Discrimination as Retaliation[*]

Till Wicker

November 3, 2025

Job Market Paper

[Link to latest version]

Abstract

Discrimination remains pervasive, yet little is known about how past personal experiences of discrimination shape one's future discriminatory behavior. This paper introduces and empirically documents *retaliatory discrimination*: a form of 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 conceptual framework that situates retaliatory discrimination alongside taste-based and statistical discrimination, I conduct experiments in Uganda and the United States. In a two-stage experiment, participants are first randomly exposed to fair or unfair task allocations from managers of varying identities: co-ethnic, non-coethnic, or neither (computer-assigned). In the second stage, I observe whether they discriminate against non-coethnic workers when placed in a managerial role. Experiencing unfair task allocations from a non-coethnic manager increases subsequent discrimination against non-coethnic workers by 78%, reducing their earnings by 15%. This effect is driven both by an increase in the number of discriminators and the intensity of discrimination. I distinguish between four pre-registered micro-foundations of retaliatory discrimination and find empirical support for motivated beliefs: participants selectively interpret unfair task allocations as discriminatory to justify retaliation. The experiments also illustrate how past experiences affect expectations of future discrimination, offering a behavioral foundation for anticipated discrimination. Finally, I show that retaliatory discrimination has meaningful policy implications: in a complementary experiment, the removal of affirmative action policies triggers a backlash that amplifies discrimination, in contrast to predictions of standard economic models of discrimination.

1 Introduction

Discrimination has been documented across many domains in both developed and developing countries (Lang and Lehmann, 2012; Bertrand and Duflo, 2017; Neumark, 2018). Individuals from both minority and majority groups also perceive widespread discrimination against their own in-group: in the USA, 24% of Black and Hispanic workers and 13% of White workers reported experiencing discrimination (NPR, 2017; Gallup, 2021).[1] However, our understanding of how perceived discrimination shapes future discriminatory behavior is limited.

The two workhorse models of discrimination, taste-based and statistical, do not allow for perceived discrimination to affect future discriminatory behavior. Taste-based discrimination argues that discrimination arises due to fixed prejudicial preferences (Becker, 1957), while statistical discrimination posits that discrimination arises from information asymmetries (Arrow, 1972a,b; Phelps, 1972). However, this stands in contrast to empirical evidence documenting (i) that perceived discrimination can affect subsequent behavior (Gagnon et al., 2025; Ruebeck, 2025), and (ii) that discrimination can be reactive, for example increasing after ethnic riots and terrorist attacks (Kaushal et al., 2007; Hjort, 2014; Shayo and Zussman, 2017; Fisman et al., 2020).

This paper introduces and empirically documents *retaliatory discrimination*: individuals increase discriminatory behavior toward a group after perceiving discrimination from its members. I develop a conceptual framework that endogenizes discriminatory preferences, making them a function of an individual's prior experiences. I then test the framework using two experiments that identify and quantify retaliatory discrimination in two distinct settings: among Eritrean refugees in Uganda, and Black and White men in the USA. Finally, extensions of the experiments highlight the broader implications of prior experiences and retaliatory discrimination on anticipated discrimination and the removal of affirmative action policies.

The conceptual framework combines retaliatory, taste-based, and statistical discrimination. In the framework, discriminatory preferences depend on both exogenous prejudicial tastes (as in Becker 1957) and past experiences: negative interactions with a group increase

---

[1] 49% and 61% of Black and White Americans perceiving discrimination say the larger problem is discrimination based on the prejudice of individual people, rather than due to laws and government policies (NPR, 2017).

animus towards its members, heightening future discrimination. Consequently, discriminatory preferences are endogenous. Importantly, the effect of past experiences on discriminatory preferences is group-specific: bias intensifies only toward the group involved, not others. I subsequently test these two propositions about how past experiences shape future discriminatory behavior through two experiments.

The first experiment, conducted among Eritrean refugees in Uganda, provides empirical support for retaliatory discrimination.[2] In both stages of the experiment, a manager must delegate eight tasks between two workers, who are paid a piece rate per completed task. In the first stage, participants are assigned the role of a worker and are paired with a Ugandan worker. I exogenously vary whether their manager is Ugandan or a Computer, and whether the manager allocates tasks either equally between the two workers or more tasks are given to the Ugandan worker. In the second stage, participants assume the managerial role, and allocate tasks between an Eritrean and a Ugandan worker. A key feature of the experimental design is that it holds taste-based and statistical discrimination fixed across treatment arms, allowing me to isolate retaliatory discrimination. By exogenously varying the source and intensity of prior negative experiences, I can examine how these affect subsequent discrimination, measured by the participant's allocation of tasks between the Eritrean and Ugandan workers in the second stage.[3]

Results show strong evidence of retaliatory discrimination. Discrimination increases by 78% when participants are randomly assigned a Ugandan manager who gives more tasks to the Ugandan worker in the experiment's first stage, compared to a treatment arm where the Ugandan manager divides the tasks fairly. This reduces the Ugandan worker's earnings in the second stage by 15%. The increase in discrimination reflects an expansion at the extensive margin (a 41% increase in the number of discriminators), and conditional on discriminating, an increase in the intensity of discrimination. In contrast, when the manager in the first stage was a computer who allocated more tasks to the Ugandan worker, subsequent discrimination against the Ugandan worker does not increase compared to when the Computer manager

---

[2]Uganda, with close to two million refugees and a progressive refugee policy (including the right to work and move freely), presents an excellent setting for this study: interactions between Ugandans and refugees are frequent, however discrimination is still widespread without the presence of hostility or violence (Loiacono and Silva Vargas, 2025).

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

fairly allocated the tasks.

An online experiment among White and Black American men reproduces results from the experiment in Uganda, increasing the finding's generalizability. The experimental design mirrors that of the experiment in Uganda, except that managers in the first stage were either coethnic or non-coethnic, and divided the tasks evenly or favored either the White or Black worker. Receiving less than half of the tasks from a non-coethnic manager in the first stage induces stronger subsequent discrimination, compared to cases when the first stage non-coethnic manager evenly splits the tasks. The non-coethnic worker's earnings fall by 6%, and retaliatory discrimination is again driven both by an increase in the number of discriminators and intensity of discrimination. Importantly, positive retaliation is not documented when the first stage non-coethnic manager assigns more tasks to the participant, highlighting an important asymmetry with respect to negative versus positive experiences.

The online experiment also distinguishes between four micro-foundations of retaliatory discrimination. I pre-registered that social preferences, Bayesian updating, memory recall, and motivated beliefs could underpin retaliatory discrimination, finding empirical support for the role of motivated beliefs. First, participants are more likely to interpret managerial in-group favoritism as discriminatory when the stage 1 manager is non-coethnic, but efficient when the manager is coethnic. Second, when there is uncertainty surrounding the ethnicity of the stage 1 manager, participants selectively interpret the manager's ethnicity from their actions in order to justify retaliatory discrimination. Additional design features of the two experiments rule out alternative explanations, including inaccurate statistical discrimination, tit-for-tat, in-group favoritism, anger, inequality aversion, and norm violations.

Within the online experiment, I show that past negative experiences can be a source of anticipated discrimination. After the two stages of the experiment, participants signal their productivity to a future hiring manager by correctly completing as many tasks as possible within 60 seconds. The future manager is non-coethnic, and participants are informed that the manager will see their name and productivity signal. Being randomly exposed to a non-coethnic manager in the first stage of the experiment who assigns the participant less than half of the tasks reduces subsequent effort in the real effort task: participants complete 12% fewer tasks.

The experiments document that perceived discrimination results in retaliatory discrimination, even when participants are unaware of the initial discriminator's motivation. I then

vary the salience of this motivation by informing participants that the stage-1 manager's decisions were guided by an affirmative-action policy that is subsequently removed. When White men receive this information, they discriminate substantially more against a Black worker in stage 2 compared to a treatment arm that provides no justification for the stage-1 managerial decision. Experiencing discrimination as a result of affirmative action policies does not increase the number of discriminators, but increases the intensity of discrimination by 51% among those who were already discriminating. The findings highlight the importance of correctly identifying the source of discrimination for policies (Bohren et al., 2025a), particularly given the ongoing widespread reversal of affirmative action policies in the public and private sector (Guardian, 2025).

Finally, the experiments also provide suggestive evidence on mitigation measures to reduce retaliatory discrimination. A sub-treatment highlighting the salience of future interactions, and hence the consequences of current discrimination, increases the number of tasks allocated to the non-coethnic worker in the second stage of the experiment. Compared to a treatment arm that does not mention the existence of future rounds, the non-coethnic worker's payoff increases by 3%, and the number of discriminators decreases by 29%. However, neither of these differences are statistically significant.

Related literature This paper contributes to three strands of literature. First, I contribute to the theoretical literature on discrimination by identifying a new source of discrimination that differs from taste-based (Becker, 1957) and statistical discrimination (Arrow, 1972a,b; Phelps, 1972). Retaliatory discrimination differs from taste-based discrimination by modeling prejudice as endogenous and thus evolving in response to past experiences.[4] It differs from statistical discrimination as retaliatory discrimination does not arise due to imperfect information. Nevertheless, the channel through which past experiences shape future discriminatory preferences and behaviors mirrors experience-based discrimination (Lepage, 2024; Benson and Lepage, 2024), where past hiring experiences induce learning about group-level productivity, giving rise to (inaccurate) statistical discrimination. Complement-

---

[4]Experimental findings cannot be explained by an exogenous distaste parameter, rejecting the taste-based discrimination definition of Becker (1957) (see Section 3). Following Buchmann et al. (2024), I therefore consider retaliatory discrimination as an new source of discrimination. Alternatively, retaliatory discrimination can be interpreted as an endogenous behavioral foundation of prejudice, which is discussed more in Section 3.

ing experience-based discrimination, I show that past experiences can shape non-pecuniary costs in addition to providing information about worker- or group-level productivity.[5] Furthermore, I provide the first formalized economic framework of how past experiences shape prejudice. This framework offers a new behavioral explanation for the emergence and persistence of discriminatory tastes (Cain, 1986) that does not rely on group differences or comparisons (Bordalo et al., 2016; Esponda et al., 2023).[6]

Second, I contribute to the literature using lab and field settings to document discrimination and its determinants (Lang and Lehmann, 2012; Bertrand and Duflo, 2017; Neumark, 2018). The novel experimental design differs from existing studies by consisting of multiple interactions in which participants can both be the victim and perpetrator of discrimination. By holding taste-based and statistical discrimination fixed across treatment arms, the experimental design provides a direct test of retaliatory discrimination. The findings, which are in contrast to predictions of other discrimination models, highlight the importance of correctly identifying the source and nature of discrimination for policy recommendations (Bohren et al., 2025a). I illustrate this by experimentally showing that the removal of affirmative action policies can induce greater subsequent discrimination against non-coethnic workers. Additionally, I contribute to the empirical literature on 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, group-specific past experiences and anticipated discrimination.

Third, I contribute to the literature on the role of past experiences on economic decisions (Malmendier, 2021; Giuliano and Spilimbergo, 2025). While these studies look at how past macro-level events (such as financial crises, or riots) shape economic decisions, I focus on individual, micro-level experiences. Specifically, I examine how past experiences affect future discriminatory behavior. Retaliatory discrimination thus offers an alternative explanation for the emergence and persistence of inter-group tensions. The framework can be applied to microeconomic interactions (Hjort, 2014; Ghosh, 2025), and macro-level rela-

---

[5]Online Appendix B1 present a theoretical model of discrimination combining both experience-based (Lepage, 2024) and retaliatory discrimination, illustrating how past experiences can micro-found *both* statistical and taste-based discrimination.

[6]Retaliatory discrimination is intricately linked to the literatures in social psychology on vicarious retribution and group generalization, which shows that individuals often generalize negative encounters from one outgroup member to the entire group, fostering support for retaliation against that group as a whole (Lickel et al., 2006; Paolini et al., 2010; Barlow et al., 2012; Paolini et al., 2024).

tionships between ethnic divisions, conflict, and economic development (Alesina and Ferrara, 2005; Arbatlı et al., 2020).[7]

This paper proceeds as follows: Section 2 develops a conceptual framework that incorporates taste-based, statistical, and retaliatory discrimination, to formalize how past individual experiences can shape future discriminatory behavior. Sections 3 and 4 present results from experiments in Uganda and the USA that causally identify retaliatory discrimination, while keeping other sources of discrimination fixed. Section 5 distinguishes between four pre-registered micro-foundations of retaliatory discrimination, before Section 6 discusses two implications: the removal of affirmative action policies, and anticipated discrimination. Section 7 explores a potential measure to reduce retaliatory discrimination, and Section 8 concludes.

2 Conceptual Framework: Retaliatory Discrimination

I develop a conceptual framework that incorporates taste-based, statistical, and retaliatory discrimination to motivate the experiments in Sections 3 and 4. 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. Therefore, the conceptual framework discussed in this section is specific to the labor market. A more general framework of discrimination is presented in Appendix A2, 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).[8]

2.1 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:

---

[7]A related literature looks at the persistence of attitudes against (minority) groups (Schindler and Westcott, 2020; Bursztyn et al., 2024).

[8]The conceptual framework models individuals as myopic, abstracting away from future and strategic interactions. I present empirical support for this in Section 7.

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 $(\theta_A, \theta_B,$ 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, and the function f is weakly increasing in both the exogenous and endogenous variable.[9],[10] 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.

In particular, the employer's utility function is:

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

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

---

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

[10]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 is discussed in Section 5.

or against a group (as in Becker 1957), and an endogenous 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, $F(\chi_{g,t})$ 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.

To incorporate statistical discrimination within the hiring decision, employers do not observe the true productivity of a worker (θ) 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 know their productivity and 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), I 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.[11] Appendix A1 discusses in greater detail how equation (1) incorporates taste-based and statistical discrimination, while Online Appendix B1 extends the theoretical framework to allow for paternalistic discrimination (Buchmann et al., 2024) and experience-based discrimination (Lepage, 2024).

There are two 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-found statistical discrimination, by providing information about group level productivity, as discussed by Lepage (2024). The second channel is through non-pecuniary costs: in addition to the group-specific 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 channel does not stem from the updating of beliefs about worker productivity but from the updating of non-pecuniary costs. The coexistence of these two mechanisms means that discrimination can persist and evolve even when employers have accurate information about

---

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

productivity distributions.

Defining Discrimination Discrimination is defined as $D_t(s, \psi_A, \psi_B) \equiv L_{A,t}(s) - L_{B,t}(s)$, 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 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 Propositions About Role of Past Experiences

The dynamic term $F(\chi_{g,t})$ implies that employers' non-pecuniary costs are not fixed, but evolve in response to past interactions. Here, I make two propositions about how past experiences shape the employer's non-pecuniary costs, and discriminatory behavior.

Without loss of generality, I assume that the employer discriminates against workers from group B, and hence $D_t(s, \psi_A, \psi_B) > 0$. Suppose an employer has a negative experience with an individual from group B. The first proposition is that 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 workers from groups A and B send identical productivity signals. The second proposition is that, because $F(\chi_{g,t})$ is defined separately for each group, these effects are group-specific. The two propositions are formalized below:

Proposition 1: (*Retaliatory Discrimination*) Ceteris paribus, more negative past experiences with workers from group B at time t ($\chi_{B,t}^{\text{mod}} < \chi_{B,t}^{\text{neg}}$) have a non-negative effect on discrimination against workers of group B at time t:

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

See Appendix A3 for the proof. If $\frac{\partial f}{\partial F(\chi_{g,t})} > 0$ (strict inequality), then more negative past experiences with workers from group B will strictly *increase* the employer's discrimination against workers of group B.

Proposition 2: (*Group-Specific Retaliatory Discrimination*) Ceteris paribus, more negative past experiences with individuals from group $g' \notin \{A, B\}$ at time t ($\chi_{g',t}^{\text{mod}} < \chi_{g',t}^{\text{neg}}$) have no

effect on discrimination against workers of group B relative to workers from group A:

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

See the proof in Appendix A3. These two propositions capture that past interactions can affect current discriminatory behavior, however these past interactions, and hence their consequences, are group-specific.

3 Main Experiment: Uganda

To test the empirical validity of retaliatory discrimination, 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. The progressive policy, coupled with Eritreans comparable levels of education and and network-based hiring, ensures that Eritrean refugees have similar economic opportunities to Ugandans in Kampala. 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. Most 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. 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 of Eritrean refugees, who have an average of only 2.52 Ugandan friends (see the Online Appendix Table B3), and 23.21% of whom felt that Ugandan firms discriminated against them.[12]

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.

---

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

3.1 Experimental Design

Figure 1 depicts the experimental design, which consisted of two stages. In both stages, a manager delegates 8 tasks between two workers. The manager is paid a fixed wage, however workers are paid a piece rate of 500 UGX per completed task.[13] Workers and managers were given alias names that revealed their nationality, but preserved their anonymity.[14]

The task consisted of making an envelope (used for a cash transfer, as in Wicker et al. 2025) 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.

Before the first stage of the experiment, 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) 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 Online Appendix B2). 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 experiment.

In the first stage of the experiment, the Eritrean participant ($\mathbf{E_1}$ in Figure 1) is assigned the role of one of the two workers. They are informed that they are paired with a Ugandan male worker U_1 (signaled by their name), and that a manager has decided the allocation of the eight tasks across the two workers. Experimental variation comes in the nature of the manager in stage 1 (M_0 in Figure 1): the manager is (i) either a Computer or a Ugandan, and (ii) either divides the eight tasks evenly across the two workers (4, 4); or assigns more tasks to the Ugandan worker (6, 2). Allocations are pre-programmed, and based on actual

---

[13]500 UGX ≈ $0.14. Additionally, participants received 2000 UGX as a show-up fee. Average compensation was 3500 UGX (equaled half a day's worth of wages), and the study lasted 20 minutes on average. Each envelope (which took less than 2 minutes per envelope) corresponded to 37 minutes of work at the minimum wage.

[14]Online Appendix B2 illustrates that during pilot work, both Eritreans and Ugandans were able to correctly identify the nationality of an individual based on the revealed name in 97% of cases.

M_0 Allocations (U_1 ; E_1)			
Computer Manager		Ugandan Manager	
(6 ; 2)	(4 ; 4)	(6 ; 2)	(4 ; 4)
T1	T2	T3	T4
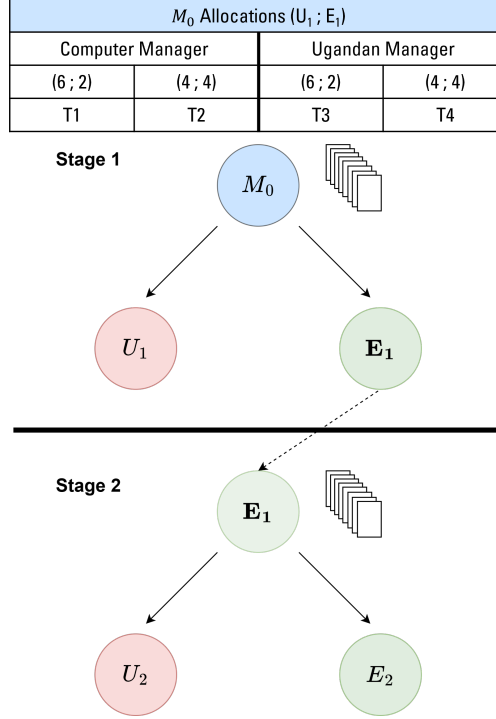
Figure 1. Experimental Design: Lab-in-the-Field in Uganda

Notes: The figure shows the experimental design of the study among Eritrean refugees in Uganda. Across both stages of the game, a manager (who is paid a flat wage) delegates eight tasks between two workers (who are paid a piece rate per completed task): a Ugandan and an Eritrean. The participant in the study is E_1, and thus is a worker in the first stage of the game, and becomes a manager in the second stage. Exogenous variation is introduced in the form of the manager in the first stage (M_0), who is either a Computer or a Ugandan, and either allocates the tasks such that they favor the Ugandan worker, or split the tasks evenly. This results in four treatment arms (T1—T4).

decisions made by Ugandans during the pilot study. As such, there are four treatment arms, as depicted in Figure 1.

Once participants learn how many tasks they had been assigned by their manager, they make the envelopes. The enumerator records how long it takes the participants to make the envelopes, and after the data collection was completed, the enumerators evaluated the quality of the envelopes based on five dimensions.[15]

The first stage of the experiment finishes once the participant is done making the envelopes, after which the second stage of the experiment commenced. Importantly, the participant does not receive any feedback regarding the quality of the envelopes they, or

---

[15]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.

their paired Ugandan worker (U_1), made. Therefore, the information set available to the participant regarding the relative productivity of Ugandans and Eritreans does not change throughout stage 1, nor across the four treatment arms.

The set-up of the second stage is identical to the first stage, except that this time, the Eritrean participant ($\mathbf{E_1}$) is the manager who has to delegate eight tasks between two male workers: one Ugandan (U_2) and one Eritrean refugee (E_2). Neither the participant, nor their previous manager (M_0), has interacted with either of the two workers before. In this stage, the Eritrean participant ($\mathbf{E_1}$) is paid a flat wage, while workers are 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: any deviation from an equal split of the eight 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 Proposition 1 of Section 2, predicts that participants randomly assigned to a Ugandan stage 1 manager who allocates fewer than half the 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 across both workers (T4). Proposition 2 argues that a negative experience with the Computer manager in stage 1 (T1) will not affect stage 2 allocations compared to when the Computer manager allocates tasks evenly (T2).

Appendix A4 presents detailed theoretical predictions of taste-based, statistical, and retaliatory discrimination, as well as other explanations (including paternalistic discrimina-

tion, systemic discrimination, social norms, fairness concerns, and experimenter preferences). None of the other 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) as the manager in the second stage. The participant had to divide eight tasks, and hence allocating four 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, averaging 3.49 tasks ($p < 0.001$). This suggests some degree of discrimination against the Ugandan worker. 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 of tasks across workers is due to taste-based discrimination, statistical discrimination, or alternative explanations (e.g. fairness considerations, see Appendix A4).

When the Computer is the stage 1 manager (referring to T1 and T2, the two bars on the left-hand-side in Figure 2), the allocation of tasks in the second stage does not differ depending on whether the participant was allocated two or four out of the eight tasks in the first stage (T1 vs. T2, $p = 0.389$). This is in line with Proposition 2, as unrelated past experiences do not affect current discriminatory actions.
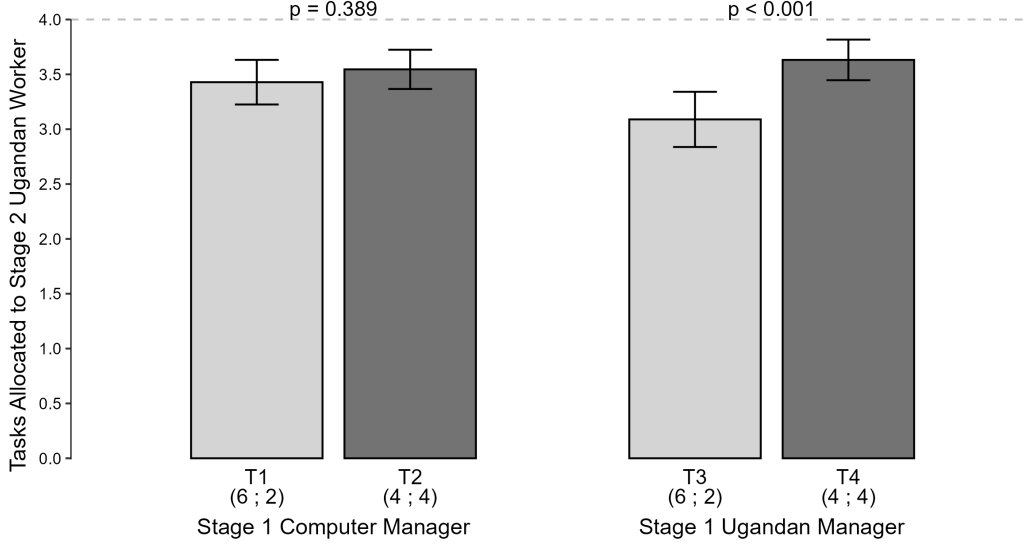
Figure 2. Task Allocation to Ugandan Worker in Stage 2 (U_2)

Notes: The Figure shows the number of tasks allocated to the Ugandan worker in stage 2 of the experiment (U_2) by the Eritrean participant (E_1). Allocating four out of the eight tasks indicates the case of no discrimination, indicated by the dashed gray line. The Figure reports average task allocations to the Ugandan worker across the four treatment arms (T1—T4), including 95% confidence intervals. P-values are based on two-sided t-tests.

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, however this difference is not statistically significant (T1 & T2 vs. T4, $p = 0.199$).[16] However, 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 nor worker (M_0 and U_1). Compared to when the Ugandan manager in the first stage evenly splits the tasks, this difference is highly statistically significant (T3 vs. T4, 78%, $p < 0.001$), and lowers the Ugandan worker's earnings by 15%. This allocation is also statistically significantly different compared to when the Computer manager allocated two

---

[16]Eritreans had different prior expectations about how many tasks 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 did not think the computer was biased ($p = 0.206$).

tasks to the participant in the first stage (T1 vs. T3, $p = 0.038$).[17] This provides support for Proposition 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 intensity 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 when their previous Ugandan manager treated them fairly (T4). When their previous Ugandan manager treated them unfairly (T3), this number jumps to 57.14%, a 17pp increase (41%, $p = 0.075$, see Appendix Table A5). Furthermore, conditional on discriminating, individuals in T3 discriminate more aggressively on average, allocating 2.50 tasks to the Ugandan worker in stage 2, compared to 2.91 tasks in T4 (46% increase in discrimination, $p < 0.001$, see Appendix Table A5).[18] This indicates that not only does retaliatory discrimination create new discriminators, but also increases the intensity of discrimination.

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 be reflective of reduced effort in response to perceived discrimination (Gagnon et al., 2025; Ruebeck, 2025).

Appendix Table A3 illustrates that the quality of the envelopes, measured along five pre-registered quality measures, decreases by $\sim 20\%$ as a result of having a Ugandan manager who assigns fewer tasks, compared with when tasks are divided evenly (T3 vs. T4). However, this difference is not statistically significant ($p = 0.159$). This provides suggestive evidence

---

[17]The regression tables underlying Figure 2 are presented in Appendix Table A1.
[18]Appendix A6.1 presents histograms of the allocations to the Ugandan worker in stage 2, across treatments 1-4.

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 placed in a consequential decision-making role. Relatedly, in line with Gagnon et al. (2025), who find that workers put less effort after perceiving discrimination, Appendix Table A2 documents statistically treatment effects on the worker's effort, defined as the time taken to complete the envelopes.[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 A4). Instead, results from the experiment align with Propositions 1 and 2 of retaliatory discrimination outlined in 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 A4 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 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 receive fewer tasks than they expected from a Ugandan manager in the first stage, they retaliate more strongly 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 Online Appendix B2.3), suggests motivated beliefs about the reasoning behind manager's choices are an important micro-foundation of retaliatory discrimination. This is discussed more in the next Section 5. Appendix A8 illustrates how retaliatory discrimination can be misinterpreted as taste-based discrimination, through a survey of 51 academics.

---

[19]Online Appendix Tables B8-B12 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.

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 copy a randomly generated sequence of letters and numbers. Secondly, the nature of the discrimination differs: workers and managers either have distinctively White or Black names.[21] Thirdly, participants are both White and Black American men, and thus participants belong to both the majority and minority group.[22] Fourth, the allocation of the eight tasks in stage 1 of the experiment are made by either a coethnic or non-coethnic manager and either favor the participant, equally split the tasks, or favor the other worker. Appendix A5 outlines the motivation for each of these design choices.
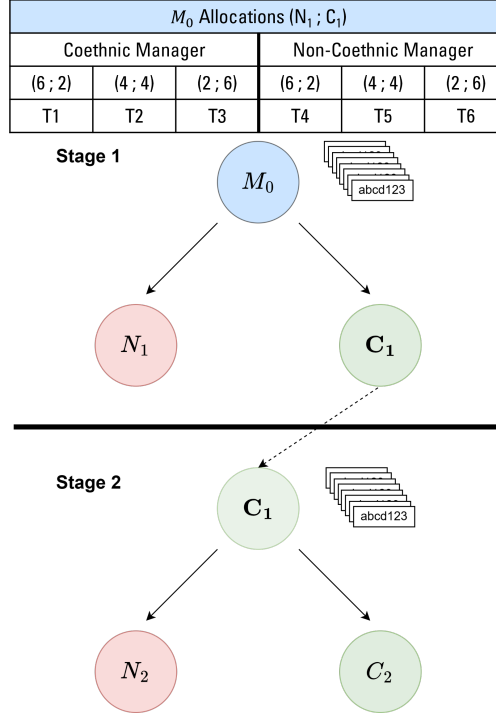
Otherwise, the experimental design mirrors that of the experiment in Uganda, with participants being a worker in the first stage before becoming a manager in the second stage. As such, the experiment consists of 6 treatment arms across which participants are randomized, as depicted in Figure 3. In treatment 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, non-coethnic worker (T1, T4), split the tasks evenly between the two workers (T2, T5), or favor the participant (T3, T6).

---

[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).

[22]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 Online Appendix Table B4).

| M_0 Allocations (N_1 ; C_1)					
Coethnic Manager			Non-Coethnic Manager		
(6 ; 2)	(4 ; 4)	(2 ; 6)	(6 ; 2)	(4 ; 4)	(2 ; 6)
T1	T2	T3	T4	T5	T6

Figure 3. Experimental Design: Online Experiment in the USA

Notes: The figure shows the experimental design of the study among White and Black American men. Across both stages of the game, a manager (who is paid a flat wage) must delegate eight tasks between a and non-coethnic worker (who are paid a piece rate. The participant in the study is C_1, and thus is a worker in the first stage of the game, and becomes a manager in the second stage. Exogenous variation is introduced in the form of the manager in the first stage (M_0), who is either a coethnic or a non-coethnic, and either allocates the tasks such that they favor either worker, or split the tasks evenly. This results in six treatment arms (T1—T6).

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. For five out of the six experimental arms, there is on average no discrimination in the allocation of tasks across the two workers ($p = 0.239 - 1.000$). However, in T4, where the participants had a non-coethnic manager that only allocated two out of the eight tasks to them, participants retaliate against a non-coethnic worker, by assigning them 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 allocates the tasks across the two workers (T4 vs. T5, $p = 0.008$), and the treatment

arm where their previous manager is coethnic, but only assigns them two of the eight tasks (T1 vs. T4, $p = 0.019$).
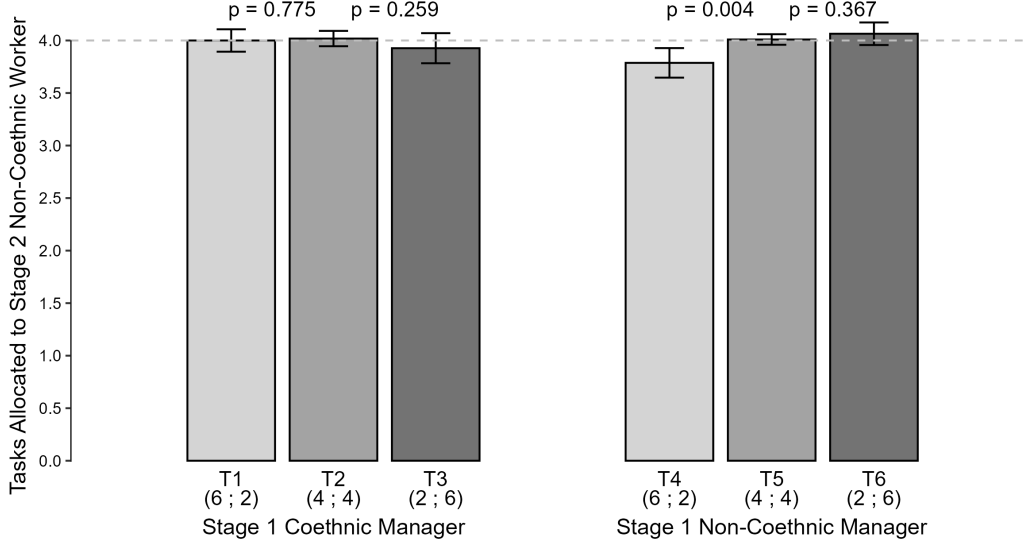
Figure 4. Task Allocation to Non-Coethnic Worker in Stage 2

Notes: The figure shows the number of tasks allocated to the Non-Coethnic worker in stage 2 of the experiment (N_2) by the participant (C_1). Allocating four out of the eight tasks indicates the case of no discrimination, indicated by the dashed gray line. The Figure reports average task allocations to the Non-Coethnic worker across the six treatment arms (T1—T6), including 95% confidence intervals. P-values are based on two-sided t-tests. See Online Appendix B11 to see heterogeneity by race and discriminatory attitudes.

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) — allocates six of the eight tasks to the participant. 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). Hence, 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 differ between members of a minority and majority group. Decomposing Figure 4 indicates that retaliatory discrimination is particularly pronounced among Black men, compared to White men (see Online Appendix Table B13). While White men on average assign 3.95

21

tasks to a worker with a Black-sounding name in T4 ($p = 0.370$), Black workers allocate 3.62 tasks to a worker with a White-sounding name in T4. This is substantially less than the setting of no discrimination ($p = 0.005$). The different allocations between White and Black participants is only statistically significant in T4 ($p = 0.019$).

In line with the results in Section 3, we again observe that retaliatory discrimination is driven by an increase in the intensive and extensive margin of discrimination. The number of individuals discriminating increases from 3.0% in T5 to 15.7% in T4 ($p < 0.001$, see Appendix Table A8). Conditional on discriminating, individuals discriminate more strongly in T4, compared to when the stage 1 non-coethnic manager fairly allocated tasks ($p < 0.001$, see Appendix Table A8).[23] This mirrors the results documented in Section 3.

4.2 Persistence of Retaliatory Discrimination

Thus far, I have documented an immediate retaliatory nature of ethnic discrimination through two experiments. 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 participants were randomized to complete a real-effort task, while the other half completed the real-effort task after the second round.[24] This results in variation in the time between the two stages of the experiment, akin to the "cooling off" literature in ultimatum games (Bosman et al., 2001). Controlling for whether participants first completed a real-effort task (which lasted ~ 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 A6 column (2)).

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

---

[23]Retaliatory discrimination is particularly pronounced for participants with below-median discriminatory attitudes ($p = 0.023$, see Online Appendix Table B14).

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

stage 2.[25] Participants again had to allocate eight tasks between a non-coethnic and coethnic worker. This allows me to test whether the managerial allocation decisions in stage one a week earlier still had an effect on the participant's own discriminatory behavior a week later.

Participants who were randomly assigned to T4 a week ago 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. Therefore, 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 both the name of their stage 1 manager and task allocation of the previous week, despite monetary incentives to do so.

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

4.3 Alternative Explanations

In this sub-section, 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 A9.

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 divides money between the two participants: a form of dictator game. As productivity does not affect allocation decisions, (inaccurate) statistical discrimination does not play a role in dictator games (List, 2006). Appendix Figure A6 illustrates that the retaliatory discrimination pattern documented in Figure 4 is replicated in the dictator game version of the experiment, ruling out accurate and inaccurate statistical discrimination as an alternative explanation.

---

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

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 only observed for T4. Furthermore, detailed beliefs were elicited from half of the participants; when asked to justify their allocation of tasks, none 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 could be attributed to in-group favoritism, participants randomized into T3 still discriminate statistically significantly more than participants 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 favored coethnic workers, and hence receive more tasks. Thirdly, no discrimination is documented in all treatment arms of the online experiment 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 experiment is identical to the online experiment of Figure 3, except that participant's group affiliation is arbitrarily determined (Red and Blue team) and a participant's ethnicity is not made salient. No discrimination, or retaliatory discrimination, is documented across the six treatment arms (see Appendix Figure A7). If reciprocity were driving the treatment effects documented in Sections 3 and 4, one would also except retaliatory discrimination to occur in the minimal group paradigm experiment (Rabin, 1993). Finally, tit-for-tat reciprocity is further ruled out as a potential mechanism by illustrating that participants retaliate more strongly ($p = 0.080$) when the non-coethnic worker in the second stage is their manager from stage 1 (direct retaliation), rather than an unrelated non-coethnic worker (retaliatory discrimination), see Appendix Table A16.

5 Micro-founding 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.[26] Extensions to the basic experiment outlined in Section 4 were designed to disentangle the underlying mechanisms. Support is found in favor of motivated beliefs, as participants selectively interpret managerial allocations in order to justify retaliation.

5.1 Memory

I find no empirical support that the recall of memories affects retaliatory discrimination.[27] 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 A10). This suggests 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. Participants were shown ten rounds of managers allocating eight tasks across a White and a Black worker. In five of the ten rounds, the manager was White, while in the other five rounds the manager was Black. Allocations of the managers — based on pilot data — are such that, on average, there was no discrimination by White or Black managers.[28] All participants are shown the same ten managerial allocations, in a randomized order. After recalling the managerial allocations (with financial incentives),

---

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

[27]Other studies have found that memories, and the biased recall of past memories, affects discrimination (Miserocchi, 2023).

[28]See Appendix A10 for an overview of the allocations.

participants are assigned the role of the manager and divide eight tasks between two workers. Mirroring stage 2 of the earlier experiments, one of the workers is White, and the other is Black. 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.

First, I find that participants more accurately recall allocations of tasks for rounds with a coethnic manager ($p = 0.068$, see Appendix Table A13 columns (1)-(2)), however this does not differ depending on whether the coethnic manager favored coethnic workers, or not.[29] 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 (see Appendix Table A13 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 (ii) a non-coethnic manager favors the non-coethnic worker (see Appendix Table A13 column (4)).[30]

Second, 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 A14).

These two findings — (i) the absence of an overall 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.

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 behavior towards other, similar individuals. Identity-dependent social preferences

---

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

[30]There is a statistically significant correlation between participants' discrimination index (elicited during the post-experimental questionnaire) and the number of previously allocated tasks recalled for White participants ($\rho = -0.110$, $p = 0.078$), but not for Black participants ($\rho = 0.058$, $p = 0.428$). This suggests that among White participants, those that had stronger discriminatory tendencies thought Black managers discriminated more against White workers. This provides further support for the role of motivated beliefs, discussed below.

could provide a micro-foundation for some of the documented results related to retaliatory discrimination, however these theoretical models have not yet been formalized. Furthermore, social preferences struggle to rationalize other findings, for example the effects of negative past experiences 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 of both experiments 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 *reduce* efficiency.

I document increased retaliation if participants can retaliate against their original manager, compared to when they can retaliate against a different worker (see Appendix Table A16), 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 affects social preferences, with participants being more altruistic towards in-group players. The differential degree of retaliatory discrimination between T1 and T4 of the online experiment ($p = 0.019$) — 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.[31] Hence, distributional social preferences that incorporate an individual's

---

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

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 formulation 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 towards other individuals with the same background or group identity. 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 intentions-based reciprocity, participants retaliate against an unrelated worker who shares the same identity as the initial manager.[32]

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.[33]

To gain relevant insights, half of the participants in the online experiment were asked

---

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

[33]By selecting a task that does not have an associated stereotype, I abstract away from stereotype-induced beliefs (Bordalo et al., 2019).

to state their beliefs throughout the experiment.[34] 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 believe that an even division of tasks is both fair (95.25%) and efficient (89.24%). Furthermore, 80.70% of participants believed other participants 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 Black 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 Black men perceived as a fair allocation in the other five treatment arms ($p = 0.059$). Perceiving fewer tasks allocated to the non-coethnic worker as fair is positively correlated with actually allocating fewer tasks to the non-coethnic worker, both for the whole sample ($\rho = 0.192$, $p < 0.001$), especially for participants 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

---

[34]There are no statistically significant differences between participants from whom beliefs were elicited versus not, see Online Appendix Table B5).

the participants 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.[35] 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. Participants in both the *Status Quo* and *Uncertain Manager* treatment arms were allocated two out of eight tasks in the first stage of the experiment. In the *Status Quo* treatment arm, the stage 1 manager was Black (equivalent to T4 of the main online experiment), while participants in the *Uncertain Manager* treatment arm were told that with 50% probability their stage 1 manager was Black and with 50% probability their stage 1 manager was White. Subsequently, participants completed the two assigned tasks and proceeded onto the 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, allowing participants to selectively interpret the managerial allocations in stage 1 in line with their prior beliefs. These motivated beliefs can subsequently 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 Black worker 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.

---

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

In the *Uncertain Manager* treatment arm, subjective posterior beliefs about the ethnicity of the manager in stage 1 are strongly correlated with their allocation of tasks in the second stage. The greater the subjective posterior 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 ($p < 0.001$, see Appendix Figure A4). This negative correlation 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.[36]

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 if they consider the probability of receiving an unfair allocation to be the same regardless of the manager's ethnicity. In that case, the allocation provides no differential information about the manager's type, and thus should have no effect on participants' discriminatory behavior as a manager in the second stage. 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). First, 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 the other worker versus 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).

Second, 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 tasks evenly, which is balanced across treatment arms (F-statistic $= 0.279$, $p = 0.924$). However, once the stage 1 manager divided the tasks,

---

[36]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).

justifications for these allocations differ across treatments. When the participant only re-
ceives two tasks from a non-coethnic manager, they cite the manager's ethnicity as a reason
in 25.58% of cases. This drops to 16.22% when the manager is a coethnic who only assigns
two tasks to the participant. 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 assigning them six tasks is a coethnic. Hence indi-
viduals 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.

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.
Instead, Online Appendix Table B14 illustrates that treatment effects are larger among
participants with below-median discriminatory tastes.

6 Implications of Retaliatory Discrimination

I illustrate the importance of past experiences and retaliatory discrimination through two
applications. First, I experimentally show that negative past experiences can give rise to
anticipated discrimination, and discuss the equilibrium consequences. Second, an experiment
simulating the reversal of affirmative action policies illustrates how retaliatory discrimination
can generate different policy conclusions than taste-based and statistical discrimination.

6.1 Micro-foundation for Anticipated Discrimination

A second extension of the role of negative past experiences on discriminatory behavior 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; 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 expectations 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 they could also affect expectations of future discrimination: 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.

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 for every sequence correctly completed.[37] Thus, the number of completed tasks is a measure of the effort the participant put into the "job application", and a signal of their productivity to the future manager.

The future manager is 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 the online experiment. This is because tastes are exogenous, and hence, with rational expectations, ones expectations of other people's tastes are also exogenous, and thus unaffected by their previous experiences with managers of the same ethnicity, provided that workers are fully informed about the population distribution of discriminatory tastes. 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 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

---

[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.

of the same and different ethnicity as the potential future manager.[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, $p = 0.030$). This presents experimental evidence that past experiences with individuals of a certain group can affect ones desire to work with individuals of the same group in the future, proxied through the effort put in to a real-effort task.[40]

Equilibrium Effects of Retaliatory Discrimination

The equilibrium effects of retaliatory discrimination as a source of 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 even if the workers are equally productive as coethnic workers. Based on these signals (referring to signal s in the conceptual framework of 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 the differential representativeness of the productivity signal of true productivity across different groups of applicants due to their past experiences.

---

[39]In Online Appendix B8, I extend the framework of retaliatory discrimination to beliefs, such that expectations of discriminatory tastes are a function of past experiences: negative past experiences with individuals of a certain group can increase an individual's expectations of prejudice among other individuals of the same group, and hence increase anticipated discrimination.

[40]I can rule out that treatment effects are driven by anger, as half of the participants are randomized to 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 ($p = 0.442$), see Appendix Table A9. Furthermore, results are not driven by less experience/practice, as otherwise one would expect negative treatment effects in T1 as well.
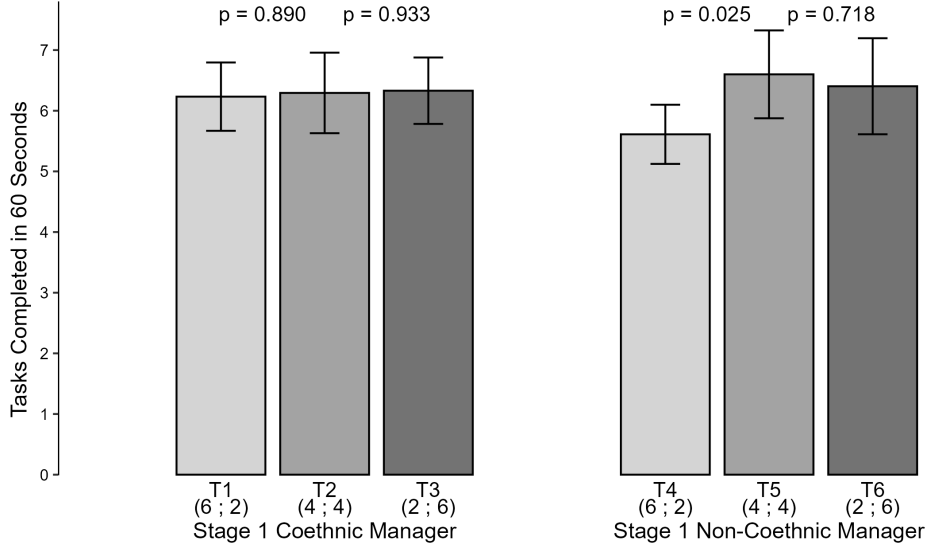
Figure 5. Tasks Completed in 60 Seconds as a Productivity Signal to a Non-Coethnic Manager

Notes: The figure shows the number of tasks completed by the participant (C_1) in a 60-second, real-effort task, which took place after the two stages of the experiment. The number of tasks participants completed in the 60 seconds, along with their pseudo-name, were shared with a non-coethnic hiring manager. The figure reports average tasks completed across the six treatment arms (T1—T6), including 95% confidence intervals. P-values are based on two-sided t-tests. See Online Appendix B11 to see heterogeneity by race and discriminatory attitudes.

6.2 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, several countries and organizations have been removing affirmative action policies and Diversity, Equity, and Inclusion (DEI) programs.[41]

Neither taste-based nor statistical discrimination predict that the introduction and removal of affirmative action policies would increase subsequent discrimination against minor-

---

[41]The White House also ordered all federal agencies to end any DEI programs as of January 2025 (The White House, 2025), and companies including Meta, Google, Amazon, and Disney have rolled back DEI policies (Guardian, 2025).

ity 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 discriminatory tastes are exogenous. 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 hired weakly more minority workers, and hence the information asymmetry about group-level productivity has weakly decreased, reducing discrimination.

Contrary to taste-based and statistical discrimination, retaliatory discrimination argues that the introduction and removal of affirmative action policies can amplify discrimination, 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). As such, AA and DEI policies can increase the number of negative past experiences with minorities, increasing prejudice as a result of pro-minority policies.[42] When affirmative action policies get removed, discrimination against minorities may subsequently actually increase.

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

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 Removal* condition, participants are informed that the manager's allocation of tasks across workers in the first stage are influenced by affirmative action policies, which have been removed before the second round.[43] Within this

---

[42]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."

[43]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 affir-

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

	Allocation of Tasks to Non-Coethnic Worker
	(1)
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 motivation of the stage 1 managerial allocations are not made explicit. Robust standard errors are in parentheses. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

experimental set-up, taste-based and statistical discrimination would not predict differences between the two treatment arms, while retaliatory discrimination would predict stronger retaliation, and hence greater discrimination, as a result of the presence of affirmative action policies in the past.

Table 1 presents the number of tasks allocated to the Black worker in the second stage of the experiment. Participants in the *Status Quo* treatment arm discriminate against the Black worker ($p = 0.072$). When the allocation of tasks in the first stage can be attributed to affirmative action policies that are subsequently abolished, the White manager retaliates more against the Black worker in the second stage of the experiment. In particular, participants in the *Affirmative Action Removal* treatment arm allocate 0.17 fewer tasks to the Black worker, equaling 0.32 standard deviations of the number of tasks allocated to the Black worker in the *Status Quo* treatment arm ($p = 0.085$).

This experiment demonstrates that retaliatory discrimination can generate policy im-

---

mative action policy has been abolished and no longer applies to your allocation decisions. You are free to distribute the tasks as you see fit."

plications that are distinct from both taste-based and statistical discrimination models. 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.

7   Mitigating Retaliatory Discrimination

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

I present one mitigating action for which I provide suggestive empirical support that it can reduce the degree of retaliatory discrimination, and hence overall discrimination: increasing the salience of future interactions.[44]

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 individual that is discriminated against in time t retaliates in future time periods.[45] 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

---

[44]I pre-registered two other mitigating measures: costly mistakes, and inefficiencies due to non-even allocation of tasks. Appendix Tables A11 and A12 illustrate that neither mitigated retaliatory discrimination.

[45]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).

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".[46] This is in contrast to predictions of taste-based and statistical 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.[47]

To experimentally investigate whether varying the salience of future rounds affects the extent to which they engage in discrimination, the online experiment is appended by an additional stage. 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 assigned two of the eight tasks, meaning that all participants are exposed to T4 of Figure 3. Afterwards, participants become the manager and allocate tasks between a White and a Black worker.

For a sub-set of the sample, I induce experimental variation in the salience of future rounds 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 you", participants in the *Future Rounds* treatment arm are informed that "there may 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)". This treatment arm thus 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 future payoff of other players, or decisions beyond the present round of the experiment, the *Future Rounds* treatment will have no effect on discriminatory behavior.

Table 2 illustrates that increasing the salience of future rounds increases the number of tasks allocated to the non-coethnic worker by 0.17 tasks, equal to 0.24 standard deviations of the division of tasks in the *Status Quo* treatment arm ($p = 0.098$). 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.

---

[46]See Online Appendix B9 for the mathematical foundations of this model.

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

Table 2: Future Rounds and Discriminatory Allocations

	Allocation of Tasks to Non-Coethnic Worker (1)
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, see Appendix Figure A2). Robust standard errors are in parentheses. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

8 Conclusion

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

Through experiments in Uganda and the USA, I empirically document retaliatory discrimination: individuals discriminate more against members of a group after previously perceiving discrimination from a member of that group. However, this retaliatory discrimination is group-specific. The new source of discrimination induces *new* discriminators as well as intensifying the intensive margin of discrimination. I distinguish between four pre-registered micro-foundations of retaliatory discrimination, finding empirical support for the role of motivated beliefs: participants interpret unfair task allocations as discriminatory in order to justify retaliation.

Identifying the right source of discrimination can have implications for policies (Bohren et al., 2025a), which also holds true for retaliatory discrimination. I illustrate this through

an experimental twist simulating the removal of affirmative action policies, which I find leads to heightened discrimination compared to a control condition where the manager's motive was unspecified. This finding is in contrast to predictions of taste-based and statistical discrimination, however in line with retaliatory discrimination.

Negative past experiences can not only affect one's own future discriminatory behavior, but also shape the expectations of discriminatory behavior of others. As such, negative past experiences can also be a micro-foundation for anticipated discrimination, as I show experimentally: 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, this presents another channel through which individual experiences of discrimination can affect future behavior. Finally, I provide suggestive evidence that retaliatory discrimination can be mitigated by highlighting the potential future consequences of current discriminatory actions.

Retaliatory discrimination 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 and behavioral foundation for the empirically documented microeconomic relationship between inter-group tensions and economic performance, as well as the macroeconomic role of ethnic divisions on conflict and economic development.

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

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.

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.

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.

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

Bordalo, P., Coffman, K., Gennaioli, N., and Shleifer, A. (2019). Beliefs about Gender. *American Economic Review*, 109(3):739–73.

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

Bosman, R., Sonnemans, J., and Zeelenberg, M. (2001). Emotions, Rejections, and Cooling Off in the Ultimatum Game. *International Journal of Modern Physics C - IJMPC*.

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

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.

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

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.

Esponda, I., Oprea, R., and Yuksel, S. (2023). Seeing What is Representative. *The Quarterly Journal of Economics*, 138(4):2607–2657.

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.

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.

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

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

Gupta, N., Rigotti, L., and Wilson, A. (2021). The Experimenters' Dilemma: Inferential Preferences over Populations.

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

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

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.

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.

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. (2019). Improving Access to Labor Markets for Refugees: Evidence from Uganda. Working paper, International Growth Centre.

Loiacono, F. and Silva Vargas, M. (2025). Matching with the Right Attitude: The Effect of Matching Firms with Refugee Workers. Working Paper.

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.

Milgrom, P. and Shannon, C. (1994). Monotone Comparative Statics. *Econometrica*, 62(1):157–180.

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.*

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.

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.

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.

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.

The White House (2025). Ending Radical and Wasteful Government DEI Programs and Preferencing. *The White House.* Presidential Action.

UNHCR (2025). Uganda - Refugee Statistics March 2025 - Active Population by Settlement. *UNHCR.*

Wicker, T. (2025). Winsorizing and Trimming with Subgroups. Working Paper.

Wicker, T., Dalton, P., and van Soest, D. (2025). Mental Accounting and Cash Transfers: Experimental Evidence from a Humanitarian Setting. Working Paper.

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.

<p style="text-align:center">Appendix to</p>

<p style="text-align:center">Discrimination as Retaliation</p>

<p style="text-align:center">by Till Wicker</p>

A1 Incorporating Other Sources 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 other models of discrimination. When only the taste component d_g matters for non-pecuniary costs, the model collapses to the taste-based discrimination model of Becker (1957). When non-pecuniary costs are absent but employers hold group-specific beliefs about productivity, we obtain statistical discrimination (Arrow, 1972a,b; Phelps, 1972).

A1.1 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). Under this specification, the employer assigns identical expected productivity to members of both groups, eliminating any informational asymmetry. 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 discriminatory taste d_g. Discriminatory tastes and behaviors are time-invariant and hence $D_t = D \; \forall t$, ceteris paribus.

A1.2 Statistical Discrimination

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

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

Employers do not observe the true productivity of the workers, 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 with past experiences (retaliatory discrimination).

A2 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 $\Pi_t(a_{i,t}, \theta_{i,t})$ from action $a_{i,t}$, and

2. Non-pecuniary costs $f(d_g, F(\chi_{g,t}))$ based on past interactions with members of group g before time t.

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

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

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 $s_{i,t} = \theta_{i,t} + \varepsilon_{i,t}$, $\varepsilon_{i,t} \sim N(0, 1/\eta_g)$. 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 (Becker, 1957).

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 in both arguments:

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

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

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: $D_t(s, \psi_A, \psi_B) \equiv \Gamma_{A,t}(s) - \Gamma_{B,t}(s)$. 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,b; 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 discriminatory that vary as a result of past experiences.

Propositions   Without loss of generality, suppose the DM discriminates against group B such that $D_t(s, \psi_A, \psi_B) > 0$. The relationship between past experiences and discriminatory behavior is grounded in two propositions:

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

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

2. (*Group-Specific Retaliatory Discrimination*) Ceteris paribus, experiences with unrelated groups $g' \notin \{A, B\}$ do not affect discrimination between A and B:

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

The proofs underlying these Propositions follow directly from Appendix A3.

A3   Theoretical Proposition Proofs

A3.1   Proposition 1:

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

Proof:

Fix time t and signal s. The employer chooses labor inputs (L_A, L_B) to maximize

$$U(L_A, L_B; F(\chi_B)) = Y(L_A, \theta_A, L_B, \theta_B) - w_A L_A - w_B L_B - L_A f(d_A, F(\chi_A)) - L_B f(d_B, F(\chi_B))$$

where $F(\chi_B)$ captures past experiences with group B, and $f(d_g, F(\chi_g))$ is increasing in its second argument: $\partial f(d_g, x) / \partial x \geq 0$. Because of this,

$$\frac{\partial^2 U}{\partial L_B \, \partial F(\chi_B)} = -\frac{\partial f(d_B, F(\chi_B))}{\partial F(\chi_B)} \leq 0$$

Hence, U satisfies the *single-crossing property* (Spence–Mirrlees condition) in $(L_B, F(\chi_B))$. By standard monotone comparative statics results (Milgrom and Shannon, 1994), the employer's optimal choice $L_B^*(F(\chi_B))$ is weakly decreasing in $F(\chi_B)$. If f is strictly increasing in its second argument and the optimum is interior, the inequality is strict. If the production function is additively separable across groups,

$$Y(L_A, \theta_A, L_B, \theta_B) = Y_A(L_A, \theta_A) + Y_B(L_B, \theta_B)$$

then the optimal choice $L_A^*(F(\chi_B))$ does not depend on $F(\chi_B)$. Defining the discrimination gap as

$$D(s, \psi_A, \psi_B \mid F(\chi_B)) = L_A^*(F(\chi_B)) - L_B^*(F(\chi_B))$$

it follows that

$$F(\chi_B)' > F(\chi_B) \implies D(s, \psi_A, \psi_B \mid F(\chi_B)') \geq D(s, \psi_A, \psi_B \mid F(\chi_B))$$

Thus, as the employer's prior experiences with group B become more negative (a higher $F(\chi_B)$), the optimal labor input for B decreases and the discrimination gap widens.

Q.E.D.

A3.2 Proposition 2:

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

Proof:

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

$$\sum_{g \in \{A,B\}} L_{g,t} f(d_g, F(\chi_{g,t}))$$

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

The first-order conditions are:

$$\frac{\partial Y}{\partial L_{A,t}} - w_A - f(d_A, F(\chi_{A,t})) = 0$$
$$\frac{\partial Y}{\partial L_{B,t}} - w_B - f(d_B, F(\chi_{B,t})) = 0$$

Since experiences with group g' (where $g' \notin \{A, B\}$) do not enter either of the previous equations, we have:

$$\frac{\partial L_{A,t}^*}{\partial F(\chi_{g',t})} = 0$$
$$\frac{\partial L_{B,t}^*}{\partial F(\chi_{g',t})} = 0$$

Therefore:

$$\frac{\partial D_t}{\partial F(\chi_{g',t})} = \frac{\partial L_{A,t}^*}{\partial F(\chi_{g',t})} - \frac{\partial L_{B,t}^*}{\partial F(\chi_{g',t})} = 0 - 0 = 0$$

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

$$D_t(s,\psi|\chi_{g',t}^{\text{mod}}) = D_t(s,\psi|\chi_{g',t}^{\text{neg}})$$

Q.E.D.

A4 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 (T1—T4). More specifically, theoretical predictions either expect 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 participant ($\mathbf{E_1}$) 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, so we would expect that the Eritrean participant ($\mathbf{E_1}$) has a greater distaste for the Ugandan worker than the Eritrean worker ($d_U > d_E$). Subsequently, the participant 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 ($f(d_g, F(\chi_{g,t})) = d_g$). Therefore, while the Eritrean participant ($\mathbf{E_1}$) 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), participants likely did not have much information or strong priors about worker- or group-level productivity. Furthermore, participant 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), and were informed that their stage 1 manager had the same information. 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.[48,49] 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:

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

Retaliatory Discrimination: Propositions 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:

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

Paternalistic Discrimination (Buchmann et al., 2024)*:* In line with the notion that refugees (and more generally, members of the minority group) are more vulnerable, paternalistic discrimination would predict that participants give *fewer* tasks to refugees, to protect them from an unpleasant situation (e.g. a paper cut). However, no differences would be expected across the different treatment arms.

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

---

[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 Online Appendix Table B3). 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.

[49]One concern could be that the stage 1 allocation could act as a signal of relative group productivity. However, in the experimental design, managers assign tasks prior to observing any worker output, and participants are explicitly shown pilot evidence indicating that average productivity is identical across groups. Thus, under Bayesian updating, the allocation contains no information about group productivity, and statistical discrimination models continue to predict no treatment differences.

Fairness Considerations: If 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. This is because the the notion of fairness (and the subsequent allocation across the two workers) is independent of *who* the manager was in the first stage.

$$U_{2,T1} = U_{2,T3} < U_{2,T2} = U_{2,T4}$$

Altruism: If the participant cares more about coethnic workers, this would result in more allocations to their fellow Eritrean worker, independent of allocations in the first stage:

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

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:

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

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:

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

Systemic Discrimination (Bohren et al., 2025b)*:* This describes the scenario where discriminatory practices are embedded within the structures and procedures of organizations. Systemic discrimination 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 in direct discrimination at a node during a fixed time. As such, systemic discrimination would not predict differential allocations across the treatment arms:

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

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:

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

A5   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, as well as 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 eight 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. Furthermore, replacing the computer manager with a coethnic manager addresses concerns surrounding computer vs. human biases and interactions.

Week 1 (n=639)

```
Stage 1: Role of Worker
```

| T1 (n=99) | T2 (n=109) | T3 (n=109) | T4 (n=108) | T5 (n=105) | T6 (n=109) |

```
Stage 2: Role of Manager
Allocating Tasks
```

Randomized Order

```
Stage 3: Real-Effort Task
```

Week 2 (n=464)

```
Recall Past Week Allocations
(Incentivized)
```

```
Role of Manager
Allocating Tasks
```

```
Memory Recall (Incentivized)
```

```
Role of Manager
Allocating Tasks
```

```
Role of Worker (Non-Coethnic
Manager; 2 / 8 tasks)
```

| Status Quo (n=78) | Future Rounds (n=74) | Direct Retaliation (n=74) | Costly Mistakes (n=78) | Other Worker (n=73) | Ineff. Allocation (n=74) |

```
Post-Experimental
Questionnaire
```

Figure A1. Overview: Experimental Design Prolific Experiment

Notes: The figure presents the experimental design of the online experiment on Prolific. In stage 1, participants are assigned the role of a worker, while in stage 2, participants become the manager who must allocates tasks between two workers. In stage 3, participants complete a 60-second, real-effort task to signal their productivity to a future hiring manager. For half of the participants, the order of stages 2 and 3 are reversed. A follow-up study is conducted one week later, in which participants are first incentivized to recall the name and task allocation of their stage 1 manager in the previous week. Subsequently, participants are assigned the role of manager, akin to stage 2 of the previous week's survey. Then, participants engage in an incentivized memory recall task, and are again assigned the role of the manager. In the penultimate stage, participants are assigned the role of a worker, whose non-coethnic manager assigns them two out of the eight tasks. However, this stage involves five different treatment arms, to test possible mitigation measures. Finally, participants complete a post-experimental questionnaire.

Subsequent Experiment (n=335)

Stage 1: Role of Worker

| *Status Quo* (n=96) | *Affirmative Action Removal* (n=98) | Uncertain Manager (n=109) |

Stage 2: Role of Manager Allocating Tasks

- -

Dictator Game (n=389)

Stage 1: Recipient

| T1 (n=66) | T2 (n=64) | T3 (n=64) | T4 (n=65) | T5 (n=65) | T6 (n=65) |

Stage 2: Role of Dictator Allocating Money

- -

Minimal Group Paradigm (n=326)

Stage 1: Role of Worker

| T1 (n=55) | T2 (n=53) | T3 (n=56) | T4 (n=53) | T5 (n=51) | T6 (n=58) |

Stage 2: Role of Manager Allocating Tasks

Figure A2. Overview: Experimental Designs of Additional Experiments

A6 Regression Tables: Uganda Experiment

Table A1: Allocation of Tasks to Ugandan Worker in Stage 2.

	Allocation of Tasks to U_2 in Stage 2
	(1)
T1	-0.20
	(0.14)
T2	-0.09
	(0.13)
T3	-0.54***
	(0.16)
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	3.63
Control Group S.D.	0.70
N	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). T1-T3 refers to Treatment arms 1-3. Control mean and standard deviation refer to the mean value and standard deviation of the outcome in T4. Robust standard errors are in parentheses.***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Table A2: Time Taken to Make Envelopes.

| | Time Taken to Make Envelopes (in seconds) | |
| | No Winsorizing | 95th percentile Winsorizing |
	(1)	(2)
T1	-64.98***	-63.621***
	(22.12)	(19.63)
T2	-0.88	-1.17
	(26.45)	(23.96)
T3	-99.05***	-98.21***
	(21.54)	(18.85)
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	311.09	308.93
Control Group S.D.	136.61	116.32
N	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). T1-T3 refers to Treatment arms 1-3. Control mean and standard deviation refer to the mean value and standard deviation of the outcome in T4. Column (1) reports results when outliers are not winsorized, while column (2) reports results when outliers are winsorized at the 95th percentile, separately per treatment arm as discussed in Wicker (2025). Robust standard errors are in parentheses.***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Table A3: Quality of Envelopes.

	Quality of Envelopes Made in Stage 1
	(1)
T1	0.01
	(0.06)
T2	-0.02
	(0.05)
T3	-0.08
	(0.05)
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	0.52
Control Group S.D.	0.28
N	224

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). T1-T3 refers to Treatment arms 1-3. Control mean and standard deviation refer to the mean value and standard deviation of the outcome in T4. Robust standard errors are in parentheses.***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Table A4: 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.

**Table A5:** Extensive and Intensive Margin of Discrimination: Uganda.

	Number of Discriminators	Allocation of Tasks to U_2 in Stage 2 Conditional on Discriminating
	(1)	(2)
T1	0.07	-0.17
	(0.09)	(0.10)
T2	-0.00	-0.09
	(0.09)	(0.10)
T3	0.17*	-0.51***
	(0.04)	(0.13)
T4 Mean	0.40	2.91
T4 S.D.	0.49	0.29
N	639	44

Notes: Intention to Treat estimates. The outcome variable is the share of discriminators (defined as assigning fewer than 4 tasks to the non-coethnic worker) in column 1, and the number of tasks assigned to the non-coethnic worker in the second stage conditional on discriminating in column 2. Control variables are selected using the post double LASSO machine learning algorithm outlined in Belloni et al. (2014). T1-T3 refers to Treatment arms 1-3. Control mean and standard deviation refer to the mean value and standard deviation of the outcome in T4. Robust standard errors are in parentheses.***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.
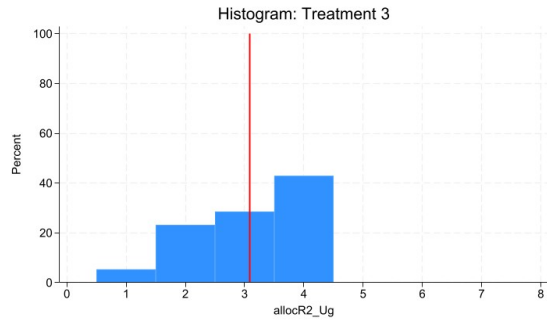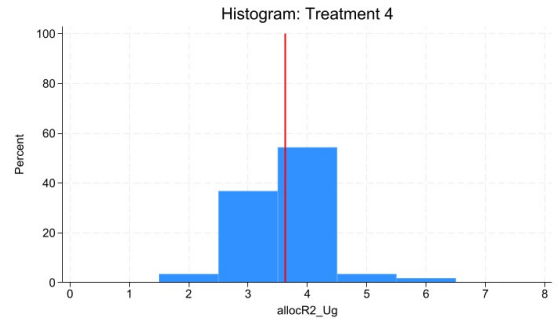
A6.1 Histogram: Uganda Experiment
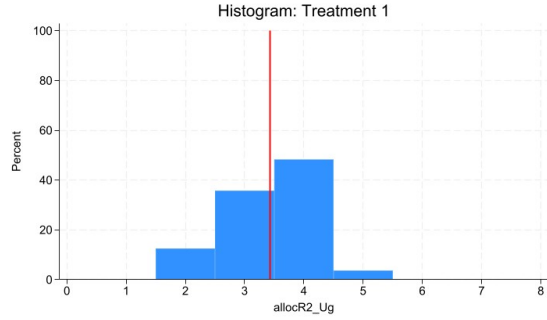
(a) Treatment 1
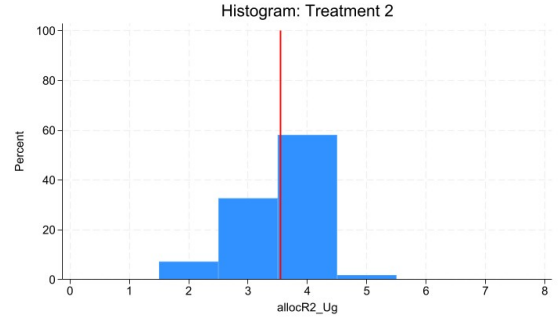
(b) Treatment 2

(c) Treatment 3

(d) Treatment 4

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

Notes: The figures present histograms of the number of tasks allocated to the Ugandan worker in stage 2 of the experiment (U_2) by the Eritrean participant (E_1). Allocating four out of the eight tasks indicates the case of no discrimination, indicated by the vertical red line.

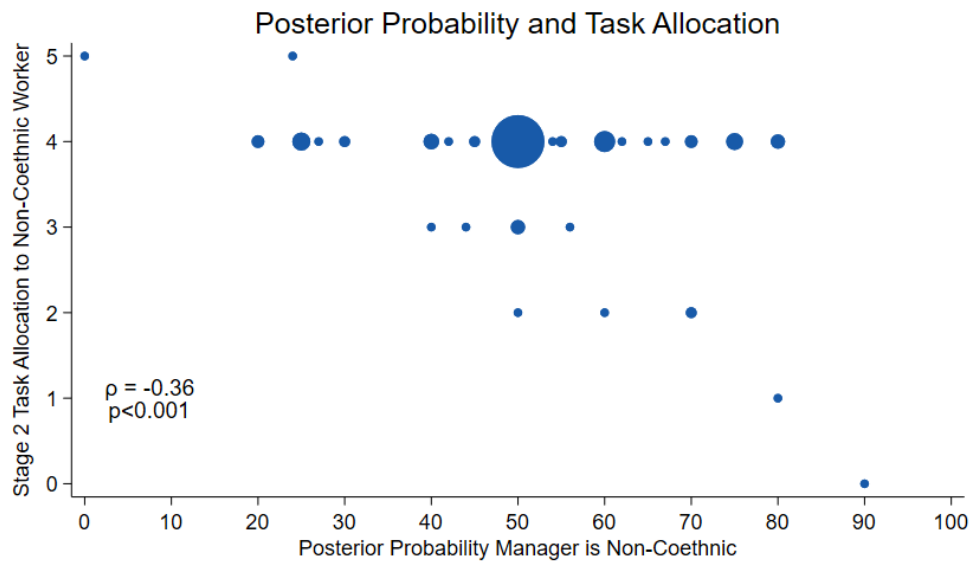## A6.2 Ex-Ante Statistical Power: Uganda

With 55-57 individuals randomized across treatment arms T1,...,T4, this study ex ante has 80% statistical power to detect a MDES of 0.47 and 0.53 standard deviations, based on a one- and two-sided t-test with $\alpha = 0.05$, respectively.

A7 Regression Tables: USA Experiment

Table A6: Allocation of Tasks to Non-Coethnic Worker in Stage 2.

	Allocation of Tasks to N_2 in Stage 2	
	(1)	(2)
T1	0.01	0.01
	(0.06)	(0.06)
T2	0.03	0.03
	(0.04)	(0.05)
T3	-0.06	-0.06
	(0.08)	(0.08)
T4	-0.20***	-0.21***
	(0.07)	(0.08)
T6	0.06	0.06
	(0.06)	(0.06)
Order Effects		-0.03
		(0.05)
Control Group Mean	3.99	3.99
Control Group S.D.	0.26	0.26
N	639	639

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). $\{T1 - T4, T6\}$ refers to Treatment arms 1-4,6. Order Effects refer to whether participants were randomized into completing Stage 2 or 3 first. Control mean and standard deviation refer to the mean value and standard deviation of the outcome in T5. 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

Notes: On the y-axis, allocations to a non-coethnic worker in the second stage of participants randomized into the *Uncertain Manager* treatment arm of the additional experiment are plotted. In this treatment arm, participants are informed that with 50% probability their manager is coethnic, and with 50% probability their manager is non-coethnic. After participants are shown that they have been assigned two of the eight tasks, they are asked to indicate with what probability they believe the manager is non-coethnic. The posterior probability of the manager being non-coethnic is plotted on the y-axis. Dots are reflective of the frequency in which the (x,y) coordinates occur: larger dots indicate greater frequency.

Table A7: Errors and Time Duration of Task in Stage 1.

	Stage 1 Task	
	Error Rate	Time Taken
T1	0.01	-23.58***
	(0.01)	(3.48)
T2	-0.01	4.59
	(0.01)	(5.88)
T3	-0.01	25.45***
	(0.01)	(5.19)
T4	0.00	-17.84***
	(0.01)	(4.25)
T6	0.02	24.24***
	(0.01)	(4.34)
T5 Mean	0.04	61.25
T5 S.D.	0.09	32.42
N	639	639

Notes: Intention to Treat estimates. The outcome variable in column (1) is the error rate per completed task in the first stage of the online experiment, and the seconds taken to complete the stage 1 tasks. Control variables are selected using the post double LASSO machine learning algorithm outlined in Belloni et al. (2014). $\{T1 - T4, T6\}$ refers to Treatment arms 1-4,6. Control mean and standard deviation refer to the mean value and standard deviation of the outcome in T5. Robust standard errors are in parentheses.***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Table A8: Extensive and Intensive Margin of Discrimination: USA.

	Number of Discriminators	Allocation of Tasks to N_2 in Stage 2 Conditional on Discriminating
	(1)	(2)
T1	0.01	-0.75
	(0.03)	(0.65)
T2	-0.00	-0.33
	(0.02)	(0.27)
T3	0.05*	-0.89***
	(0.03)	(0.33)
T4	0.13***	-0.65***
	(0.04)	(0.20)
T6	0.04	-0.25
	(0.03)	(0.23)
T5 Mean	0.03	3.00
T5 S.D.	0.17	0.00
N	639	44

Notes: Intention to Treat estimates. The outcome variable is the share of discriminators (defined as assigning fewer than 4 tasks to the non-coethnic worker) in column 1, and the number of tasks assigned to the non-coethnic worker in the second stage conditional on discriminating in column 2. Control variables are selected using the post double LASSO machine learning algorithm outlined in Belloni et al. (2014). $\{T1 - T4, T6\}$ refers to Treatment arms 1-4,6. Control mean and standard deviation refer to the mean value and standard deviation of the outcome in T5. Robust standard errors are in parentheses.***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Table A9: Errors and Effort of Real Effort Task (Stage 3).

	Error Rate (1)	Number of Tasks Completed (2)
T1	0.01	-0.33
	(0.02)	(0.45)
T2	0.01	-0.35
	(0.02)	(0.49)
T3	0.01	-0.21
	(0.02)	(0.45)
T4	0.01	-0.92**
	(0.02)	(0.43)
T6	0.02	-0.23
	(0.02)	(0.54)
Order Effect	-0.01	0.20
	(0.01)	(0.26)
T5 Mean	0.06	6.60
T5 S.D.	0.13	3.74
N	639	639

Notes: Intention to Treat estimates. The outcome variable in column (1) is the error rate per completed task in the third stage of the online experiment, and the number of tasks completed in the third stage. Control variables are selected using the post double LASSO machine learning algorithm outlined in Belloni et al. (2014). $\{T1 - T4, T6\}$ refers to Treatment arms 1-4,6. Order Effects refer to whether participants were randomized into completing Stage 2 or 3 first. Control mean and standard deviation refer to the mean value and standard deviation of the outcome in T5. Robust standard errors are in parentheses.***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Table A10: Effects of Recall on Persistence.

	Number of Tasks Allocated to Non-Coethnic Worker Week 2
	(1)
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 A11: Costly Mistakes and Discriminatory Allocations

	Number of Tasks Allocated to Non-Coethnic Worker
	(1)
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, see Appendix Figure A2). Robust standard errors are in parentheses. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Table A12: Inefficient Allocations and Discriminatory Allocations

	Number of Tasks Allocated to Non-Coethnic Worker (1)
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, see Appendix Figure A2). Robust standard errors are in parentheses. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Table A13: Memory Recall of Past Rounds

	Correctly Recalled Allocations		Discrepancy: Recall Tasks for Coethnic Worker	
	(1)	(2)	(3)	(4)
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 A14: Memory Recall on Retaliatory Discrimination

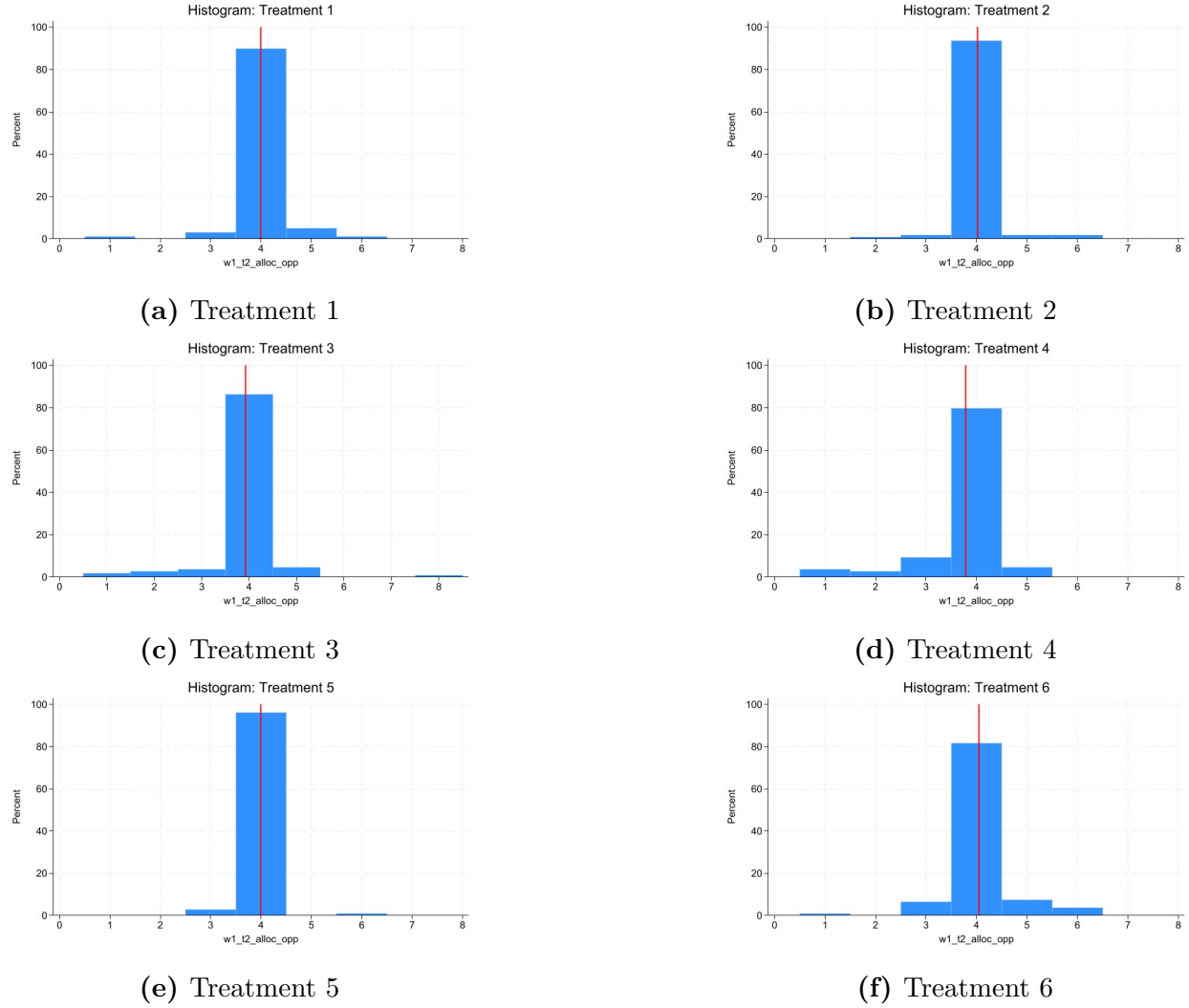|	Number of Tasks Allocated to Non-Coethnic Worker	
	(1)	(2)
Correctly Recalled Rounds	0.03	
	(0.10)	
Average Discrepancy of Recall		0.07
		(0.07)
T1 Mean	4.05	4.05
T1 S.D.	0.69	0.69
N	451	451

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.

Table A15: Extensive and Intensive Margin of Discrimination: Removal Affirmative Action.

	Number of Discriminators	Allocation of Tasks to N_2 in Stage 2 Conditional on Discriminating
	(1)	(2)
Treatment: *AA Removal*	-0.03	-0.80*
	(0.04)	(0.45)
Status Quo Mean	0.11	2.72
Status Quo S.D.	0.32	0.47
N	96	11

Notes: Intention to Treat estimates. The outcome variable is the share of discriminators (defined as assigning fewer than 4 tasks to the non-coethnic worker) in column 1, and the number of tasks assigned to the non-coethnic worker in the second stage conditional on discriminating in column 2. Control variables are selected using the post double LASSO machine learning algorithm outlined in Belloni et al. (2014). $\{T1 - T4, T6\}$ refers to Treatment arms 1-4,6. Control mean and standard deviation refer to the mean value and standard deviation of the outcome in T5. Robust standard errors are in parentheses.***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

A7.1 Histogram: USA Experiment

(a) Treatment 1

(b) Treatment 2

(c) Treatment 3

(d) Treatment 4

(e) Treatment 5

(f) Treatment 6

Figure A5. Histograms: Prolific Experiment

Notes: The figures present histograms of the number of tasks allocated to the Non-Coethnic worker in stage 2 of the experiment (N_2) by the participant (C_1). Allocating four out of the eight tasks indicates the case of no discrimination, indicated by the vertical red line.

A7.2 Ex-Ante Statistical Power: USA

With 99-109 individuals randomized across treatment arms T1,...,T6, this study ex ante has 80% statistical power to detect a MDES of 0.34 and 0.39 standard deviations, based on a one- and

two-sided t-test with $\alpha = 0.05$, respectively.

A8 Misinterpreting Retaliatory Discrimination

The results from stage 2 of the experiment in Uganda can also be used to illustrate how retaliatory discrimination may be misclassified as taste-based discrimination à la Becker (1957) when only behavior in round t is considered, without considering rounds $t - i$, $i > 0$. In a separate survey, 51 academics were asked to identify the source of discrimination in the second stage of the experiment. After receiving an overview of taste-based and statistical discrimination, as well as the experimental set-up of stage 2, half the participants were randomized to see the participant's division of tasks across the Ugandan and refugee worker in stage two of T3. The other half were shown the division of tasks in T4.

Despite the differences in both the source and intensity of discrimination between the two treatment arms, experts overwhelming identify the source of the discrimination as being taste-based in both treatment arms. 65.4% and 64.0% of academics shown task allocations in T3 and T4 identified the source of discrimination as taste-based, respectively ($p = 0.920$).[50] This illustrates how retaliatory discrimination can be mis-identified as taste-based discrimination if individuals do not take earlier interactions into consideration.

---

[50]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.

A9 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 Online Appendix Table B2.[51] This is a frequently used approach in experimental studies to minimize the role of (inaccurate) statistical discrimination (for example, see Chan 2025).
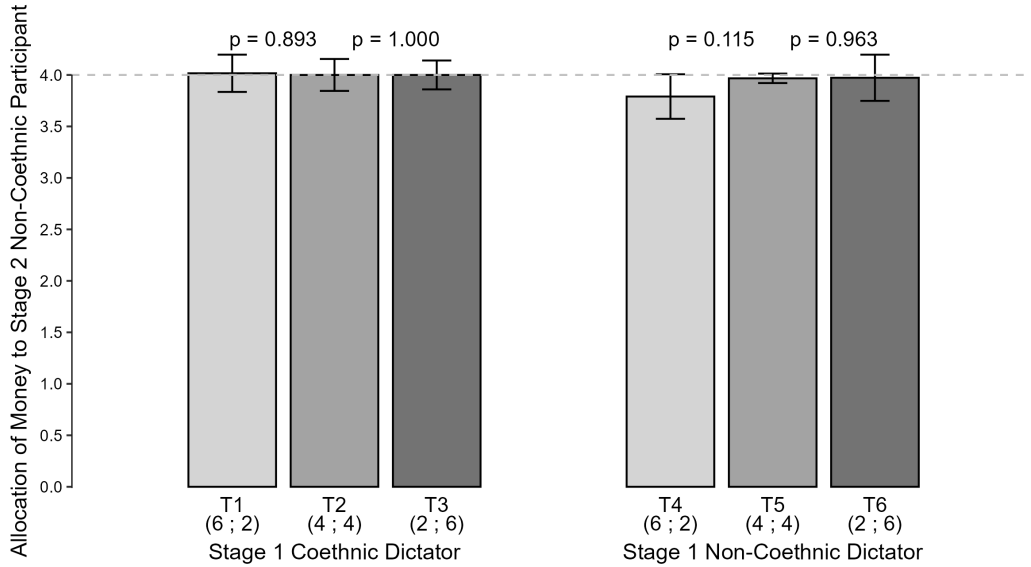
Figure A6. Money Allocation to Non-Coethnic Worker in Stage 2: Dictator Game

Notes: The figure shows the amount of money (in the form of quantities of 10 cents) allocated to the Non-Coethnic participant in stage 2 of the dictator game experiment (N_2) by the participant (C_1). Allocating four out of the eight sets of 10 cents indicates the case of no discrimination, indicated by the dashed gray line. The figure reports average money allocations to the Non-Coethnic worker across the six treatment arms (T1—T6), including 95% confidence intervals. P-values are based on two-sided t-tests.

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.

---

[51]In the online experiment, participants were further informed that stage 1 managers received the same information prior to making their allocation decisions.

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.

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. Reciprocity models (Rabin, 1993) predict that direct retaliation will be stronger than indirect retaliation. In line with this, 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 A16). However, this sub-treatment also rules out a generalized reciprocity model where reciprocity would extend equally to uninvolved others who share the perpetrator's group identity.

Table A16: Direct Retaliation and Discriminatory Allocations

	Number of Tasks Allocated to Non-Coethnic Worker (1)
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, see Appendix Figure A2). 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

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. This also rules out that treatment effects observed in Sections 3 and 4 are driven by an updating of beliefs other than discriminatory preferences. If this were the case, we would have expected a similar pattern in Appendix Figure A7 as in Figure 4.
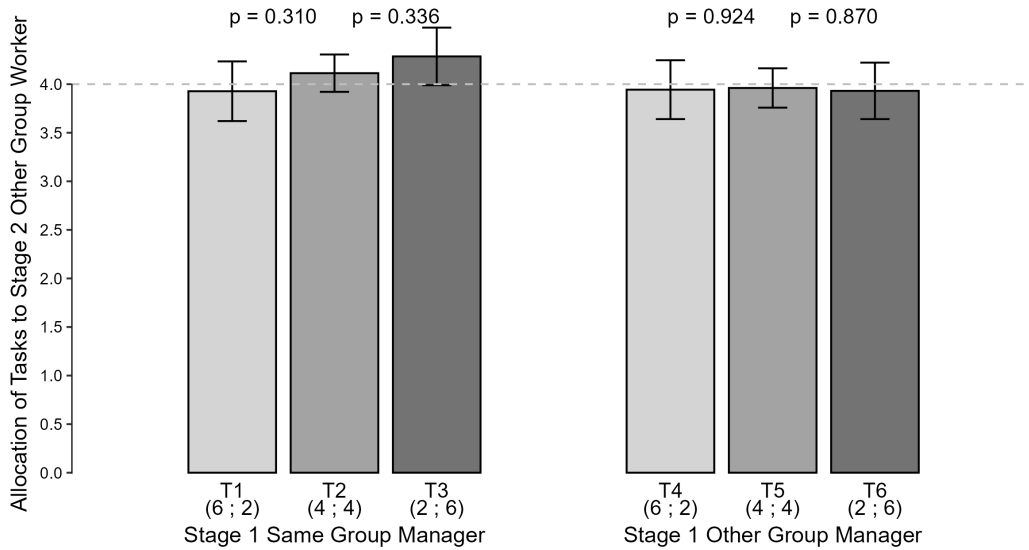
Figure A7. Task Allocation to Other Group Worker in Stage 2: Minimal Group Paradigm

Notes: The figure shows the number of tasks allocated to the worker of the other group (Red vs. Blue) in stage 2 of the minimal group paradigm experiment by the participant. Allocating four out of the eight tasks indicates the case of no discrimination, indicated by the dashed gray line. The figure reports average task allocations to the worker of the other group across the six treatment arms (T1—T6), including 95% confidence intervals. P-values are based on two-sided t-tests.

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 in $\{T1, T3, T4, T6\}$ compared to treatments where the stage 1 manager split the tasks evenly ($p < 0.001$), average allocations do not differ significantly ($p = 0.818$).[52]

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 A10). 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 (despite the increase in frequency of norm violations), indicating that the memory recall, and associated heightened salience of norm violations, is unlikely to have caused 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. I do document that participants randomized to the Computer manager in stage 1 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 in $\{T1, T2, T4\}$ 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

---

[52]Results are similar for the dictator and minimum group paradigm games: the number of allocations that deviate from an even split (4 ; 4) is higher in $\{(T1, T3, T4, T6\}$ than in in $\{(T2, T5\}$ ($p = 0.026$ and $p < 0.001$); however average allocations do not differ significantly ($p = 0.469$ and $p = 0.894$, for the dictator and minimum group paradigm games respectively).

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.

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 Black 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 Black workers.[53]

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.

---

[53]We also don't observe heterogeneity by perceptions of discrimination in Uganda (see Online Appendix Table B11), and find that participants with below-median discriminatory perceptions the USA have larger treatment effects (see Online Appendix Table B14), in contrast to predictions of a social planner.

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 A6 column (2), I illustrate that having to first complete a real-effort task, that takes ~ 3 minutes before making allocation decisions 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.[54] As such, we would expect retaliation in T1, however we do not observe this. Furthermore, 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 6.1, nor the mitigating effects of highlighting the salience of future rounds (Section 7). Lastly, if anger were driving the treatment effects, we would expect to find results in the minimal group paradigm experiment, which we do not (see Appendix Figure A7).

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. First, 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. Second, 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

---

[54]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, $p = 0.252$).

in-person sample, and with tasks that vary in their usefulness, I minimize the role of experimenter demand effects.

There are two other pieces of evidence that suggest experimenter demand effects do not play a major role. First, if participants in the first experiment cared about the quality of the envelopes produced in order to please the researchers, we would not expect the quality of the envelopes to be different across treatment arms. This is in contrast to findings of Appendix Table A3. Second, 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 (see Appendix Figure A7).

A10 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 A17: Rounds Shown to Participants for Memory Recall