Worker productivity for group g is drawn from $\theta_g \sim N(\mu_g, 1/\tau_g)$ with true signal precision η_g.

Subjective Beliefs: Employers hold potentially inaccurate subjective beliefs $\hat{\psi} \equiv (\hat{\mu}_g, \hat{\tau}_g, \hat{\eta}_g)$ about group g's productivity distribution and signal precision, where:

$$\hat{\theta}_g \sim N(\hat{\mu}_g, 1/\hat{\tau}_g) \tag{2}$$

$$\text{Subjective signal precision:} \quad \hat{\eta}_g \geq 0, \quad \hat{\eta}_g \neq \eta_g \tag{3}$$

Inaccurate Statistical Discrimination occurs when $(\hat{\mu}_g, \hat{\tau}_g, \hat{\eta}_g) \neq (\mu_g, \tau_g, \eta_g)$ for some group g.

Following Bohren et al. (2025a), employers make hiring decisions based on their subjective posterior beliefs. After observing worker's group identity g and signal s, the employer forms a

posterior belief using Bayes' rule with subjective distributions:

$$\hat{\mu}_{g,t}(s) = \frac{\hat{\tau}_g \hat{\mu}_g + \hat{\eta}_g s}{\hat{\tau}_g + \hat{\eta}_g} \tag{4}$$

B.4 Systemic Discrimination

Systemic discrimination captures how discrimination in other decisions indirectly contributes to disparities by affecting relevant attributes for a given decision, which in turn generates disparities in outcomes. This extension demonstrates how retaliatory discrimination at the focal node can coexist with systemic discrimination arising from other nodes in the decision system.

Following Bohren et al. (2025b), we embed the retaliatory discrimination model within a broader system of interconnected decision nodes. The system consists of a set of nodes $N \equiv \{1, ..., N\} \cup \{n^*\}$, where n^* represents the focal hiring node from our baseline model.
Each non-focal node $n = 1, ..., N$ represents a decision task where:

- Worker i has productivity $\theta\_i^n \in \Theta^n$ for task n
- An employer observes the worker's group G_i and signal $S_i^n \in \mathcal{S}^n$
- The employer selects action $A_i^n \in \mathcal{A}^n$ according to action rule $A^n(G_i, S_i^n)$. The action is to either hire the worker, or not.

At the focal node n^* (our baseline hiring decision):

- Worker productivity is $\theta_i^* \in \Theta^*$
- Signal is $S_i^* \in \mathcal{S}^*$
- Action is $A_i^* \in \mathcal{A}^*$

As discussed in Bohren et al. (2025b), actions at other nodes can affect productivity and signals at the focal node. For example, S_i^* may include performance evaluations from other nodes, or focal-node productivity θ_i^* may depend on training received at node n.

Incorporating this within the retaliatory discrimination model, the employer's utility function at the focal node becomes:

$$\max_{L_{A,t}, L_{B,t}} Y(L_{A,t}, \theta_A, L_{B,t}, \theta_B) - \sum_{g \in \{A,B\}} L_{g,t} w_g - \sum_{g \in \{A,B\}} L_{g,t} f(d_g, F(e_{g,t}))$$

However, the signal S_i^* and/or productivity Y_i^* now depend on actions at other nodes:

$$S_i^* = S^*(G_i, A_i^1, A_i^2, ..., A_i^N, \xi_i)\theta_i^* \qquad = \theta^*(G_i, A_i^1, A_i^2, ..., A_i^N, \zeta_i) \tag{5}$$

where ξ_i and ζ_i represent other factors affecting signals and productivity.

As such, retaliatory discrimination at one node can have consequences for future nodes, and hence lead to more, systematized discrimination.

C Theoretical Prediction Proofs

C.1 Prediction 1:

$$\chi_{B,t}^{\text{mod}} < \chi_{B,t}^{\text{neg}} \Longrightarrow D_t(s, \psi | \chi_{B,t}^{\text{mod}}) \leq D_t(s, \psi | \chi_{B,t}^{\text{neg}})$$

Proof:

From equation (1), the employer's utility maximization problem is:

$$\max_{L_{A,t}, L_{B,t}} Y(L_{A,t}, \theta_A, L_{B,t}, \theta_B) - \sum_{g \in \{A,B\}} L_{g,t} w_g - \sum_{g \in \{A,B\}} L_{g,t} f(d_g, F(e_{g,t})) \tag{C.1.1}$$

The first-order conditions with respect to $L_{A,t}$ and $L_{B,t}$ are:

$$\frac{\partial Y}{\partial L_{A,t}} - w_A - f(d_A, F(e_{A,t})) = 0 \tag{C.1.2}$$

$$\frac{\partial Y}{\partial L_{B,t}} - w_B - f(d_B, F(e_{B,t})) = 0 \tag{C.1.3}$$

From these conditions, the optimal hiring levels are implicitly defined as:

$$L_{A,t}^* = L_A^*(w_A, d_A, F(e_{A,t}), \theta_A, \theta_B) \tag{C.1.4}$$

$$L_{B,t}^* = L_B^*(w_B, d_B, F(e_{B,t}), \theta_A, \theta_B) \tag{C.1.5}$$

The discrimination measure is defined as:

$$D_t(s, \psi) = L_{A,t}^* | s - L_{B,t}^* | s \tag{C.1.6}$$

To prove Prediction 1, we need to show that:

$$\frac{\partial D_t}{\partial F(e_{B,t})} \geq 0 \tag{C.1.7}$$

From the implicit function theorem applied to equations (C.1.2) and (C.1.3):

$$\frac{\partial L_{A,t}^*}{\partial F(e_{B,t})} = 0 \tag{C.1.8}$$

$$\frac{\partial L^*_{B,t}}{\partial F(e_{B,t})} = -\frac{\frac{\partial f}{\partial F(e_{B,t})}}{\frac{\partial^2 Y}{\partial L^2_{B,t}}} \tag{C.1.9}$$

Given the assumptions that:

- $\frac{\partial f}{\partial F(e_{g,t})} \geq 0$ (f is weakly increasing in $F(e_{g,t})$)
- $\frac{\partial^2 Y}{\partial L^2_{B,t}} < 0$ (diminishing marginal productivity)

We have:

$$\frac{\partial L^*_{B,t}}{\partial F(e_{B,t})} \leq 0 \tag{C.1.10}$$

Therefore:

$$\frac{\partial D_t}{\partial F(e_{B,t})} = \frac{\partial L^*_{A,t}}{\partial F(e_{B,t})} - \frac{\partial L^*_{B,t}}{\partial F(e_{B,t})} = 0 - (\leq 0) \geq 0 \tag{C.1.11}$$

Since $F(e_{g,t})$ is a weakly increasing function of past experiences, and more negative experiences ($\chi^{\text{mod}}_{B,t} < \chi^{\text{neg}}_{B,t}$) imply $F(\chi^{\text{mod}}_{B,t}) \geq F(\chi^{\text{neg}}_{B,t})$, we conclude:

$$D_t(s, \psi|\chi^{\text{mod}}_{B,t}) \leq D_t(s, \psi|\chi^{\text{neg}}_{B,t}) \quad \blacksquare \tag{C.1.12}$$

C.2 Prediction 2:

$$\chi^{\text{mod}}_{g',t} < \chi^{\text{neg}}_{g',t} \Longrightarrow D_t(s, \psi|\chi^{\text{mod}}_{g',t}) = D_t(s, \psi|\chi^{\text{neg}}_{g',t})$$

Proof:

From the model specification in equation (1), the non-pecuniary costs are group-specific:

$$\sum_{g \in \{A,B\}} L_{g,t} f(d_g, F(e_{g,t})) \tag{C.2.1}$$

This means that the cost function for group A depends only on $F(e_{A,t})$, and the cost function for group B depends only on $F(e_{B,t})$.

The first-order conditions are:

$$\frac{\partial Y}{\partial L_{A,t}} - w_A - f(d_A, F(e_{A,t})) = 0 \tag{C.2.2}$$

$$\frac{\partial Y}{\partial L_{B,t}} - w_B - f(d_B, F(e_{B,t})) = 0 \tag{C.2.2}$$

Since experiences with group g' (where $g' \notin \{A, B\}$) do not enter either equation (C.5) or (C.6), we have:

$$\frac{\partial L^*_{A,t}}{\partial F(e_{g',t})} = 0 \tag{C.2.3}$$

$$\frac{\partial L^*_{B,t}}{\partial F(e_{g',t})} = 0 \tag{C.2.4}$$

Therefore:

$$\frac{\partial D_t}{\partial F(e_{g',t})} = \frac{\partial L^*_{A,t}}{\partial F(e_{g',t})} - \frac{\partial L^*_{B,t}}{\partial F(e_{g',t})} = 0 - 0 = 0 \tag{C.2.5}$$

This implies that discrimination $D_t(s, \psi)$ is invariant to past experiences with groups other than A and B:

$$D_t(s, \psi | \chi^{\text{mod}}_{g',t}) = D_t(s, \psi | \chi^{\text{neg}}_{g',t}) \quad \blacksquare \tag{C.2.6}$$

D Pilot Data Insights

D.1 Recognition of Nationality by Name

Table A1: Correctly Identified Nationality by Name During Pilot.

	Name and Nationality: Pilot	
	Ugandan Name	Eritrean Name
	(1)	(2)
Correctly Identified by Eritrean	97.33%	96.00%
Incorrectly Identified by Eritrean	2.67%	4.00%
Correctly Identified by Ugandan	100.00%	94.67%
Incorrectly Identified by Ugandan	0.00%	5.33%
N	50	50

Notes: Refers to data collected in a pilot study in December 2024 with 25 Eritrean refugees, and 25 Ugandans. None of these individuals participated in the final study. Pilot participants were shown the name of their fellow co-worker and manager (in stage 1 of the experiment), and of two workers (in stage 2 of the experiment).

D.2 Average Quality of Envelopes Made

Table A2: Quality of Envelopes and Time Taken During Pilot.

	Quality of Envelope	Time Taken to Make Envelopes
	(1)	(2)
Eritrean Workers	0.53	383.91
Ugandan Workers	0.51	380.61
N	50	50

Notes: Refers to data collected in a pilot study in December 2024 with 25 Eritrean refugees, and 25 Ugandans. None of these individuals participated in the final study. Quality of Envelope is based on five defined categories, each evaluated on a $\{0, 1\}$ scale, and averaged. Time taken to make envelopes is measured in seconds. On average, individuals made 4 envelopes. This also refers to the data shared with participants in the study.

D.3 Word Clouds of Reasons Why Given Set Number of Tasks

During the pilot study, detailed beliefs were elicited about the expectations of the number of tasks the participant would receive, and why they believed they ultimately received the number of tasks that they did. These were converted into word clouds (using a free word cloud generator). Below, I illustrate the word clouds for when the participants received two, four, and six out of eight possible tasks:

Figure A1. Word Clouds

(a) Assigned **Two** Tasks

(b) Assigned **Four** Tasks

(c) Assigned **Six** Tasks

E Theoretical Model Predictions

Based on equation (1), the nature of discrimination results in different allocations (A) across the Ugandan and Eritrean worker in stage 2 ($\{U_2, E_2\}$) across the four treatment arms ($T_1 - T4$). More specifically, theoretical predictions either predict more tasks allocated to the Ugandan worker ($\{U_2 > E_2\}$), an equal number of tasks allocated to both workers ($\{U_2 = E_2\}$), more tasks allocated to the Ugandan worker ($\{U_2 < E_2\}$), or no directional prediction ($\{U_2 \ ? \ E_2\}$):

No Discrimination: Equal allocations to both workers in the second stage, hence giving four tasks to both workers. This is independent of allocations in the first stage. Therefore, the Eritrean manager ($\mathbf{E_1}$) will allocate an equal number of tasks to the Eritrean worker (E_2) and the Ugandan worker (U_2), and this will not differ across the four treatment arms:

$$A_{T1} = A_{T2} = A_{T3} = A_{T4} = \{U_2, E_2\} = \{4, 4\}$$

Taste-Based Discrimination: Becker (1957) argues that employers have a distaste for workers of other groups. As such, we would expect that the Eritrean manager has a greater distaste for the Ugandan worker than the Eritrean worker ($d_U > d_E$). Subsequently, the manager should allocate more tasks to the Eritrean worker than the Ugandan worker when they are the manager. However, as the taste for discrimination is a fixed preference, it is independent of past experiences, and hence independent of allocations in the first stage ($f(d_g, F(e_{g,t})) = d_g$). Therefore, while the Eritrean manager ($\mathbf{E_1}$) will allocate more tasks to the Eritrean worker (E_2) than the Ugandan worker (U_2), this will not differ across the four treatment arms:

$$A_{T1} = A_{T2} = A_{T3} = A_{T4} = \{U_2 < E_2\}$$

Statistical Discrimination: Under statistical discrimination, decision-makers rely on group-level observations to draw inferences about individual workers' productivity, when individual productivity is not perfectly observable. As this task is novel (no participant had made envelopes before), managers likely did not have much information or strong priors about worker- or group-level productivity. Furthermore, managers were informed that Ugandan and Eritrean workers were equally productive at making envelopes during the pilot study (both in terms of the average time taken, and quality of the envelope). This approach has been used by other studies to minimize the scope for (inaccurate) statistical discrimination (Bohren et al., 2025b; Chan, 2025; Montoya et al., 2025).

Changes in statistical discrimination arise as a result of the employer obtaining new informa-

tion about group-level productivity. However, the productivity-related information set available to participants remains constant across the four treatments, and remains unchanged throughout the experiment.[48] As such, while participants may have priors about group's relative productivity, given that participants do not differentially learn about worker- or group-level productivity across the treatment arms, statistical discrimination — based on both accurate and inaccurate beliefs — would not result in a differential allocation across the four treatments arms:

$$A_{T1} = A_{T2} = A_{T3} = A_{T4} = \{U_2 \ ? \ E_2\}$$

In summary, both taste-based and statistical discrimination would predict no differences between treatment arms, as managerial tastes are fixed and participants do not (differentially) learn about worker's productivity in the first stage of the experiment across treatment arms.

Retaliatory Discrimination: Predictions 1 and 2 from Section 2 predict that negative past experiences, such as those as a worker in stage 1 of the experiment, can increase non-pecuniary costs in the current period, resulting in greater discrimination. However, these tastes are group-specific. As such, a past (negative) experience with a Computer manager should not affect current decisions between a Ugandan and Eritrean worker. Conversely, a past negative experience with a Ugandan manager will result in a non-positive retaliation against an (unrelated) Ugandan worker, generating discrimination:

$$A_{T1} = A_{T2} = \{U_2 \ ? \ E_2\};$$
$$A_{T3} \neq A_{T4}, \text{ specifically: } U_{2,T3} \leq U_{2,T4} \quad \Leftrightarrow \quad E_{2,T3} \geq E_{2,T4}$$

E.1 Paternalistic Discrimination

Paternalistic discrimination (Buchmann et al., 2024) — in line with the notion that refugees (and more generally, members of the minority group) are more vulnerable — would predict that managers give *fewer* tasks to refugees, to protect them from an unpleasant situation (e.g. a paper cut).

[48] Experience-based discrimination (LePage, 2024), where past hiring experiences provide information about group-level productivity, can be a micro-foundation of statistical discrimination. It arises due to managers decreasing hiring and learning about workers from group g after negative initial experiences. While participants will have prior experiences coming into the experiment, these are balanced across treatment arms (see Appendix Table A3). As participants are not differentially learning about group-level productivity across the treatment arms, experience-based discrimination would predict no differential allocations across the treatment arms.

However, no differences would be expected across the different treatment arms.

$$A_{T1} = A_{T2} = A_{T3} = A_{T4} = \{U_2 > E_2\}$$

E.2 Fairness Considerations and Altruism

Fairness concerns, where the participant cares about overall equality of pay between refugees and Ugandans, would mean that the manager allocates more tasks to the Eritrean worker when they were given two tasks in stage 1, compared with four tasks: $U_{2,T1} = U_{2,T3} > U_{2,T2} = U_{2,T4}$. However, the notion of fairness (and the subsequent allocation across the two workers) is independent of *who* the manager was in the first stage. Similarly, altruism for coethnic workers would result in more allocations to their fellow Eritrean worker, independent of allocations in the first stage:

$$A_{T1} = A_{T2} = A_{T3} = A_{T4} = \{U_2 < E_2\}$$

E.3 Social Norms

If the social norm is to split the eight tasks evenly between two workers, Treatments 1 and 3 would imply a norm violation. This norm violation may induce participants to also be more likely to deviate from the norm, compared to Treatments 2 and 4. As such, allocations in Treatments 1 and 3 would be the same, as would allocations in Treatments 2 and 4, however these two sets of allocations do not equal each other:

$$\begin{aligned} A_{T1} = A_{T3} &= \{U_2 \ ? \ E_2\}; \\ A_{T2} = A_{T4} &= \{U_2 \ ? \ E_2\} \\ A_{T1} = A_{T3} &\neq A_{T2} = A_{T4} \end{aligned}$$

E.4 Experimenter Demand Effects

The participants in the study may not only care about their own monetary payoff, but also the quality of the envelopes, as they were used by the researcher and a partner NGO. As such, they may want to allocate more tasks to the worker who they believe is more productive. However, this allocation will be unaffected by the first stage, and hence will remain constant across the four experimental arms. Predictions would be the same as those of statistical discrimination:

$$A_{T1} = A_{T2} = A_{T3} = A_{T4} = \{U_2 \ ? \ E_2\}$$

E.5 Systemic Discrimination

Systemic discrimination (Bohren et al., 2025b), where discriminatory practices are embedded within the structures and procedures of organizations, could result in differential allocations between the Ugandan and Eritrean worker — for example if participants replicate patterns they have observed elsewhere. However, this study is designed to measure differences direct discrimination at a node during a fixed time. As such, systemic discrimination would not predict differential allocations across the treatment arms:

$$A_{T1} = A_{T2} = A_{T3} = A_{T4} = \{U_2 \ ? \ E_2\}$$

E.6 Income Effects

The existence of Treatments 1 and 2 (with the Computer Manager) mitigate concerns surrounding income effects resulting from receiving either two or four tasks in the first stage. Nevertheless, the participant's own income earned in the first round may affect their behavior in round 2:

$$A_{T1} = A_{T3} = \{U_2 \ ? \ E_2\};$$
$$A_{T2} = A_{T4} = \{U_2 \ ? \ E_2\}$$

F Design Choices: Prolific Experiment

Below I justify each of the four deviations from the lab-in-the-field experiment conducted in Uganda:

1. The task differs: following Gagnon et al. (2025), participants had to copy a randomly generated sequence of letters and numbers: This was done because the envelopes could not be reproduced online, however also to use a task that had no intrinsic value, in order to reduce experimenter demand effects (de Quidt et al., 2018).

2. The nature of the discrimination (and hence workers and managers) differed: they either had white- or black-sounding names: this was due to the different nature of discrimination, given the context. This further increases the external validity of the study's findings.

3. Participants were both White and Black American men, and thus participants belonged to both the majority and minority group: this helps address issues surrounding social planner concerns, as well as documenting the widespread nature of this phenomena.

4. The allocation of the 8 tasks in stage 1 of the experiment were either favoring the participant, equally splitting the tasks, or favoring the other worker: this addresses the (a)symmetry of the results, by highlighting that retaliatory discrimination does not apply to situations of positive past experiences.

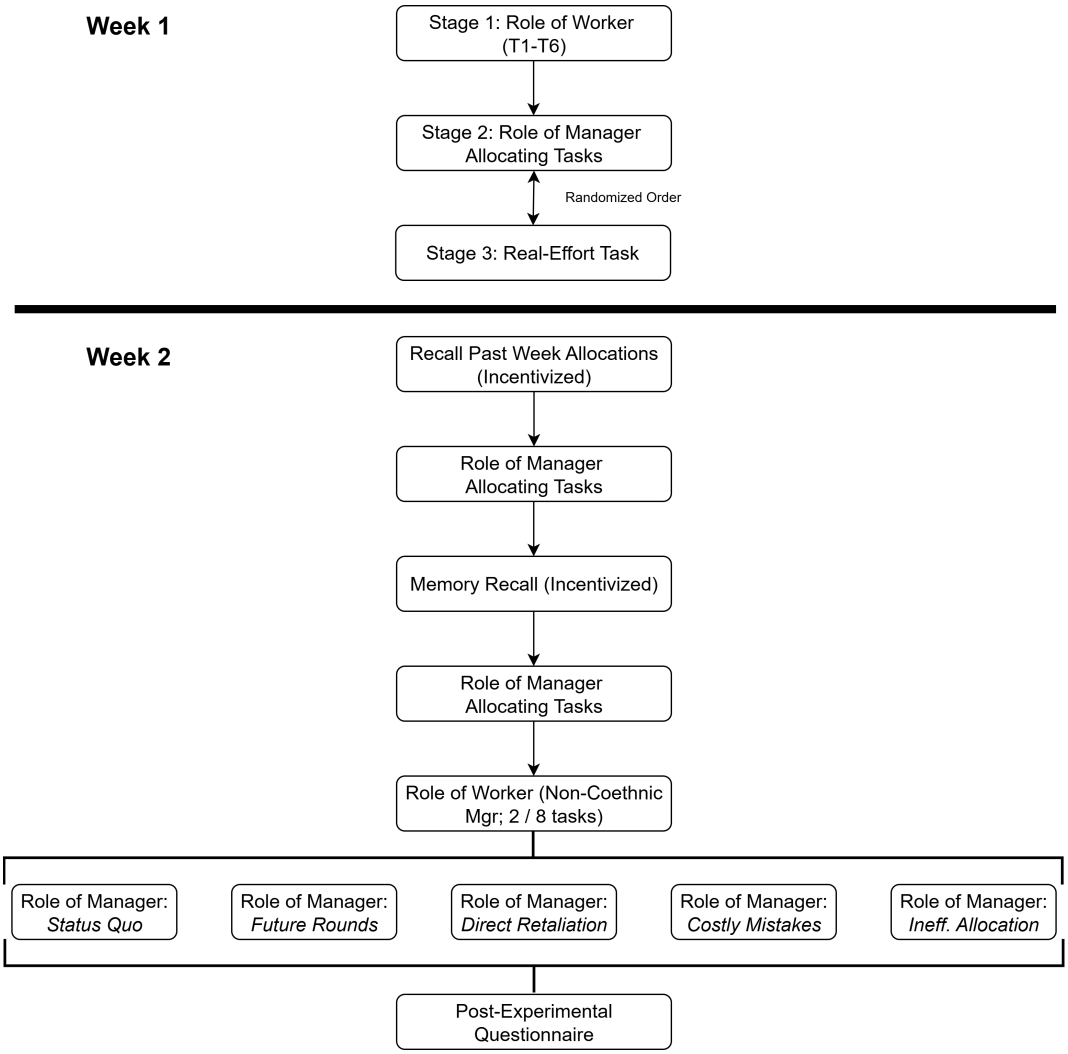


Figure A2. Overview: Experimental Design Prolific Experiment

G Balance Table

G.1 Uganda: Lab-in-the-Field

Table A3: Balance Table: Uganda.

	(T1)		(T2)		(T3)		(T4)		F-test	
	Computer Manager				***Ugandan Manager***					
	(2,6)		***(4,4)***		***(2,6)***		***(4,4)***			
Variable	N	Mean/(SD)	N	Mean/(SD)	N	Mean/(SD)	N	Mean/(SD)	N	F-stat/P-value
Age	56	30.714	55	30.418	56	31.071	57	30.772	224	0.094
		(7.586)		(5.570)		(6.760)		(5.846)		0.963
Ugandan friends	56	2.589	55	2.509	56	2.661	57	2.333	224	0.486
		(1.797)		(1.373)		(1.552)		(1.314)		0.692
Arrival Year	56	2017.232	55	2016.418	56	2017.214	57	2016.544	224	0.513
		(4.884)		(4.003)		(5.263)		(3.689)		0.674
Attitudes Towards Ugandans	56	-0.120	55	0.122	56	-0.022	57	0.017	224	1.885
		(0.642)		(0.557)		(0.457)		(0.499)		0.133
Empathy Index	56	0.135	55	-0.137	56	-0.010	57	-0.105	224	3.707**
		(0.552)		(0.438)		(0.442)		(0.454)		0.012
Retaliation Index	56	-0.103	55	0.105	56	0.227	57	0.200	224	2.215*
		(0.826)		(0.606)		(0.728)		(0.825)		0.087

Notes: Columns (T1), (T2), (T3), and (T4) show the average value (and standard deviation) for respondents in each of the four treatment arms: Ugandan manager (2,6), Ugandan manager (4,4), Computer manager (2,6), and Computer manager (4,4), where values in parentheses indicate the allocation of the manager in the first stage of the game. The F-test reports the joint test for orthogonality, including both the F-statistic and associated p-value. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

G.2 America: Online Experiment

Table A4: Balance Table: USA.

	(T1)		(T2)		(T3)		(T4)		(T5)		(T6)		F-test	
		(2,6)		***Coethnic Manager (4,4)***		***(6,2)***		***(2,6)***		***Non-Coethnic Manager (4,4)***		***(6,2)***		
Variable	N	Mean/(SD)	N	Mean/(SD)	N	Mean/(SD)	N	Mean/(SD)	N	Mean/(SD)	N	Mean/(SD)	N	F-stat/P-value
Age	108	40.824	105	39.648	109	40.872	99	38.960	109	38.697	109	38.550	639	1.145
		(10.340)		(9.779)		(9.613)		(9.788)		(10.863)		(10.321)		0.335
Total Approvals	108	2650.694	105	2201.610	109	2412.633	99	2638.556	109	2035.055	109	2517.532	639	1.158
		(2665.565)		(2215.184)		(2502.470)		(2419.970)		(1798.169)		(2529.070)		0.329
Ethnicity: African American	108	0.481	105	0.495	109	0.477	99	0.485	109	0.523	109	0.450	639	0.249
		(0.502)		(0.502)		(0.502)		(0.502)		(0.502)		(0.500)		0.940
USA National	108	1.000	105	0.981	109	0.982	99	0.990	109	0.963	109	1.000	639	1.515
		(0.000)		(0.137)		(0.135)		(0.101)		(0.189)		(0.000)		0.183
Student	108	0.093	105	0.133	109	0.055	99	0.071	109	0.119	109	0.128	639	1.235
		(0.291)		(0.342)		(0.229)		(0.258)		(0.326)		(0.336)		0.291
Employed	108	0.093	105	0.086	109	0.073	99	0.061	109	0.101	109	0.064	639	0.372
		(0.291)		(0.281)		(0.262)		(0.240)		(0.303)		(0.246)		0.868
Detailed Elicitation	108	0.509	105	0.457	109	0.495	99	0.455	109	0.523	109	0.523	639	0.401
		(0.502)		(0.501)		(0.502)		(0.500)		(0.502)		(0.502)		0.848

Notes: Columns (T1), (T2), (T3), (T4), (T5), and (T6) show the average value (and standard deviation) for respondents in each of the six treatment arms: Same Ethnicity Manager (2,6), Same Ethnicity Manager (4,4), Same Ethnicity Manager (5,2) Other Ethnicity Manager (2,6), Other Ethnicity Manager (4,4), and Other Ethnicity Manager (6,2), where values in parentheses indicate the allocation of the manager in the first stage of the game. The F-test reports the joint test for orthogonality, including both the F-statistic and associated p-value. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Table A5: Balance Table: USA, Detailed Belief Elicitation.

	No		Detailed Beliefs		t-test	
Variable	N	Mean/(SD)	N	Mean/(SD)	N	p-value
Age	323	40.050	316	39.139	639	0.257
		(10.357)		(9.907)		
Total Approvals	323	2456.734	316	2355.522	639	0.590
		(2416.707)		(2330.703)		
Ethnicity: African American	323	0.489	316	0.481	639	0.837
		(0.501)		(0.500)		
USA National	323	0.978	316	0.994	639	0.100
		(0.146)		(0.079)		
Student	323	0.093	316	0.108	639	0.536
		(0.291)		(0.310)		
Employed	323	0.068	316	0.092	639	0.271
		(0.252)		(0.289)		

Notes: Columns show the average value (and standard deviation) for respondents who either provided detailed beliefs, or did not. The t-test reports the associated p-value. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

G.3 Minimal Group Paradigm

The Minimal Group Paradigm study was conducted on Besample, an online survey platform similar to Prolific that surveys participants across many countries. For this study, 320 men were recruited from Kenya, Ethiopia, Ghana and Nigeria. Appendix Table A6 presents the balance table for this sample:

Table A6: Balance Table: Minimal Group Paradigm.

	(T1)		(T2)		(T3)		(T4)		(T5)		(T6)		F-test	
							Treatment Groups							
Variable	N	Mean/(SD)	N	Mean/(SD)	N	Mean/(SD)	N	Mean/(SD)	N	Mean/(SD)	N	Mean/(SD)	N	F-stat/P-value
Urban	53	0.849	51	0.725	58	0.845	55	0.891	53	0.887	56	0.786	326	1.539
		(0.361)		(0.451)		(0.365)		(0.315)		(0.320)		(0.414)		0.177
Employed	53	0.585	51	0.431	58	0.500	55	0.382	53	0.547	56	0.482	326	1.182
		(0.497)		(0.500)		(0.504)		(0.490)		(0.503)		(0.504)		0.318
Age	53	27.887	51	26.882	58	28.914	55	27.673	53	29.698	56	28.929	326	1.190
		(6.883)		(5.945)		(7.373)		(6.449)		(7.360)		(6.954)		0.314
Nationality: Ghana	53	0.340	51	0.451	58	0.345	55	0.400	53	0.321	56	0.411	326	0.568
		(0.478)		(0.503)		(0.479)		(0.494)		(0.471)		(0.496)		0.725
Nationality: Kenya	53	0.208	51	0.118	58	0.190	55	0.182	53	0.151	56	0.125	326	0.522
		(0.409)		(0.325)		(0.395)		(0.389)		(0.361)		(0.334)		0.759
Nationality: Nigeria	53	0.245	51	0.255	58	0.293	55	0.291	53	0.264	56	0.304	326	0.150
		(0.434)		(0.440)		(0.459)		(0.458)		(0.445)		(0.464)		0.980
Highest Schooling: Primary	53	0.000	51	0.000	58	0.000	55	0.018	53	0.000	56	0.000	326	0.985
		(0.000)		(0.000)		(0.000)		(0.135)		(0.000)		(0.000)		0.427
Highest Schooling: Secondary	53	0.264	51	0.275	58	0.259	55	0.255	53	0.189	56	0.268	326	0.276
		(0.445)		(0.451)		(0.442)		(0.440)		(0.395)		(0.447)		0.926
Highest Schooling: Bachelors	53	0.547	51	0.627	58	0.603	55	0.600	53	0.679	56	0.643	326	0.446
		(0.503)		(0.488)		(0.493)		(0.494)		(0.471)		(0.483)		0.816
Highest Schooling: Masters	53	0.132	51	0.059	58	0.086	55	0.073	53	0.094	56	0.071	326	0.442
		(0.342)		(0.238)		(0.283)		(0.262)		(0.295)		(0.260)		0.819
Highest Schooling: Ph.D.	53	0.019	51	0.000	58	0.017	55	0.000	53	0.000	56	0.018	326	0.573
		(0.137)		(0.000)		(0.131)		(0.000)		(0.000)		(0.134)		0.721
Highest Schooling: Vocational	53	0.038	51	0.020	58	0.034	55	0.055	53	0.038	56	0.000	326	0.643
		(0.192)		(0.140)		(0.184)		(0.229)		(0.192)		(0.000)		0.667

Notes: Columns (T1)–(T6) show the average value (and standard deviation) for respondents in each of the six treatment arms. The F-test reports the joint test for orthogonality, including both the F-statistic and associated p-value. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

G.4 Dictator Game

Table A7: Balance Table: Dictator Game.

	(T1)		(T2)		(T3)		(T4)		(T5)		(T6)		F-test	
		Coethnic Manager					***Non-Coethnic Manager***							
	(2,6)		***(4,4)***		***(6,2)***		***(2,6)***		***(4,4)***		***(6,2)***			
Variable	N	Mean/(SD)	N	Mean/(SD)	N	Mean/(SD)	N	Mean/(SD)	N	Mean/(SD)	N	Mean/(SD)	N	F-stat/P-value
Age	62	39.081	61	37.656	62	36.758	62	38.919	61	38.475	61	39.230	369	0.558
		(9.479)		(9.832)		(9.634)		(9.958)		(9.944)		(11.920)		0.732
Total Approvals	62	3597.194	61	2571.344	62	2553.984	62	2426.806	61	2648.803	61	3200.656	369	2.229*
		(2893.761)		(2150.441)		(2255.177)		(2179.148)		(2632.518)		(2373.807)		0.051
Ethnicity: African American	62	0.226	61	0.230	62	0.226	62	0.242	61	0.246	61	0.246	369	0.032
		(0.422)		(0.424)		(0.422)		(0.432)		(0.434)		(0.434)		0.999
USA National	62	1.000	61	1.000	62	1.000	62	1.000	61	1.000	61	1.000	369	
		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		
Student	62	0.129	61	0.000	62	0.097	62	0.065	61	0.115	61	0.098	369	1.728
		(0.338)		(0.000)		(0.298)		(0.248)		(0.321)		(0.300)		0.127
Employed	62	0.129	61	0.148	62	0.048	62	0.097	61	0.066	61	0.082	369	1.018
		(0.338)		(0.358)		(0.216)		(0.298)		(0.250)		(0.277)		0.406

Notes: Columns (T1), (T2), (T3), (T4), (T5), and (T6) show the average value (and standard deviation) for respondents in each of the six treatment arms: Same Ethnicity Manager (2,6), Same Ethnicity Manager (4,4), Same Ethnicity Manager (5,2) Other Ethnicity Manager (2,6), Other Ethnicity Manager (4,4), and Other Ethnicity Manager (6,2), where values in parentheses indicate the allocation of the manager in the first stage of the game. The F-test reports the joint test for orthogonality, including both the F-statistic and associated p-value. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

H Regression Tables - Uganda Experiment

Table A8: Allocation of Tasks to Ugandan Worker in Stage 2.

	Allocation of Tasks to U_2 in Stage 2	
	Computer Manager in Stage 1	Ugandan Manager in Stage 1
	(1)	(2)
Stage 1: Negative	-0.17	-0.53***
	(0.13)	(0.15)
Control Group Mean	3.55	3.63
Control Group S.D.	0.66	0.70
N	111	113

Notes: Intention to Treat estimates. The outcome variable is the number of tasks allocated to the Ugandan worker by the participant in the second stage of the experiment, and ranges from 0 to 8. Control variables are selected using the post double LASSO machine learning algorithm outlined in Belloni et al. (2014). *Stage 1: Negative* is a dummy variable equal to 1 if the allocation of the manager in the first round was (6 ; 2), and hence refers to treatments T1 and T3. Column (1) reports results for the sub-sample who had a Computer manager in stage 1 (T1 and T2), while column (2) reports results for the sub-sample who had a Ugandan manager in stage 1 (T3 and T4). Control mean and standard deviation refer to the mean value and standard deviation of the outcome in the control group (T2 and T4, respectively). Robust standard errors are in parentheses.***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Table A9: Allocation of Tasks to Ugandan Worker in Stage 2.

	Allocation of Tasks to U_2 in Stage 2	
	(1)	(2)
Stage 1: Ugandan Manager	-0.05	0.10
	(0.10)	(0.12)
Stage 1: Negative		-0.25*
		(0.13)
Interaction Term		-0.31
		(0.19)
p-value: T1 vs. T2		0.39
p-value: T3 vs. T4		0.00
p-value: T1 vs. T3		0.04
p-value: T2 vs. T4		0.50
p-value: T1 & T2 vs. T3		0.01
Control Group Mean	3.49	3.55
Control Group S.D.	0.71	0.66
N	224	224

Notes: Intention to Treat estimates. The outcome variable is the number of tasks allocated to the Ugandan worker by the participant in the second stage of the experiment, and ranges from 0 to 8. Control variables are selected using the post double LASSO machine learning algorithm outlined in Belloni et al. (2014). *Stage 1: Ugandan Manager* is a dummy variable equal to 1 if the manager in the first round was Ugandan, and hence refers to treatments T3 and T4. *Stage 1: Negative* is a dummy variable equal to 1 if the allocation of the manager in the first round was (6 ; 2), and hence refers to treatments T1 and T3. The *Interaction Term* refers to *Stage 1: Ugandan Manager* interacted with *Stage 1: Negative*. Control mean and standard deviation refer to the mean value and standard deviation of the outcome in the control group (Treatment: Computer manager with (4, 4) allocation in the first stage). Robust standard errors are in parentheses.***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Table A10: Time Taken to Make Envelopes.

	Time Taken to Make Envelopes (in seconds)			
	No Winsorizing		95th percentile Winsorizing	
	(1)	(2)	(3)	(4)
Stage 1: Ugandan Manager	-15.87	0.88	-15.99	1.17
	(16.61)	(26.21)	(15.36)	(23.74)
Stage 1: Negative		-64.11***		-62.45***
		(22.90)		(21.81)
Interaction Term		-34.94		-35.76
		(31.30)		(28.71)
p-value: T1 vs. T2		0.01		0.01
p-value: T3 vs. T4		0.00		0.00
p-value: T1 vs. T3		0.05		0.04
p-value: T2 vs. T4		0.97		0.96
Control Group Mean	277.87	310.21	276.26	307.76
Control Group S.D.	124.97	143.08	119.19	136.06
N	224	224	224	224

Notes: Intention to Treat estimates. The outcome variable is the number of seconds the participant took to make the allocated number of envelopes in the first stage of the experiment. Control variables are selected using the post double LASSO machine learning algorithm outlined in Belloni et al. (2014). *Stage 1: Ugandan Manager* is a dummy variable equal to 1 if the manager in the first round was Ugandan, and hence refers to treatments T3 and T4. *Stage 1: Negative* is a dummy variable equal to 1 if the allocation of the manager in the first round was (6 ; 2), and hence refers to treatments T1 and T3. The *Interaction Term* refers to *Stage 1: Ugandan Manager* interacted with *Stage 1: Negative*. Columns (1) and (2) report results when outliers are not winsorized, while columns (3) and (4) reports results when outliers are winsorized at the 95th percentile, separately per treatment arm as discussed in Wicker (2025). Control mean and standard deviation refer to the mean value and standard deviation of the outcome in the control group (Treatment: Computer manager with (4, 4) allocation in the first stage). Robust standard errors are in parentheses.***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Table A11: Quality of Envelopes.

	Quality of Envelopes Made in Stage 1	
	Computer Manager in Stage 1	Ugandan Manager in Stage 1
	(1)	(2)
Stage 1: Negative	0.03	-0.10*
	(0.06)	(0.05)
Control Group Mean	0.50	0.52
Control Group S.D.	0.26	0.28
N	111	113

Notes: Intention to Treat estimates. The outcome variable is the average quality of the envelopes produced by the participant in the first stage of the experiment, and ranges from 0 to 1. The five pre-registered components of *Envelope Quality* were: sides of envelope have a finger width; triangle fold is in the middle; creases are tight and straight; glue still sticks; and top fold is sharp. Control variables are selected using the post double LASSO machine learning algorithm outlined in Belloni et al. (2014). *Stage 1: Negative* is a dummy variable equal to 1 if the allocation of the manager in the first round was (6 ; 2), and hence refers to treatments T1 and T3. Column (1) reports results for the sub-sample who had a Ugandan manager in stage 1 (T3 and T4), while column (2) reports results for the sub-sample who had a Ugandan manager in stage 1 (T1 and T2). Control mean and standard deviation refer to the mean value and standard deviation of the outcome in the control group (T2 and T4, respectively). Robust standard errors are in parentheses.***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Table A12: Quality of Envelopes.

	Quality of Envelopes Made in Stage 1	
	(1)	(2)
Stage 1: Ugandan Manager	-0.01	0.03
	(0.04)	(0.05)
Stage 1: Negative		0.01
		(0.06)
Interaction Term		-0.09
		(0.08)
p-value: T1 vs. T2		0.61
p-value: T3 vs. T4		0.16
p-value: T1 vs. T3		0.18
p-value: T2 vs. T4		0.63
Control Group Mean	0.51	0.50
Control Group S.D.	0.31	0.26
N	224	224

Notes: Intention to Treat estimates. The outcome variable is the number of tasks allocated to the Ugandan worker by the participant in the second stage of the experiment, and ranges from 0 to 8. Control variables are selected using the post double LASSO machine learning algorithm outlined in Belloni et al. (2014). *Stage 1: Ugandan Manager* is a dummy variable equal to 1 if the manager in the first round was Ugandan, and hence refers to treatments T3 and T4. *Stage 1: Negative* is a dummy variable equal to 1 if the allocation of the manager in the first round was (6 ; 2), and hence refers to treatments T1 and T3. The *Interaction Term* refers to *Stage 1: Ugandan Manager* interacted with *Stage 1: Negative*. Control mean and standard deviation refer to the mean value and standard deviation of the outcome in the control group (Treatment: Computer manager with (4, 4) allocation in the first stage). Robust standard errors are in parentheses.***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Table A13: Placebo Test: Expected Number of Envelopes in First Stage.

	Expected Number of Envelopes in Stage 1	
	(1)	(2)
Stage 1: Ugandan Manager	0.27**	0.13
	(0.13)	(0.17)
Stage 1: Negative		-0.25
		(0.19)
Interaction Term		0.27
		(0.26)
p-value: T1 vs. T2		0.03
p-value: T3 vs. T4		0.87
p-value: T1 vs. T3		0.01
p-value: T2 vs. T4		0.47
Control Group Mean	3.87	4.09
Control Group S.D.	1.05	1.02
N	224	224

Notes: Intention to Treat estimates. The outcome variable is the number of tasks allocated to the Ugandan worker by the participant in the second stage of the experiment, and ranges from 0 to 8. Control variables are selected using the post double LASSO machine learning algorithm outlined in Belloni et al. (2014). *Stage 1: Ugandan Manager* is a dummy variable equal to 1 if the manager in the first round was Ugandan, and hence refers to treatments T3 and T4. *Stage 1: Negative* is a dummy variable equal to 1 if the allocation of the manager in the first round was (6 ; 2), and hence refers to treatments T1 and T3. The *Interaction Term* refers to *Stage 1: Ugandan Manager* interacted with *Stage 1: Negative*. Control mean and standard deviation refer to the mean value and standard deviation of the outcome in the control group (Treatment: Computer manager with (4, 4) allocation in the first stage). Robust standard errors are in parentheses.***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Table A14: Discrepancy of Expected vs. Actual Envelopes on Stage 2 Allocations.

	Allocation of Tasks to U_2 in Stage 2	
	(1)	(2)
Discrepancy: Expected - Actual Envs.	-0.14***	-0.04
	(0.04)	(0.06)
Stage 1: Ugandan Manager		0.16
		(0.12)
Interaction Term		-0.18**
		(0.08)
Control Group Mean	3.68	3.67
Control Group S.D.	0.68	0.66
N	224	224

Notes: Intention to Treat estimates. The outcome variable is the number of tasks allocated to the Ugandan worker by the participant in the second stage of the experiment, and ranges from 0 to 8. Control variables are selected using the post double LASSO machine learning algorithm outlined in Belloni et al. (2014). *Stage 1: Ugandan Manager* is a dummy variable equal to 1 if the manager in the first round was Ugandan, and hence refers to treatments T3 and T4. *Discrepancy* is the difference between the expected number of envelopes, and the actual number of envelopes the participant received in Stage 1. A positive value implies that the participant received *fewer* tasks than they expected. The *Interaction Term* refers to *Stage 1: Ugandan Manager* interacted with *Discrepancy*. Control mean and standard deviation refer to the mean value and standard deviation of the outcome in the control group (Treatment: Computer manager with 0 discrepancy between expected and received envelopes). Robust standard errors are in parentheses.***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

H.1 Heterogeneity

Table A15: HTE: Ugandan Friends
Allocation of Tasks to Ugandan Worker in Stage 2.

	Allocation of Tasks to U_2 in Stage 2	
	Ugandan Manager in Stage 1	Computer Manager in Stage 1
	(1)	(2)
Stage 1: Negative	-0.50***	-0.44*
	(0.18)	(0.23)
Ugandan Friends	-0.38**	0.15
	(0.18)	(0.19)
Interaction Term	-0.18	0.29
	(0.25)	(0.27)
Control Group Mean	3.63	3.55
Control Group S.D.	0.70	0.66
N	113	111

Notes: Intention to Treat estimates. The outcome variable is the number of tasks allocated to the Ugandan worker by the participant in the second stage of the experiment, and ranges from 0 to 8. Control variables are selected using the post double LASSO machine learning algorithm outlined in Belloni et al. (2014). *Stage 1: Negative* is a dummy variable equal to 1 if the manager in the first round was Ugandan, and hence refers to treatments T3 and T4. *Ugandan Friends* refers to the number of Ugandan friends the participant reported to have. The *Interaction Term* refers to *Stage 1: Negative* interacted with *Ugandan Friends*. Column (1) reports results for the sub-sample who had a Ugandan manager in stage 1 (T3 and T4), while column (2) reports results for the sub-sample who had a Ugandan manager in stage 1 (T1 and T2). Control mean and standard deviation refer to the mean value and standard deviation of the outcome in the control group (T2 and T4, respectively). Robust standard errors are in parentheses.***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Table A16: HTE: Empathy
Allocation of Tasks to Ugandan Worker in Stage 2.

	Allocation of Tasks to U_2 in Stage 2	
	Ugandan Manager in Stage 1	Computer Manager in Stage 1
	(1)	(2)
Stage 1: Negative	-0.70***	-0.25
	(0.20)	(0.25)
Empathy	0.11	0.12
	(0.24)	(0.20)
Interaction Term	0.11	-0.14
	(0.28)	(0.28)
Control Group Mean	3.63	3.55
Control Group S.D.	0.70	0.66
N	113	111

Notes: Intention to Treat estimates. The outcome variable is the number of tasks allocated to the Ugandan worker by the participant in the second stage of the experiment, and ranges from 0 to 8. Control variables are selected using the post double LASSO machine learning algorithm outlined in Belloni et al. (2014). *Stage 1: Negative* is a dummy variable equal to 1 if the manager in the first round was Ugandan, and hence refers to treatments T3 and T4. *Empathy* refers to an inverse-covariance weighted index of five 5-item Likert scale questions: "Other people's misfortunes do not disturb me a great deal"; "It upsets me to see someone being treated disrespectfully"; "I am not really interested in how other people feel"; "When I see someone being treated unfairly, I do not feel very much pity for them"; "When I see someone being taken advantage of, I feel protective towards him/her." The *Interaction Term* refers to *Stage 1: Negative* interacted with *Empathy*. Column (1) reports results for the sub-sample who had a Ugandan manager in stage 1 (T3 and T4), while column (2) reports results for the sub-sample who had a Ugandan manager in stage 1 (T1 and T2). Control mean and standard deviation refer to the mean value and standard deviation of the outcome in the control group (T2 and T4, respectively). Robust standard errors are in parentheses.***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Table A17: HTE: Retaliation
Allocation of Tasks to Ugandan Worker in Stage 2.

	Allocation of Tasks to U_2 in Stage 2	
	Ugandan Manager in Stage 1	Computer Manager in Stage 1
	(1)	(2)
Stage 1: Negative	-0.67***	-0.15
	(0.18)	(0.18)
Retaliate	-0.16	0.13
	(0.20)	(0.19)
Interaction Term	0.03	-0.42
	(0.28)	(0.27)
Control Group Mean	3.63	3.55
Control Group S.D.	0.70	0.66
N	113	111

Notes: Intention to Treat estimates. The outcome variable is the number of tasks allocated to the Ugandan worker by the participant in the second stage of the experiment, and ranges from 0 to 8. Control variables are selected using the post double LASSO machine learning algorithm outlined in Belloni et al. (2014). *Stage 1: Negative* is a dummy variable equal to 1 if the manager in the first round was Ugandan, and hence refers to treatments T3 and T4. *Retaliate* refers to an inverse-covariance weighted index of two 7-item Likert scale questions: "If someone does me a favor, I am ready to return it to them"; "If someone treats me unfairly, I'll take the opportunity to get back at them." The *Interaction Term* refers to *Stage 1: Negative* interacted with *Retaliate*. Column (1) reports results for the sub-sample who had a Ugandan manager in stage 1 (T3 and T4), while column (2) reports results for the sub-sample who had a Ugandan manager in stage 1 (T1 and T2). Control mean and standard deviation refer to the mean value and standard deviation of the outcome in the control group (T2 and T4, respectively). Robust standard errors are in parentheses.***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Table A18: HTE: Attitudes Towards Ugandans
Allocation of Tasks to Ugandan Worker in Stage 2.

	Allocation of Tasks to U_2 in Stage 2	
	Ugandan Manager in Stage 1	Computer Manager in Stage 1
	(1)	(2)
Stage 1: Negative	-0.46**	-0.15
	(0.18)	(0.24)
Attitudes Towards Ugandans	0.28	0.34*
	(0.17)	(0.21)
Interaction Term	-0.41	-0.15
	(0.26)	(0.27)
Control Group Mean	3.63	3.55
Control Group S.D.	0.70	0.66
N	113	111

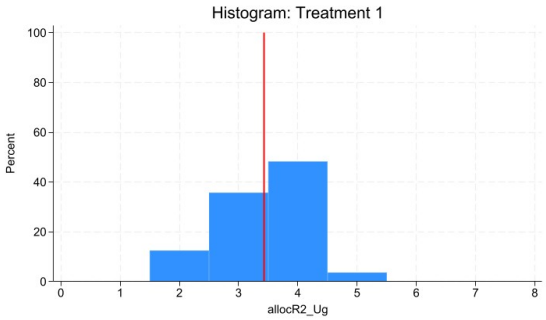
Notes: Intention to Treat estimates. The outcome variable is the number of tasks allocated to the Ugandan worker by the participant in the second stage of the experiment, and ranges from 0 to 8. Control variables are selected using the post double LASSO machine learning algorithm outlined in Belloni et al. (2014). *Stage 1: Negative* is a dummy variable equal to 1 if the manager in the first round was Ugandan, and hence refers to treatments T3 and T4. *Attitudes Towards Ugandans* refers to an inverse-covariance weighted index of four 5-item Likert scale questions: "Ugandans are friendly and good people"; "Eritreans are well integrated with Ugandans"; "Ugandan employers discriminate against me because I am an Eritrean"; "I have just as many opportunities to find formal work as my Ugandan neighbors." The *Interaction Term* refers to *Stage 1: Negative* interacted with *Attitudes Towards Ugandans*. Column (1) reports results for the sub-sample who had a Ugandan manager in stage 1 (T3 and T4), while column (2) reports results for the sub-sample who had a Ugandan manager in stage 1 (T1 and T2). Control mean and standard deviation refer to the mean value and standard deviation of the outcome in the control group (T2 and T4, respectively). Robust standard errors are in parentheses.***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Table A19: HTE: Years in Uganda
Allocation of Tasks to Ugandan Worker in Stage 2.

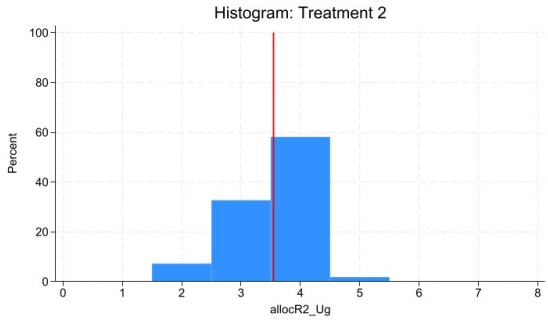
	Allocation of Tasks to U_2 in Stage 2	
	Ugandan Manager in Stage 1	Computer Manager in Stage 1
	(1)	(2)
Stage 1: Negative	-0.50***	-0.31
	(0.16)	(0.21)
Years in Uganda	0.08	-0.03
	(0.20)	(0.21)
Interaction Term	-0.31	-0.03
	(0.27)	(0.28)
Control Group Mean	3.63	3.55
Control Group S.D.	0.70	0.66
N	113	111

Notes: Intention to Treat estimates. The outcome variable is the number of tasks allocated to the Ugandan worker by the participant in the second stage of the experiment, and ranges from 0 to 8. Control variables are selected using the post double LASSO machine learning algorithm outlined in Belloni et al. (2014). *Stage 1: Negative* is a dummy variable equal to 1 if the manager in the first round was Ugandan, and hence refers to treatments T3 and T4. *Years in Uganda* refers to the number of years the participant has lived in Uganda. The *Interaction Term* refers to *Stage 1: Negative* interacted with *Years in Uganda*. Column (1) reports results for the sub-sample who had a Ugandan manager in stage 1 (T3 and T4), while column (2) reports results for the sub-sample who had a Ugandan manager in stage 1 (T1 and T2). Control mean and standard deviation refer to the mean value and standard deviation of the outcome in the control group (T2 and T4, respectively). Robust standard errors are in parentheses.***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

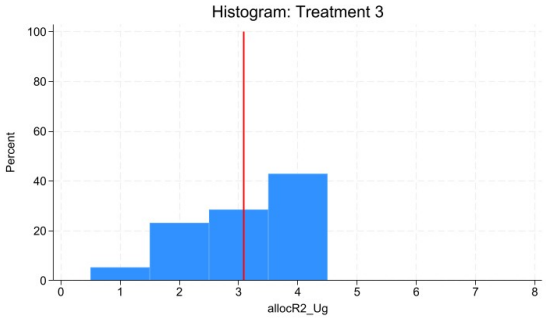
H.2 Histograms



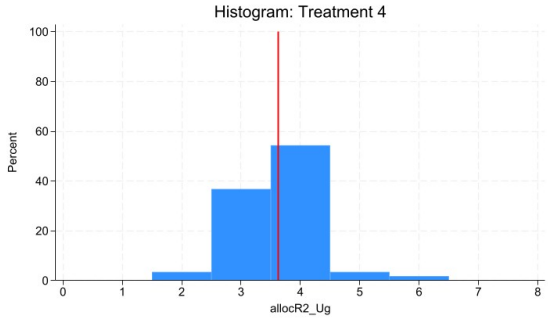
(a) Treatment 1



(b) Treatment 2



(c) Treatment 3



(d) Treatment 4

Figure A3. Histograms: Uganda Lab-in-the-Field Experiment

I Regression Tables - America Experiment

Table A20: Errors and Effort of Real Effort Task.

	(1) Error Rate	(2) Number of Tasks Completed
Stage 1: Non-Coethnic Manager	-0.01	0.35
	(0.02)	(0.49)
Stage 1: Positive	0.00	0.15
	(0.02)	(0.43)
Stage 1: Negative	0.00	0.02
	(0.02)	(0.43)
Stage 1: Non-Coethnic & Positive	0.02	-0.37
	(0.03)	(0.69)
Stage 1: Non-Coethnic & Negative	0.01	-0.94
	(0.03)	(0.61)
Order Effects	-0.01	0.20
	(0.01)	(0.26)
T1 Mean	0.06	6.29
T1 S.D.	0.13	3.50
N	639	639

Notes: The outcome variable is the number of errors, and number of tasks completed during the real-effort task, where participants had 60 seconds to complete as many tasks as possible in order to increase their likelihood of being hired by a Non-Coethnic Manager. Control variables are selected using the post double LASSO machine learning algorithm outlined in Belloni et al. (2014). *Stage 1: Non-Coethnic Manager* is a dummy variable equal to 1 if the manager in the first round was non-coethnic, and hence refers to treatments T4-T6. *Stage 1: Negative* is a dummy variable equal to 1 if the allocation of the manager in the first round was (6 ; 2), and hence refers to treatments T1 and T4. *Stage 1: Positive* is a dummy variable equal to 1 if the allocation of the manager in the first round was (2 ; 6), and hence refers to treatments T3 and T6. The *Interaction Terms* refers to *Stage 1: Non-Coethnic Manager* interacted with *Stage 1: Negative*, and *Stage 1: Positive*, respectively. Control mean and standard deviation refer to the mean value and standard deviation of the outcome in the control group (Coethnic manager with (4 ; 4) allocation in the first stage). Robust standard errors are in parentheses. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

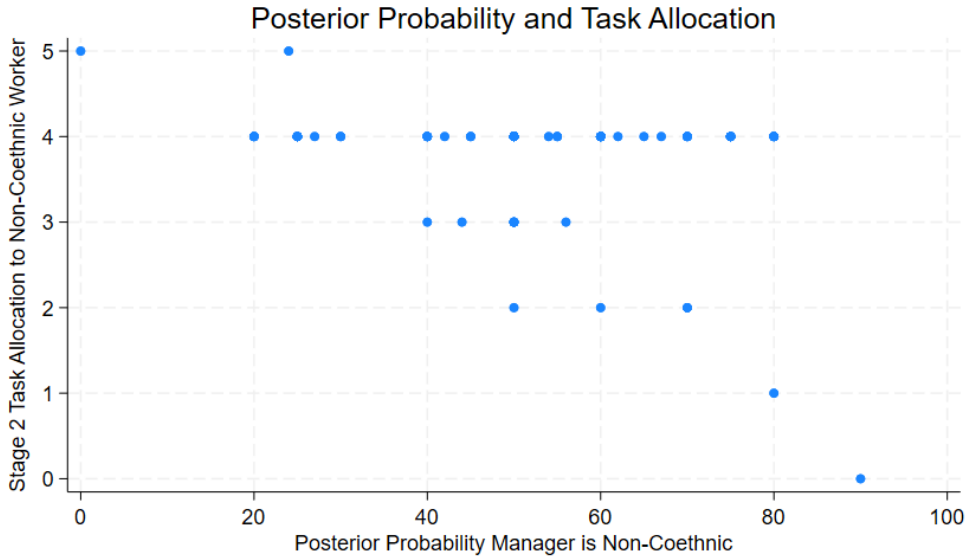


Figure A4. Role of Posterior Beliefs on Subsequent Managerial Allocation

Table A21: Effects of Recall on Persistence.

	(1) Number of Tasks Allocated to Non-Coethnic Worker Week 2
Recalled Non-Coethnic Manager	0.03
	(0.18)
Number of Tasks Recalled	0.01
	(0.04)
Interaction Term	0.00
	(0.04)
Mean	3.99
S.D.	0.61
N	460

Notes: The outcome variable is the number of tasks allocated to the Non-Coethnic worker by the participant during the follow-up experiment one week later, and ranges from 0 to 8. Control variables are selected using the post double LASSO machine learning algorithm outlined in Belloni et al. (2014). *Recalled Non-Coethnic Manager* is a dummy variable equal to 1 if the participant successfully recalled the name of their previous manager, from a multiple-choice list. *Number of Tasks Recalled* is a dummy variable equal to 1 if the participant successfully recalled the number of tasks assigned to them by their previous manager. The *Interaction Term* refers to *Recalled Non-Coethnic Manager* interacted with *Number of Tasks Recalled*. Control mean and standard deviation refer to the mean value and standard deviation of the outcome of participants who neither recalled their previous manager nor the number of allocated tasks. Robust standard errors are in parentheses. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Table A22: Costly Mistakes and Discriminatory Allocations

	(1) Number of Tasks Allocated to Non-Coethnic Worker
Treatment: *Costly Mistakes*	0.08
	(0.10)
Status Quo Mean	3.90
Status Quo S.D.	0.72
N	153

Notes: Intention to Treat estimates. Control variables are selected using the post double LASSO machine learning algorithm outlined in Belloni et al. (2014). *Costly Mistakes* refers to the treatment arm where mistakes by the workers would reduce the payoff of the managers. *Status Quo* mean and standard deviation refer to the mean value and standard deviation of the outcome in the treatment arm where the salience of future rounds was not made salient (and hence equivalent to T4 of Figure 3). Robust standard errors are in parentheses. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Table A23: Inefficient Allocations and Discriminatory Allocations

	(1) Number of Tasks Allocated to Non-Coethnic Worker
Treatment: *Inefficient Allocation*	0.02
	(0.11)
Status Quo Mean	3.90
Status Quo S.D.	0.72
N	149

Notes: Intention to Treat estimates. Control variables are selected using the post double LASSO machine learning algorithm outlined in Belloni et al. (2014). *Inefficient Allocation* refers to the treatment arm where the most efficient division of tasks entailed an even division of tasks, as tasks got increasingly more complex. *Status Quo* mean and standard deviation refer to the mean value and standard deviation of the outcome in the treatment arm where the salience of future rounds was not made salient (and hence equivalent to T4 of Figure 3). Robust standard errors are in parentheses. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Table A24: Memory Recall of Past Rounds

	(1)	(2)	(3)	(4)
	Correctly Recalled Allocations		Discrepancy: Recall Tasks for Coethnic Worker	
Coethnic Manager	0.02*		0.01	
	(0.01)		(0.04)	
Coethnic Mgr. Pref Coethnic Worker		0.03		-0.65***
		(0.02)		(0.07)
Coethnic Mgr. Pref Non-Coethnic Worker		0.02		2.57***
		(0.02)		(0.06)
Non-Coethnic Mgr. Pref Coethnic Worker		-0.01		2.32***
		(0.02)		(0.07)
Non-Coethnic Mgr. Pref Non-Coethnic Worker		0.00		-0.48***
		(0.02)		(0.06)
Coethnic Mgr. No Pref		0.03		0.79***
		(0.03)		(0.08)
Non-Coethnic Mgr. No Pref: Mean	0.41	0.41	-0.01	-0.01
Non-Coethnic Mgr. No Pref: S.D.	0.49	0.49	1.92	1.92
N	4025	4025	4025	4025

Notes: The outcome variables are whether the participant correctly recalled the allocation of tasks by managers during the memory recall task; and the discrepancy in the recall. Control variables are selected using the post double LASSO machine learning algorithm outlined in Belloni et al. (2014). *Coethnic Manager* is a dummy variable equal to 1 if the manager in the shown round was coethnic. *Coethnic Manager Pref Coethnic Worker* is a dummy variable equal to 1 if the manager in the shown round was Coethnic and allocated more tasks to the Coethnic worker. *Coethnic Manager Pref Non-Coethnic Worker* is a dummy variable equal to 1 if the manager in the shown round was Coethnic and allocated more tasks to the Non-Coethnic worker. *Non-Coethnic Manager Pref Coethnic Worker* is a dummy variable equal to 1 if the manager in the shown round was Non-Coethnic and allocated more tasks to the Coethnic worker. *Non-Coethnic Manager Pref Non-Coethnic Worker* is a dummy variable equal to 1 if the manager in the shown round was Non-Coethnic and allocated more tasks to the Non-Coethnic worker. *Coethnic Manager No Pref* is a dummy variable equal to 1 if the manager in the shown round was Coethnic and allocated the tasks evenly between both workers. Control mean and standard deviation refer to the mean value and standard deviation of the outcome when the shown manager was Non-Coethnic and allocated the tasks evenly between both workers. Robust standard errors are in parentheses. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Table A25: Memory Recall on Retaliatory Discrimination

	(1)	(2)
	Number of Tasks Allocated to Non-Coethnic Worker	
Correctly Recalled Rounds	0.03	
	(0.10)	
Average Discrepancy of Recall		0.07
		(0.07)
T1 Mean	4.05	4.05
T1 S.D.	0.69	0.69
N	451	451

Notes: The outcome variable is the number of tasks allocated to the Non-Coethnic worker by the participant after the memory recall task, and ranges from 0 to 8. Control variables are selected using the post double LASSO machine learning algorithm outlined in Belloni et al. (2014). *Correctly Recalled Rounds* is a variable that counts the number of correctly recalled rounds, out of 10. *Average Discrepancy of Recall* is a variable that reports the average discrepancy between the recalled, and actual, managerial allocations. Control mean and standard deviation refer to the mean value and standard deviation of the outcome variable. Robust standard errors are in parentheses. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

I.1 Heterogeneity

Table A26: HTE: Discriminatory Attitudes
Allocation of Tasks to Non-Coethnic Worker in Stage 2.

	(1)	(2)	(3)
		Number of Tasks Allocated to Non-Coethnic Worker	
	Whole Sample	Below Median Discrim. Index	Above Median Discrim. Index
Stage 1: Non-Coethnic Manager	-0.03	-0.00	-0.12
	(0.04)	(0.04)	(0.11)
Stage 1: Positive	-0.09	-0.04	-0.30*
	(0.08)	(0.10)	(0.15)
Stage 1: Negative	-0.02	0.07	-0.09
	(0.06)	(0.06)	(0.12)
Stage 1: Non-Coethnic & Positive	0.16	0.10	0.34**
	(0.10)	(0.17)	(0.17)
Stage 1: Non-Coethnic & Negative	-0.18*	-0.40**	0.14
	(0.10)	(0.18)	(0.14)
T1 Mean	4.02	4.02	4.02
T1 S.D.	0.38	0.38	0.38
N	639	228	234

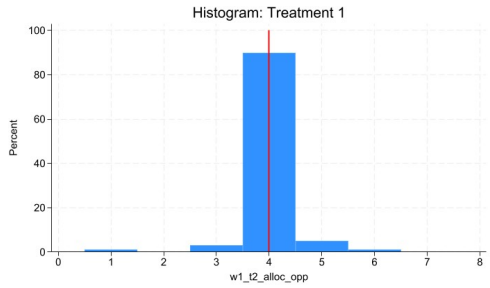
Notes: The outcome variable is the number of tasks allocated to the Non-Coethnic worker by the participant in the second stage of the experiment, and ranges from 0 to 8. Regression results are reported separately for (1) the whole sample; (2) participants with a below-median discrimination score; (3) participants with an above-median discrimination score. Control variables are selected using the post double LASSO machine learning algorithm outlined in Belloni et al. (2014). *Stage 1: Non-Coethnic Manager* is a dummy variable equal to 1 if the manager in the first round was non-coethnic, and hence refers to treatments T4-T6. *Stage 1: Negative* is a dummy variable equal to 1 if the allocation of the manager in the first round was (6 ; 2), and hence refers to treatments T1 and T4. *Stage 1: Positive* is a dummy variable equal to 1 if the allocation of the manager in the first round was (2 ; 6), and hence refers to treatments T3 and T6. The *Interaction Terms* refers to *Stage 1: Non-Coethnic Manager* interacted with *Stage 1: Negative*, and *Stage 1: Positive*, respectively. Control mean and standard deviation refer to the mean value and standard deviation of the outcome in the control group (Coethnic manager with (4 ; 4) allocation in the first stage). Robust standard errors are in parentheses. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Table A27: HTE: Future Rounds
African American and White Men.

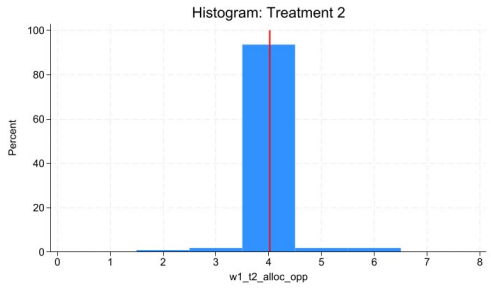
	(1)	(2)	(3)
	Number of Tasks Allocated to Non-Coethnic Worker		
	Whole Sample	African American Men	White Men
Treatment: *Future Rounds*	0.17*	0.10	0.19
	(0.11)	(0.09)	(0.20)
Status Quo Mean	3.90	3.91	3.88
Status Quo S.D.	0.72	0.51	0.94
N	148	85	63

Notes: Intention to Treat estimates. The outcome variable is the number of tasks allocated to the Non-Coethnic worker by the participant in the second stage of the experiment, and ranges from 0 to 8. These are reported separately for (1) the whole sample; (2) African American men; (3) White men. Control variables are selected using the post double LASSO machine learning algorithm outlined in Belloni et al. (2014). *Future Rounds* refers to the treatment arm where the experimental instructions heightened the salience of future rounds. *Status Quo* mean and standard deviation refer to the mean value and standard deviation of the outcome in the treatment arm where the salience of future rounds was not made salient (and hence equivalent to T4 of Figure 3). Robust standard errors are in parentheses. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

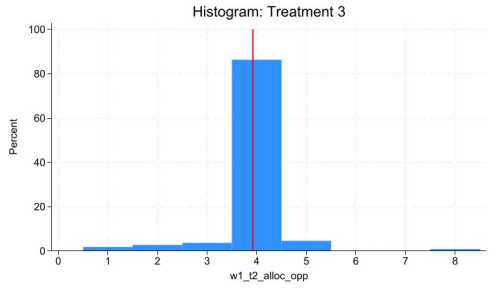
I.2 Histograms



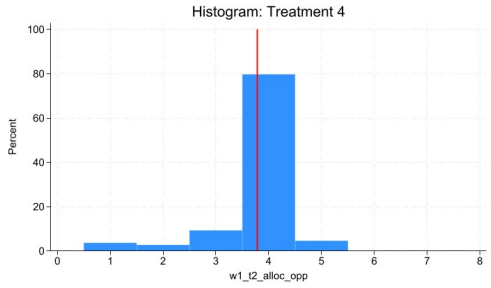
(a) Treatment 1



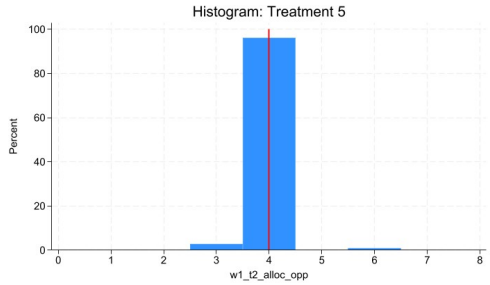
(b) Treatment 2



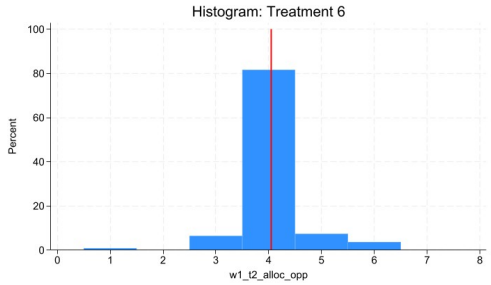
(c) Treatment 3



(d) Treatment 4



(e) Treatment 5



(f) Treatment 6

Figure A5. Histograms: Prolific Experiment

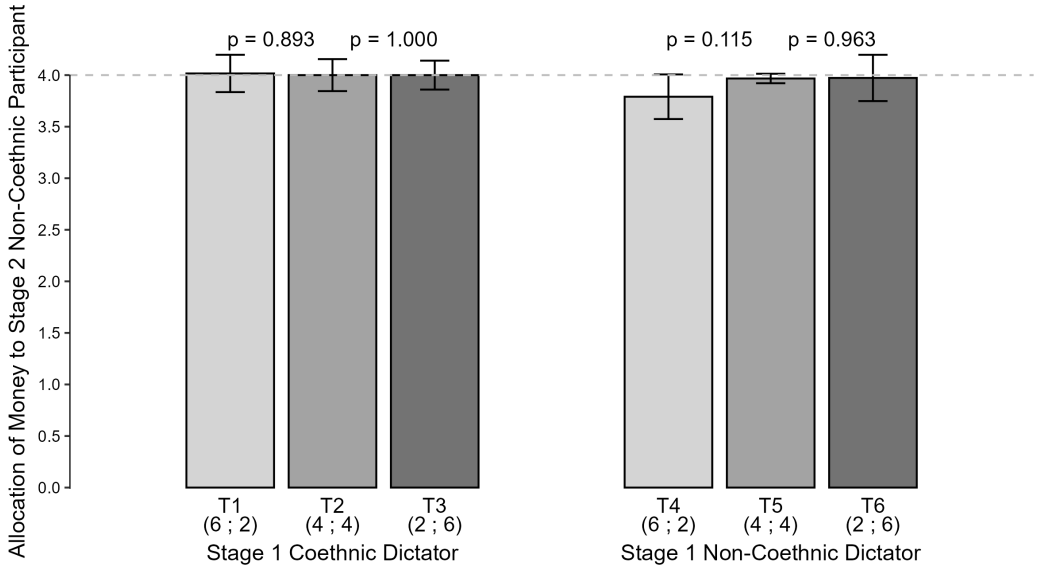


Figure A6. Money Allocation to Non-Coethnic Worker in Stage 2: Dictator Game

J Ruling Out Alternative Mechanisms

(Inaccurate) Statistical Discrimination

One alternative explanation is that participants had inaccurate beliefs about the productivity of workers of different groups, which impacted their allocation of tasks. To minimize this mechanism, prior to the start of the experiment, participants were informed that "Pilot study data showed that on average, individuals from different ethnicities and genders are equally fast and accurate." In the lab-in-the-field experiment in Uganda, participants were even shown numbers to support this claim, see Appendix Table A2. This is a frequently used approach in experimental studies to minimize the role of (inaccurate) statistical discrimination (for example, see Chan (2025)).

To further rule out statistical discrimination — both accurate and inaccurate (Bohren et al., 2025a) — I replicate the experimental design of Figure 3 with six treatment arms as a dictator game. Hence, instead of completing tasks (where beliefs about productivity may play a role), individuals simply divide money. This approach rules out statistical discrimination, as individuals do not need to form beliefs about worker productivity. Appendix Figure A6 illustrates that the pattern documented in Figure 4 is replicated in the dictator game version of the experiment, ruling

out accurate and inaccurate statistical discrimination as a mechanism.

Tit-for-Tat and Reciprocity

The initial models of social preferences such as fairness considerations, other-regarding preferences, and reciprocity (see Rabin (1993); Fehr and Schmidt (1999)) do not consider the role of identity or group affiliation. As such, these models would predict (negative) reciprocity not only in T4 of Figure 4, but also T1, when individuals could reciprocate after perceiving discrimination by a manager of their same ethnicity. We furthermore observe no positive reciprocity, see T3 and T6 of Figure 4.

Furthermore, in a sub-treatment of the online experiment, participants get the opportunity to retaliate directly against their stage 1 manager when they become the manager in stage 2, rather than retaliating against a different non-coethnic worker. Individuals retaliate far more aggressively against their previous manager, compared to a member of the same ethnicity as the manager ($p = 0.080$, see Appendix Table A28) in contrast to predictions of the theoretical models of reciprocity.

Table A28: Direct Retaliation and Discriminatory Allocations

	(1) Number of Tasks Allocated to Non-Coethnic Worker
Treatment: *Direct Retaliation*	-0.38*
	(0.22)
Status Quo Mean	3.90
Status Quo S.D.	0.72
N	151

Notes: Intention to Treat estimates. Control variables are selected using the post double LASSO machine learning algorithm outlined in Belloni et al. (2014). *Direct Retaliation* refers to the treatment arm where participants could directly retaliate against their stage 1 manager, when they become manager in stage 2. *Status Quo* mean and standard deviation refer to the mean value and standard deviation of the outcome in the treatment arm where the salience of future rounds was not made salient (and hence equivalent to T4 of Figure 3). Robust standard errors are in parentheses. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Lastly, to illustrate that the salience of group differences and the salience of group-based

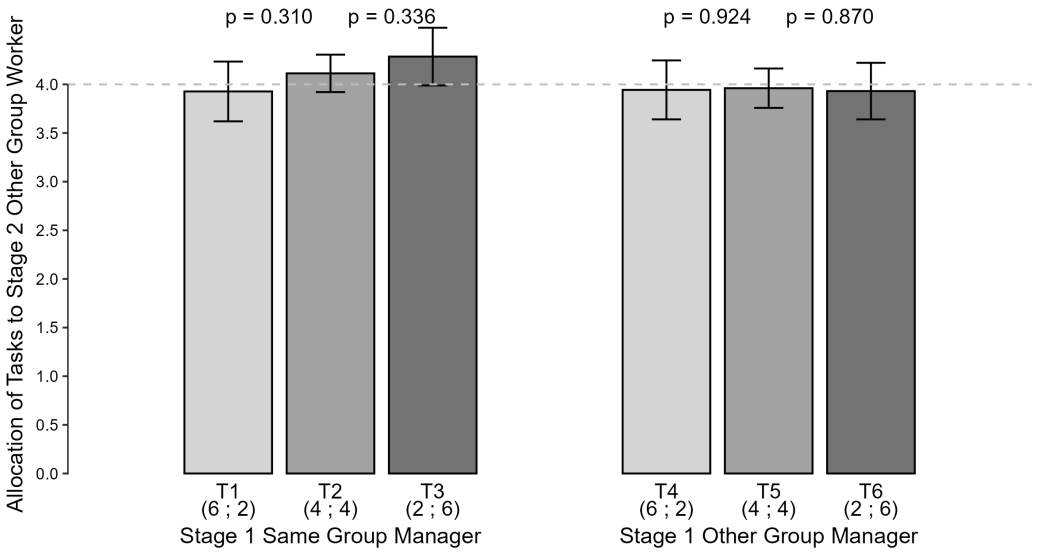


Figure A7. Task Allocation to Other Group Worker in Stage 2: Minimal Group Paradigm

discrimination matters, the online experiment of Figure 3 is replicated among a new sample with one variation: rather than exploring task allocations among the racial ethnicity dimension, participants are arbitrarily divided into a Red and Blue team. This is based on the minimal group paradigm of social psychology (Tajfel, 1970). No retaliatory discrimination is documented in the minimal group paradigm setting (see Appendix Figure A7), suggesting that artificially invoking group status is not enough to induce discriminatory preferences, contrary to predictions of reciprocity and tit-for-tat strategies.

Norm Violation

An alternative explanation could be that, rather than documenting retaliatory discrimination, Figure 4 captures norm violations: having observed managers deviate from the fair allocation of tasks $(4; 4)$, participants are more likely to do so once they become managers. If this were the case, we would expect average task allocations to differ from an even split in treatment arms where participants observed their manager deviating from the social norm of fairness (T1, T3, T4, T6). While the number of allocations that deviate from an even split $(4\ ;\ 4)$ is higher compared to treatments where the stage 1 manager split the tasks evenly ($p < 0.001$), average allocations do

not differ significantly ($p = 0.818$).[49]

Furthermore, if social norms were driving the treatment effects, we would expect to document treatment effects after the memory recall task. Eight of the ten managerial allocations participants were asked to recall deviated from the even split (4 ; 4) norm (see Appendix M). As such, we would expect that participants would be more willing to deviate from the social norm after observing several previous managers do the same. The percentage of participants deviating from the social norm increases to 21.17% ($p < 0.001$), however norm violations are not unidirectional: in 85% of norm violation cases, the norm violation was in *favor* of the worker of the different ethnicity. This goes against predictions that the negative treatment effects observed in T4 are due to norm violations.

Finally, no discriminatory behavior is documented after the memory recall task, indicating that the memory recall, and associated heightened salience of norm violations, did not cause the observed discriminatory behavior of Figures 2 and 4.

In-Group Favoritism

The treatment effects could arise not as a result of retaliatory behavior against out-group members, but instead due to in-group favoritism. The *Computer* Manager treatment arm from the lab-in-the-field experiment can help rule out that the mechanism is indeed in-group favoritism.

If treatment effects documented in Figures 2 and 4 are due to in-group favoritism, we would expect Eritrean participants to also favor the Eritrean worker when their previous manager was a Computer. While we do document that they favor Eritrean workers, allocating statistically significantly more than four tasks to the Eritrean worker ($p < 0.001$), However, compared to T3, allocations to the Eritrean worker are significantly less (0.40 fewer tasks, $p = 0.003$). This is in contrast to predictions of in-group favoritism and in line with the notion of retaliatory discrimination. As such, we rule out in-group favoritism as a potential explanation.

Secondly, if in-group favoritism were driving the results, we would expect the presence of discrimination across all treatment arms. However, in the online experiment among American men, discrimination (defined as an allocation of tasks differing from an even split) is only observed in T4.

[49] Results are similar for the dictator and minimum group paradigm games: $p = 0.026$ and $p < 0.001$; and $p = 0.469$ and $p = 0.894$, respectively.

Social Planner and Preference for Equality

A further concern could be that participants act as social planners — in particular with refugees in Uganda — and hence want to allocate more tasks to workers who are less well off. This could explain why tasks are unevenly distributed across all four treatments in the Ugandan experiment, as refugees are typically perceived to be more vulnerable than Ugandans. The same reasoning would be expected to hold for African American men, who have historically been disadvantaged in the labor market (Lang and Lehmann, 2012). However, as Figure 4 illustrates, there is no systemic favoring of African American workers.[50]

Furthermore, if participants were acting as social planners, we would expect a similar treatment effect as the one observed in T4 to also be documented in T1 in Figure 4. However, we do not observe this, as treatment effects are statistically significantly different ($p = 0.019$). I try to minimize the likelihood that participants feel that workers have been discriminated against in past activities, by highlighting in the introduction that "all workers, including yourself, have not participated in these tasks before". As such, participants should not have ex-ante expectations that workers with a particular pseudo-name have been discriminated against in earlier rounds of the game. In line with this, when participants are asked to justify their allocation across the two workers, no participant cited reasons related to workers having been discriminated against in the past, and hence acting as a social planner.

Closely related to the idea of being a social planner that equals out past individual injustices, the participant could also have a preference for equality across groups. In this case, participants would want to reverse the allocations made in stage 1 when they become managers in stage 2, in order to balance out aggregate tasks (and hence earnings) across the two ethnic groups. However, only 0.9% of participants did this. Furthermore, we would subsequently anticipate that participants will award fewer tasks to workers of their same ethnicity if they received more than four tasks in the first stage. This is only documented in 8.26% of cases.

Anger

Rather than discriminatory preferences being the driving mechanism, an alternative explanation is that participants were angry, and hence retaliated. Anger is typically thought of as a System-1 response, and hence impulsive (Kahneman, 2011). In Section 4.2 and Table A29, I illustrate that first having to complete a real-effort task, that takes $\sim$ 3 minutes before making allocation decisions

[50] We also don't observe heterogeneity by perceptions of discrimination in Uganda (see Appendix Table A18), and find that participants with below-median discriminatory perceptions the USA have larger treatment effects (see Appendix Table A26), in contrast to predictions of a social planner.

does not affect retaliatory discrimination, contrary to what would be expected if impulsive anger were the driving mechanism. Furthermore, anger is likely invoked as a result of getting assigned fewer tasks than expected. As such, T1 in Figure 4 should also induce anger, as participants also receive two tasks.[51] As such, we would expect retaliation in T1, however we do not observe this ($p = 1.000$). Lastly, anger as a micro-foundation for the treatment effects observed in T4 is unable to rationalize the treatment effects on the real-effort task discussed in Section 6.2.

Experimenter Demand Effects

A concern with experiments hosted in non-natural settings is that participants behave differently than they would in real life, and respond as they believe the researcher would want them to. I adopt several approaches to minimize this. By conducting experiments in-person and online and on different populations, I increase the external validity of the findings, reducing the likelihood that participants across different samples both give socially desirable answers. As de Quidt et al. (2018) discuss, online experiments — where individuals can complete the experiment on their own devices without the physical presence of the experimenter — reduce the potential for experimenter demand effects. Secondly, I vary the usefulness of the tasks across the in-person experiment in Uganda, and the online experiment. In Uganda, participants made envelopes that were used by an NGO for a cash transfer program, and hence the task was useful. Participants may have had an incentive to appease the researcher and allocate tasks such that envelopes were of the highest quality. This is ruled out in the online experiment: following Gagnon et al. (2025), I have participants complete a task that is of no use to anyone. I furthermore explicitly state in the instructions: "The experimenters will not derive any earnings from your decisions. The lines of numbers and/or letters that are entered have no further use for anyone." By consistently finding similar results among an online sample, and an in-person sample, and with tasks that vary in their usefulness, I minimize the role of experimenter demand effects. Finally, if experimenter demand effects played a major role, we would have expected to find results in the minimal group paradigm experiment, which we do not.

[51] In line with this, participants expected to receive more tasks in T1 than T4, however this difference is not statistically significant (4.53 vs. 4.26, $p = 0.252$).

Table A29: Controlling for Order Effects:
Allocation of Tasks to Non-Coethnic Worker in Stage 2.

	(1) Number of Tasks Allocated to Non-Coethnic Worker
Stage 1: Non-Coethnic Manager	-0.03
	(0.05)
Stage 1: Positive	-0.09
	(0.08)
Stage 1: Negative	-0.02
	(0.06)
Stage 1: Non-Coethnic & Positive	0.16
	(0.10)
Stage 1: Non-Coethnic & Negative	-0.19*
	(0.10)
Order Effects	-0.03
	(0.05)
T1 Mean	4.02
T1 S.D.	0.38
N	639

Notes: The outcome variable is the number of tasks allocated to the Non-Coethnic worker by the participant in the second stage of the experiment, and ranges from 0 to 8. Control variables are selected using the post double LASSO machine learning algorithm outlined in Belloni et al. (2014). *Stage 1: Non-Coethnic Manager* is a dummy variable equal to 1 if the manager in the first round was non-coethnic, and hence refers to treatments T4-T6. *Stage 1: Negative* is a dummy variable equal to 1 if the allocation of the manager in the first round was (6 ; 2), and hence refers to treatments T1 and T4. *Stage 1: Positive* is a dummy variable equal to 1 if the allocation of the manager in the first round was (2 ; 6), and hence refers to treatments T3 and T6. The *Interaction Terms* refers to *Stage 1: Non-Coethnic Manager* interacted with *Stage 1: Negative*, and *Stage 1: Positive*, respectively. Control mean and standard deviation refer to the mean value and standard deviation of the outcome in the control group (Coethnic manager with (4 ; 4) allocation in the first stage). Robust standard errors are in parentheses. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

K Anecdotal Evidence of Retaliatory Discrimination

After the online experiment, participants were asked whether the notion of retaliatory discrimination resonated with their own past experiences. Below are some of the responses of participants:

> "I have experienced situations where I felt unfairly treated or discriminated against by someone from a different ethnic group. Sometimes, this led to feelings of frustration or resentment. In some cases, I noticed that I unconsciously responded by being stricter or less cooperative toward others from the same group as the person who treated me unfairly, even though they had nothing to do with the original incident."

> "I have felt treated unfairly at work or school because of my ethnicity. Even if I didn't confront the person, it sometimes affected how I acted toward others in similar situations."

> "I once experienced being treated unfairly by a supervisor, which later made me feel less inclined to cooperate with another colleague from the same background, even though they weren't personally responsible."

> "I once faced bias from a supervisor, and later felt tempted to be less cooperative with another coworker from their group."

> "I once felt unfairly treated by a manager during a group project, he consistently gave me the least desirable tasks. Later when I had to assign roles in a different setting to someone from his same background, I found myself feeling tempted to be less fair."

> "If a Black boss acts aggressively toward me, I may replicate the same behavior toward my Black subordinates."

> "In a previous job, after being treated unfairly by a manager from a certain background, I caught myself being less cooperative with another colleague from that same background."

> "In a previous job,a manager from from one ethnic group consistently gave me fewer opportunities. Later, when I had authority over someone from the same background, I had to consciously check my own bias to avoid unfair treatment. It was a wake up call about how resentment can linger if we're not self aware."

> "In my final year at university, I once felt a lecturer graded my work unfairly compared to my classmates from his own ethnic group. A week later, I was in a group assignment where one teammate was from the same group as that lecturer. I noticed I was initially less willing to cooperate, but I made a conscious effort not to let the previous incident influence me."

L Retaliatory Discrimination Micro-Foundations

Prior to the launch of the online experiment on Prolific, I pre-registered four theoretical micro-foundations of retaliatory discrimination on the AEA RCT Registry (AEARCTR-0016047). The function $f\Big(d_g, F(\chi_{g,t})\Big)$ specified in Section 2 is increasing in both arguments, however, the functional form of f is important for understanding the dynamic evolution of discriminatory tastes. Below I present four different specifications of $f(d_g, F(\chi_{g,t}))$ — Retaliatory Tit-for-Tat (social preferences), Bayesian Updating, Motivated Beliefs, and Memory Recall.

L.1 Reciprocity-Based Tit-for-Tat (Social Preferences)

$f(d_g, F(e_{g,t}))$ is simplified to only consider the experience in the last period, with an individual of group g: $f(d_g, e_{g,t-1})$. In particular, it takes an additively separable form consisting of the individual's discriminatory prior, d_g, and a tit-for-tat reciprocity-based update, following Rabin (1993):

$$f(d_g, \chi_{g,t-1}) = d_g + \underbrace{\phi[d_{g,t}(\chi_{g,t-1}) - d_g]}_{\text{tit-for-tat}}$$

where $d_{g,t}(\chi_{g,t-1})$ captures the (perceived) discrimination as a result of interactions with individuals of group g in the previous period, $t-1$. ϕ captures the re-activeness of the tit-for-tat response, where a value of $\phi = 1$ corresponds to fully retaliatory behavior, and $\phi = 0$ no retaliatory behavior (and hence a return to the static, taste-based model of Becker (1957)). Values in between ($0 < \phi < 1$) capture partial tit-for-tat, and $\phi > 1$ captures over-retaliation.

L.2 Bayesian Updating

Employers have a prior distribution of their discriminatory distaste for workers from group g distributed according to the density function $h(d_g)$ of population-level discriminatory preferences with a prior mean $\bar{d}_g$. Employers interact with workers of group g, and these interactions inform the information set of the employer, $I_{g,t}$. Through each interaction, an i.i.d. signal is drawn from the group-specific distaste distribution $D_g|d_g \sim G(x_g)$, characterized by the mean d_g, finite variance, and density function $g(x_g)$. Conditional on d_g, experiences with workers of group g are independent ($x_{k,g}|d_g \overset{\text{i.i.d.}}{\sim} p(x_g|d_g)$), and hence the joint likelihood of the signals is $p(I_{g,t}|d_g) = \prod_{k \in I_{g,t}} g(x_{k,g})$. As a result, the posterior distribution of d_g given the observed signals $I_{g,t}$ is:

$$d_g|I_{g,t} = \frac{\prod_{k \in I_{g,t}} g_{d_g}(x_{k,g}) h(d_g)}{\int \prod_{k \in I_{g,t}} g_{d_g}(x_{k,g}) h(d_g)\, dd_g}$$

and hence

$$f(d_g, F(\chi_{g,t})) = E\Big[\frac{\prod_{k\in I_{g,t}} g_{d_g}(x_{k,g})h(d_g)}{\int \prod_{k\in I_{g,t}} g_{d_g}(x_{k,g})h(d_g)\, dd_g}\Big]$$

To derive a closed-form solution for the posterior mean of the employer's discriminatory distastes, $f(d_g, F(\chi_{g,t}))$, I will assume that both the prior distribution $h(d_g)$ and likelihood function $p(x_g|d_g)$ follow a Gaussian normal distribution. In particular, $h(d_g) \sim \mathcal{N}(\bar{d}_g, \frac{1}{\tau_{d_g}})$, and $p(x_g|d_g) \sim \mathcal{N}(\bar{x}_{g,t}, \frac{1}{\tau_{x_{g,t}}})$.[52] Subsequently,

$$f(d_g, F(\chi_{g,t})) = \frac{\bar{d}_g \cdot \tau_{d_g} + \bar{x}_{g,t} \cdot \tau_{x_{g,t}} \cdot n_t}{\tau_{d_g} + \tau_{x_{g,t}} \cdot n_t}$$

where $n_t = |I_{g,t}|$, the number of signals the employer received up until time t of individuals in group g - each with the same signal precision τ_{d_g}.[53]

Given the standard setting with Gaussian prior and signal distributions, the posterior distribution can also be written as a linear combination of the two:

$$f(d_g, F(\chi_{g,t})) = \omega_g \bar{d}_g + (1-\omega) p(x_g|d_g)$$

where ω_g is the weight on the prior mean for group g, and defined as:

$$\omega_g = \frac{\tau_{d_g}}{n_t \cdot \tau_{x_{g,t}} + \tau_{d_g}}$$

and hence increasing in the precision of the prior distribution versus the signal.

L.3 Motivated Beliefs

Following the large theoretical and empirical literature documenting that individuals do not always update as a Bayesian does, I incorporate motivated beliefs within the Bayesian updating model.[54] In particular, a motive function is introduced, in line with Thaler (2024). The motive function is applied to individual signal $x_{k,g}$ received by the employer through an experience with individuals

[52] By setting $\tau_{d_g} \to \infty$, the prior distribution of tastes becomes a single value, akin to Becker (1957).

[53] The assumption of equal signal precision can easily be dropped, as is the case with motivated beliefs, see below.

[54] Motivated beliefs can also be referred to as the confirmation bias (Rabin and Schrag, 1999) and reference-dependent preferences (Kőszegi and Rabin, 2006).

from group g. In particular, the motive function affects the precision of the signal:

$$\tilde{\tau}_{k,g} = \tau_{x_{g,t}} \cdot M(x_{k,g}, \bar{d}_g)$$

where $M(x_{k,g}, \bar{d}_g)$ is defined as:

$$M(x_{k,g}, \bar{d}_g) = \begin{cases} 1+\alpha & \text{if } \text{sign}(x_{k,g}) \cdot \text{sign}(\bar{d}_g) \geq 0 \quad \text{[confirming signal]} \\ 1-\beta & \text{if } \text{sign}(x_{k,g}) \cdot \text{sign}(\bar{d}_g) < 0 \quad \text{[contradicting signal]} \end{cases}$$

$\alpha > 0$ and $\beta \in (0,1)$ capture the additional weight or discount applied to signals confirming and contradicting the employer's prior $\bar{d}_g$, compared with the Bayesian standard. $\alpha > \beta$ captures motivated reasoning, as an employer places more weight on signals that align with their discriminatory priors $\bar{d}_g$ than signals that go against their priors. As such, the likelihood function is now based on the motive-adjusted precision: $p(x_{k,g}|d_g, \tilde{\tau}_{k,g})$. As a consequence, the posterior distribution of d_g given the observed signals $I_{g,t}$ is:

$$d_g|I_{g,t} = \frac{\prod_{k\in I_{g,t}} p(x_{k,g}|d_g, \tilde{\tau}_{k,g}) h(d_g)}{\int \prod_{k\in I_{g,t}} p(x_{k,g}|d_g, \tilde{\tau}_{k,g}) h(d_g)\, dd_g}$$

and hence

$$f(d_g, F(\chi_{g,t})) = E\Big[\frac{\prod_{k\in I_{g,t}} p(x_{k,g}|d_g, \tilde{\tau}_{k,g}) h(d_g)}{\int \prod_{k\in I_{g,t}} p(x_{k,g}|d_g, \tilde{\tau}_{k,g}) h(d_g)\, dd_g}\Big]$$

Using the same Gaussian assumption as under the Bayesian Updating model to obtain a closed form solution, the final posterior mean becomes:

$$f(d_g, F(\chi_{g,t})) = \frac{\bar{d}_g \cdot \tau_{d_g} + \sum_{k\in I_{g,t}} \tau_{x_{g,t}} \cdot M(x_{k,g}, \bar{d}_g) \cdot \bar{x}_{k,g}}{\tau_{d_g} + \sum_{k\in I_{g,t}} \tau_{x_{g,t}} \cdot M(x_{k,g}, \bar{d}_g)} = \frac{\bar{d}_g \cdot \tau_{d_g} + \sum_{k\in I_{g,t}} \tilde{\tau}_{k,g} \cdot \bar{x}_{k,g}}{\tau_{d_g} + \sum_{k\in I_{g,t}} \tilde{\tau}_{k,g}}$$

L.4 Recall of Memories

In line with Bordalo et al. (2024), the average *distaste* parameter is a weighted average of the employer's "true", static taste for discrimination, and their experience-based discriminatory tastes. The relative weighting of the static $(1-\rho)$ and dynamic component (ρ) is assumed to be exogenous, however this assumption can be relaxed.[55] $d_g(\kappa_{g,t})$ captures the situation-specific discriminatory factor that is influenced by the recall of past experiences.

[55] For example, an employer with no experience may only rely on their distaste, and hence $\rho = 0$. As employers get more experience, they may place more emphasis on past experiences, $\rho > 0$.

$$f(d_g, F(\chi_{g,t})) = \underbrace{\sum_{g \in \{A,B\}} L_{g,t}\Big((1-\rho)\cdot d_g + \rho \cdot d_g(\kappa_{g,t})\Big)}_{\text{Distaste}}$$

The current decision of the employer comes with a cue $\kappa_{g,t}$, and the employer has a database $F(\tilde{e})$ of relevant past experiences, consisting of experiences with individuals of group g ($F(\tilde{e}_g)$), and other groups g' ($F(\tilde{e}_{g'})$. The cue $\kappa_{g,t}$ is characterized by several defining attributes, including the group affiliation of the individual engaging with (g), and context (c).[56] Following Kahana (2012); Bordalo et al. (2020, 2024); Miserocchi (2023), recall of experiences is characterized by their similarity to the current setting, and interference. Interference refers to the case when the recall of a given memory is weakened by other memories that are more similar to the cue of the current situation, $\kappa_{g,t}$. The similarity of an experience with an individual of group g in time period $t-k$, $\chi_{g,t-k} \equiv (d_g, c)$ to the cue κ_t of the current decision is given by the multiplicatively separable distance:

$$S(\chi_{g,t-k}, \kappa_{g,t}) \equiv S_1(d_g - d_{g,t-k}|)S_2(|c_t - c_{t-k}|)$$

$S_j : R_+ \Rightarrow R_+$ is decreasing for $j = 1, 2$. A more tractable expression is the exponential specification, with the form:

$$S(\chi_{g,t-k}, \kappa_{g,t}) = exp\{-\delta[(d_g - d_{g,t-k})^2 + (c_t - c_{t-k})^2]\}$$

where $\delta \geq 0$ captures the importance of similarity in recall. The likelihood of past experiences being recalled is a function of how similar/relevant they are. Similarity is a function of how close the past experience was to the discriminatory taste of the employer (d_g), and the contextual relatedness (c_t). It follows that the weight assigned to memory $\chi_{g,t-k}$ after the cue $\kappa_{g,t}$ is given is as follows:

$$w(\chi_{g,t-k}, \kappa_{g,t}) = \frac{S(\chi_{g,t-k}, \kappa_{g,t})}{\int S(\tilde{e}, \kappa_{g,t}) dF(\tilde{e})}$$

where $F(\tilde{e})$ captures the entire distribution of past experiences. These experiences are not group-specific and do not have to be exclusively domain-specific: for example, discrimination perceived in the housing market can be a relevant past experience even if the cue $\kappa_{g,t}$ refers to a labor market situation. Aggregating over past memories results in the memory-based discrimination taste of cue

[56] For simplicity, I drop the price (q) and quantity (q) cues, which were included in Bordalo et al. (2020), from the equations. I therefore focus only on the group affiliation and context. In line with Bordalo et al. (2020), context also captures non-hedonic attributes, such as the timing of the experience.

κ_t:[57]

$$d_g(\kappa_{g,t}) \equiv \int d_{g,t-k} w(\chi_{g,t-k}, \kappa_{g,t}) dF(\tilde{e})$$

Bias in the recall of memories across groups g can occur for two distinct reasons:

1. Fewer memories for a particular group, $F(\tilde{e}_g) \neq F(\tilde{e}'_g)$;
2. Differential recall of positive and negative experiences across groups due to differing discriminatory tastes $d_g \neq d_{g'}$.

The first bias can arise for a variety of reasons: self-fulfilling prophecies (Coate and Loury, 1993; Glover et al., 2017; Gagnon et al., 2025), limited experimentation (LePage, 2024), systemic discrimination (Bohren et al., 2025b), or the reliance on stereotypes (Miserocchi, 2023), among others. The differential availability of memories is likely particularly pronounced in cases where one of the two groups is a minority with whom interaction is limited.

The second bias highlights how, even in the case of an identical underlying distribution of experiences across both groups ($F(\tilde{e}_g) = F(\tilde{e}'_g)$), the weight assigned to memories differs across groups g and g'. This is because the similarity of, and hence weight assigned to, memories is a function of the employer's group-specific distaste parameter d_g. Without loss of generality, we assume that for a given employer e, the fixed distaste d_g is greater for group B than group A: $d_A < d_B$. When recalling memories, even if the memory set for both groups is identical ($F(\tilde{e}_A) = F(\tilde{e}_B)$), the employer is more likely to recall memories in line with their discrimination taste, d_g. Given $d_A < d_B$, the employer will recall more negative past experiences of Group B individuals than Group A individuals, skewing the memory recall to be more negative against Group B individuals. This results in a larger memory-based discrimination parameter ($d_b(\kappa_{b,t}) > d_a(\kappa_{a,t})$, and hence stronger discriminatory behavior in the current decision with cue $\kappa_{g,t}$.

[57] This specification does not explicitly differentiate between the timing of when occurred, as timing is implicitly captured in the context parameter, c_t. This can be addressed in two ways: by including the duration distance in the similarity function, of by incorporating a time-specific weighting function, as in Malmendier and Wachter (2024).

M Memory Recall - Online Experiment

Participants were shown 10 allocations of a manager to two workers. The 10 allocations are the following. (W) and (B) denote a White- and Black-sounding name, respectively.

Round	Manager Name	Worker#1 Name	Worker#2 Name	Allocation: Worker#1	Allocation: Worker#2
1	Brendan (W)	Joshua (W)	Marquis (B)	2	6
2	Matthew (W)	Terrance (B)	Jay (W)	4	4
3	Jacob (W)	Adam (W)	Reginald (B)	3	5
4	Nathan (W)	Tyrone (B)	Scott (W)	3	5
5	Jeremy (W)	John (W)	Donnell (B)	6	2
6	DeAndre (B)	Tremayne (B)	Justin (W)	6	2
7	Terrell (B)	Neil (W)	Demarcus (B)	3	5
8	Lamarion (B)	Maurice (B)	Geoffrey (W)	4	4
9	Antwan (B)	Robert (W)	Devonte (B)	5	3
10	Jermaine (B)	Rasheed	Daniel (W)	2	6

Table A30: Rounds Shown to Participants for Memory Recall

N Theory Model: Anticipated Discrimination

Following Buchmann et al. (2024), worker i supplies labor if the expected utility from working is weakly greater than their outside option, which is normalized to zero. While the focus is on the extensive margin of labor supply (working vs. not), insights directly translate to the intensive margin. Workers of group g receive a wage from their employer equivalent to w_g. Expected costs of working for employer k are defined as c_{igk}, and the disutility of working $u_i(.)$ is continuously differentiable and $\frac{\partial u_i}{\partial c_{igk}} > 0$, with $u_i(0) = 0$.

The cost of working for worker i of group g for employer k is c_{igk}, which is a linear combination of the group-specific costs c_g, individual-specific costs c_i, and employer-group specific costs, c_{gk}: $c_{igk} = c_i + c_g + c_{gk}$. Employer-group specific costs are unknown to workers, however workers form beliefs based on the group g of the employer, as well as their past experiences. In particular, employer-group specific costs are larger when the employer is of a different group than the worker ($k_g \neq i_g$). Empirical support for worker's preference to work for employers of a similar background to theirs comes from Hellerstein and Neumark (2008) and Giuliano et al. (2009).

Assuming that the outside option is zero, and job applications are costless, worker i of group g only supplies labor if:

$$\mathbb{E}[w_g - u_i(c_{igk})] \geq 0$$

The worker's expectation of c_{gk} also depends on their past experiences with employers of the same group as employer k. In particular, negative past experiences with employers of the same group g as employer k increase the worker's expectations of the employer-group specific costs: $\frac{\partial c_{gk}}{\partial \chi_{g,k,t}} > 0$. As such, negative past experiences with an employer of the same group g as employer k can increase the employer-group specific costs for worker i (c_{gk}), and hence the cost of working c_{igk}. Given $u_i(.)$ is monotonically increasing in c_{igk}, this increases the expected disutility from work, and hence reduces the labor supplied by worker i. This illustrates how retaliatory discrimination (and negative past experiences) can provide a micro-foundation for anticipated discrimination.

O Theory Model: Future Rounds

In each round, two players of different groups ($g \in \{A, B\}$) are randomly paired. Each player only plays the game once. Players can take one of two actions: they can either discriminate (D), or not discriminate (N). Akin to a prisoners dilemma, both players discriminating generates the worst outcome:

	Opponent: N	**Opponent: D**
Player: N	$(R,\, R)$	$(S,\, T)$
Player: D	$(T,\, S)$	$(P,\, P)$

Table A31: Payoffs in the Discrimination Prisoner's Dilemma.

with $T > R > P > S$ and $2R > T + S$. If the game is a one-stage game, D is the dominant strategy, and hence both the Player and Opponent will discriminate, resulting in payoff P.

Beliefs and Learning: Considering now that the game will be played across multiple rounds (but each player only plays once themselves), each group h holds a belief $\mu_{t,g}$ about the probability that a member of group g discriminates. Beliefs are updated via a Beta distribution prior with mass $m = \alpha_g + \beta_g$, where α_g is interpreted as the "prior number of discriminatory acts" observed from group g, while β_g can be interpreted as the "prior number of cooperative (non-discriminatory) acts" observed from group g. After observing one new action $x \in \{0, 1\}$ (with $x = 1$ if D), the posterior mean is

$$\mu_{t+1,g} = \frac{m\mu_{t,g} + x}{m + 1}.$$

In expectation, $x = \sigma(\mu_t^g)$, where σ is the strategy of group g.

Social Preferences and Reputational Cost: Individuals care not only about their own payoff but also about the expected payoffs of future in-group members. Let $\lambda \geq 0$ be the weight on in-group welfare, $\delta \in (0, 1)$ the discount factor, and $r > 0$ the retaliation strength (how strongly the out-group responds to the discriminatory reputation of a group). The effective reputational cost C of discriminating is

$$C \;\equiv\; \lambda \cdot \frac{\delta}{1 - \delta} \cdot \frac{r}{m + 1}.$$

The cost depends positively on λ, δ, r, and depends negatively on m, the strength of the prior (and hence number of observations/experiences).

Best Responses: The myopic, one-round gain from discrimination against belief p is

$$G(p) = (1-p)(T-R) + p(P-S).$$

However, if the game is played over multiple rounds, an individual discriminates if and only if $G(p) \geq C$. This defines a cutoff belief

$$p^\star = \frac{(T-R)-C}{(T-R)-(P-S)}, \qquad p^\star \in [0,1].$$

If $p > p^\star$, the player discriminates, while if $p < p^\star$ the player does not discriminate, and the player is indifferent if $p = p^\star$

Assuming that each group's discrimination probability is a linear response:$\sigma(\mu) = \beta\mu + (1-\beta)p^\star, \qquad \beta \in (0,1)$, belief dynamics then follow: $\mu_{t+1} = \frac{m\mu_t + \sigma(\mu_t)}{m+1}$.

This leads to the following proposition:

Proposition 1. *The unique symmetric steady state of the belief dynamics is $\mu^* = p^\star$. Convergence is linear with rate $\frac{m+\beta}{m+1} \in (0,1)$.*

The proof is available upon request. We therefore have the following comparative statics:

$$\frac{\partial \mu^*}{\partial \lambda} < 0, \qquad \frac{\partial \mu^*}{\partial \delta} < 0, \qquad \frac{\partial \mu^*}{\partial r} < 0, \qquad \frac{\partial \mu^*}{\partial m} > 0.$$

Hence, stronger social preferences for in-group members (λ), discount rate (δ), or retaliation (r) reduce steady-state discrimination, while stronger priors (m) increase it by making reputations less responsive to individual actions.

The *Future Rounds* treatment changed the game from a one-stage to a multiple-rounds game, and as a result introduced a non-zero discount factor. If participants did not have social preferences for in-group members ($\lambda = 0$), we would not expect treatment differences between the *Status Quo* and *Future Rounds* treatment arms. However, we document differences between the two treatment arms, providing further support for the role of social preferences as a micro-foundation for retaliatory discrimination.

P Research Transparency

In this appendix, I outline deviations from the pre-registration, and provide justifications.

Lab-in-the-Field in Uganda

- I pre-registered that I would also include a third stage in the experiment: participants would have 7 minutes time to create as many envelopes as possible. A Ugandan manager would then be shown a list of names and the number of envelopes made in 7 minutes, before deciding who to hire for a larger task. The aim was to document retaliatory discrimination as a micro-foundation for anticipated discrimination. However, after piloting, feedback from the enumerators was that the additional task took too long, and was too cumbersome for participants. Hence this leg of the experiment was dropped, and instead incorporated in the online experiment.
- Some questions were dropped from the pre-experimental questionnaire after the pilot, due to their sensitive nature. An example is "The Ugandan government discriminates against me because I am an Eritrean."

Prolific Experiment Among American Men

-

Q LLM Usage Disclosure

I used ChatGPT for two tasks:

1. Assistance with coding and data cleaning
2. Correcting grammar mistakes, using the following prompt:

> Proofread the attached PDF for objective grammar/punctuation only (no style rewrites). Use American English. Output an excel table with columns: Location | Old | New (bold changes) | Rule. Sentence-level, minimal edits. Don't alter equations, citations, LaTeX, quotes, or meaning. Fix only clear errors (agreement, articles, prepositions, punctuation, capitalization, parallelism, line-break hyphens, typos). After the table, list up to 5 common error patters and a short "Flags (ambiguous/Meaning-Dependent)" list if needed. If a page has no issues, state "No issues on p. X."