Attention Cost of Unfair Treatment

Daniel J. Banko-Ferran*

November 29, 2025†

Click here for most recent version

Abstract

Unfair treatment is common in the workplace, where it may manifest through channels such as unequal wage compensation, task misallocation, or overt discrimination. This paper provides causal evidence that perceived unfairness reduces attention, impairing performance on workplace-relevant tasks. Using a large-scale online experiment with over 3,300 participants, I find that unfair treatment significantly reduces attention scores by roughly 7%, with especially pronounced effects when unfairness is attributed to discrimination. These results highlight the cognitive costs of unfairness, with implications for productivity and workplace design.

Keywords: Discrimination, Productivity, Cognition
JEL Classification: J71, J24, M54, D91

*`db1699@georgetown.edu`. McCourt School of Public Policy, Georgetown University.

†Institutional Review Board approval for human subjects from the University of Pittsburgh (STUDY20090122). This study's analysis plans were pre-registered on AsPredicted. I thank David Huffman, Osea Giuntella, Silvia Saccardo, and Lester Lusher for helpful and supportive advising on this project. I would like to thank attendees at the University of Pittsburgh's behavioral/experimental economics brown bag, Economic Science Association's North America and Asia-Pacific meetings, BEEMA9, and C-BEAM for their thoughtful and engaging questions. I also thank Astrid Hopfensitz, Alex Eble, David Dickinson, Christopher Roth, Sevgi Yuksel, and Joakim Weill for useful suggestions and comments. All errors are my own. Data and analysis materials (Stata code) will be made available online upon publication.

1 Introduction

Unfair treatment can occur in many forms in the workplace. Examples include unequal wages for similarly productive workers, or a co-worker free-riding and taking credit for another worker's efforts. These actions violate central notions of fairness, captured by models of social preferences such as egalitarianism, inequality aversion, and reciprocity (Bartling et al., 2009; Cappelen et al., 2022; Falk and Fischbacher, 2002; Fehr and Schmidt, 1999; Houser and Kurzban, 2002; Kahneman, Knetsch and Thaler, 1986). A body of research has shown that unfair treatment can prompt individuals to deliberately take costly actions to punish the unfair actor or restore fairness, such as reducing effort or lowering quality (Cohn et al., 2011; Krueger and Mas, 2004). For instance, Krueger and Mas find that workers deliberately reduced output to retaliate against wage cuts, while Cohn et al. show that workers reduce performance more strongly when wages are unfairly cut for them than when they are uniformly reduced for all. Such responses, though costly, can be utility-maximizing through a direct utility benefit of punishing unfairness and by deterring future unfairness.

This paper considers another channel through which unfair treatment can reduce productivity: not through deliberate punishment of others but through involuntary reductions in focus and effort, ultimately harming individual and organizational welfare. Specifically, this paper tests the hypothesis that unfair treatment may tend to act as a distraction to those who are treated unfairly, and reduce attentional resources needed for job performance. This can lead to lower performance ratings and earnings for those who are treated unfairly, even when their worse productivity does not directly affect unfair actors.

To illustrate, imagine an employee whose colleague takes credit for their contribution on a project. Even if the employee chooses not to retaliate, they may continue to ruminate about the unfair incident and have difficulty concentrating on other tasks. Psychological research provides evidence that rumination (repetitive thinking about negative experiences) impairs working memory and attentional control (Whitmer and Gotlib, 2013), and disrupts the efficiency of the goal-directed attentional system (Eysenck et al., 2007). Yet, whether unfair treatment itself can impair attention and performance in a work setting has not been systematically tested.

I address this gap using a large pre-registered online experiment (N = 3,364) designed to mimic unfair interactions in the workplace. Workers are randomly assigned to perform work-related tasks under procedures that differ in the presence of unfair treatment. The experimental design includes three treatment conditions that isolate distinct mechanisms linking fairness and distraction. The *Control* condition rules out the possibility that any randomly revealed information could reduce attention; this isolates the effect of unfairness content itself rather than generic distraction from new information. The *Fair-OutGroup* condition tests whether receiving fair treatment might also be distracting to determine whether distraction is specific to negative fairness violations. Finally, the *Unfair In-Group* condition explores whether unfair treatment is more or less distracting when it comes from an in-group member (for example, someone with a shared gender or political identity) as opposed to an out-group member.

In addition to the attention measures, I include validated measures of cognitive performance: a four-item variant of the Cognitive Reflection Test (Frederick, 2005) and a four-item Raven's Progressive Matrices test (John and Raven, 2003). These tests are validated measures of cognitive performance that are predictive of other outcomes related to intelligence (Brañas-Garza, Kujal and Lenkei, 2019; Burfurd and Wilkening, 2022; Carpenter, Just and Shell, 1990; Pennycook and Rand, 2021). I also measure perceived discrimination, affect, and beliefs relating to experiencing unfair treatment. I focus on gender and political affiliation out-group pairings, but allow participants to perceive and report discrimination based on other demographic factors, including racial and age discrimination. By observing whether workers performed worse under procedures of unfair treatment through random assignment, I can measure the average treatment effect of unfairness on cognitive performance and its impact on earnings.

This study makes several contributions to the literature. First, it identifies a novel pathway linking unfair treatment to reduced productivity (through cognitive distraction rather than strategic retaliation), bridging economic models of fairness preferences with psychological models of attention and emotion regulation. Second, it complements prior work on the behavioral consequences of unfair treatment (Cohn et al., 2011; Fehr and Schmidt, 1999; Gächter, Nosenzo and Sefton, 2012; Xiao and Houser, 2005) by showing that even when retaliation is impossible, fairness violations can harm performance. Third, it connects to research on the cognitive and financial costs of discrimination (Becker, 2010; Bertrand and Duflo, 2017; Blau and Kahn, 2017; Fershtman and Gneezy, 2001; Kline, Rose and Walters, 2022; Mullainathan and Shafir, 2013; Neumark, 2018; Ruebeck, 2025; Salvatore and Shelton, 2007; Sloan, 2012; Zahodne et al., 2020), highlighting that attention may be a mechanism through which unfair environments perpetuate inequality.

The rest of the paper proceeds as follows. Section 2 presents the theoretical framework linking unfair treatment and attention. Section 3 describes the experimental design. Section 4 presents the results, and Section 5 concludes.

2 Theoretical Framework

2.1 Background

Building on the literature on rational inattention (Maćkowiak, Matějka and Wiederholt, 2023), attention can be modeled as a scarce resource that individuals must allocate across tasks under information-processing constraints. This theoretical perspective is based on the directed cognition model of (Gabaix and Laibson, 2005), which conceptualizes attention as the outcome of option-value calculations under scarcity. In their framework, agents choose cognitive operations by comparing the expected benefit of further analysis to its cost. In my framework, attention carries an implicit cost, and the allocation of attention is governed by a trade-off between the expected benefits of additional information and the mental resources required to process it.

Unfair treatment enters this model as a parameter that increases the effective cost of attention. Psychologically, this cost arises because unfair treatment induces cognitive load and rumination which expend mental resources and lower available bandwidth.[1] In

[1]According to the American Psychological Association, cognitive load is the relative demand imposed by a particular task in terms of mental resources required, whereas rumination is thinking involving

my setting, when unfair treatment increases the shadow cost of attention, the threshold for allocating effort to productive tasks rises, leading to suboptimal search, premature stopping, or misallocation of mental effort. Moreover, rumination can be understood as a distortion in the information-value calculation: instead of allocating attention to information with the highest expected return, individuals may allocate excessive bandwidth to processing or rehearsing unfair experiences, even when these yield no instrumental benefit.

This section develops a simple framework to formalize the attention costs of unfair treatment. I build on the rational inattention literature (Maćkowiak, Matějka and Wiederholt, 2023) and models of directed cognition (Gabaix and Laibson, 2005), treating attention as a scarce economic resource that individuals allocate under cognitive constraints. Unfair treatment is modeled as a parameter that inflates the cost of attention, thereby lowering the optimal attention level for information-search tasks.

2.2 Rational Inattention with Unfairness

Let Θ denote the set of possible true states of the world, $\mathcal{S}$ the set of possible signals perceived by a worker, and T the set of possible attention levels. Consider a worker who chooses an attention level $\tau \in T$ in order to observe a noisy signal $s \in \mathcal{S}$ about an underlying state $\theta \in \Theta$. The state θ is a task-relevant quantity (for example, the amount of effort required to complete a task) that the worker wishes to know to perform a task but cannot directly observe it. Obtaining information about a signal (through increasing τ) increases the quality of the signal but also incurs an information-processing cost.

Attention is represented by the parameter $\tau \geq 0$: as τ increases, the conditional distribution of signals becomes more concentrated around $s = \theta$, so the signal becomes more informative. The worker chooses an attention level that maximizes conditional expected utility. A higher τ means allocating more attention to generate a clearer signal about the state, which sharpens the posterior about θ given s, so the action selected better matches the true state and raises expected utility.

excessive, repetitive thoughts or themes that interfere with other forms of mental activity (source: https://dictionary.apa.org/).

The first of two key aspects of the model is that the cost of processing information increases depending on the intensity of unfair treatment, $\mu \geq 0$. To capture the notion that unfair treatment is a violation of fairness norms (Cohn et al., 2011; Kahneman, Knetsch and Thaler, 1986), $\mu = 0$ is the baseline where a decision-maker is not treated unfairly. That is, $\mu = 0$ when either the individual is treated fairly or not interacting with other individuals (i.e., no social preferences).

The second key aspect of the model is that unfair treatment μ affects attention through rumination and cognitive load. The cost of processing information is represented by the parameter $\lambda(\mu) \geq 0$: as μ increases, optimal τ will decrease because attention is more costly.

The worker chooses an attention level τ that maximizes expected utility net of information-processing costs:

$$\max_{\tau \geq 0} \gamma\tau \; - \; \tfrac{1}{2}\lambda(\mu)\tau^2 \tag{1}$$

where

$$\lambda(\mu) = \lambda_0(1 + \alpha\mu), \qquad \lambda_0 > 0, \; \alpha > 0 \tag{2}$$

is the unit cost of processing information. λ_0 represents the baseline cost of attention and α measures how sensitive that cost is to unfair treatment. $\gamma > 0$ represents the productivity of attention, i.e. how much the expected utility benefit improves with additional precision of the signal. The cognitive cost of increasing precision through the attention parameter τ is quadratic because sustaining higher precision requires disproportionately more mental control and effort.

The first-order condition is

$$\frac{\partial}{\partial \tau}\left[\gamma\tau - \tfrac{1}{2}\lambda(\mu)\tau^2\right] = 0 \quad \Rightarrow \quad \gamma - \lambda_0(1 + \alpha\mu)\tau = 0$$

$$\Rightarrow \tau^*(\mu) = \frac{\gamma}{\lambda_0(1 + \alpha\mu)}. \tag{3}$$

Differentiating $\tau^*(\mu)$ yields:

$$\frac{d\tau^*}{d\mu} = -\frac{\gamma\alpha}{\lambda_0(1+\alpha\mu)^2} < 0 \tag{4}$$

Unfairness increases the marginal cost of τ, leading to lower optimal attention allocation. This can lead to a reduction in performance on a work-related task, thereby reducing expected earnings. In my experimental design, I estimate the change in τ due to unfair treatment from a baseline (control) condition using a treatment arm. I also measure τ in a second treatment arm where participants are fairly treated. The model yields two testable predictions. First, unfair treatment should reduce attention relative to a baseline group that is not unfairly treated. Second, fair treatment should not reduce performance relative to the baseline, as the associated attention cost parameter is the same because $\mu = 0$ for both control and fairly treated participants. Consequently, participants in the fair treatment group are expected to perform comparably to those in the control group on attention tasks, and significantly better than those in the unfair treatment group.

3 Research Design

I conducted a pre-registered, between-subjects experiment on Prolific in 2024–2025 with $N = 3,364$ participants. Participants were randomly assigned to one of two roles, *Dictators* and *Recipients*, and placed into pairs based on opposing gender and political affiliation. The purpose of these pairings was to simulate interactions among coworkers in the workplace.

The procedure for inducing unfair treatment was as follows. *Recipients* completed a keyboard-pressing task to earn a monetary bonus, while *Dictators* decided whether to take the bonus or not and skip doing the task themselves. The dictator made the decisions before participants were paired, as the pairings were made based on the decisions of the *Dictators*and treatment assignment of the *Recipients*. *Recipients* were assigned to one of three treatments: *Unfair-OutGroup*, *Fair-OutGroup*, or *Control.* Participants in the *Unfair-OutGroup* were paired with dictators who decided to take their financial bonus.

Participants in the *Fair-OutGroup* condition were assigned to dictators who declined to take their financial bonus. Participants in the *Control* condition were not paired with another participant, and performed the task without the chance of winning a financial bonus. Note that participants in the *Dictator* role are not included in the analysis sample.

After learning whether their bonus was taken, all participants completed a series of incentivized attention tasks. The primary outcome measures are performance on these attention tasks.

3.1 Setting and Sample

The study design and analysis approach was pre-registered on aspredicted.org (Analysis plans #186456 and Analysis plan #209646). The study was conducted on Prolific, an online platform that provides high-quality, diverse samples for academic research (Peer et al., 2022). The main sample was collected in three waves.[2] Participants had to be adults aged 18 and older and living in the US. All participants consented before taking part in the study. Participants completed a series of comprehension questions, attention checks, and minimum effort checks. Participants received full information and the order of tasks was counterbalanced across participants so that half of participants completed the tasks in reverse order. The *Main* sample consists of 2,303 participants and includes the *Control*, *Unfair-OutGroup*, and *Fair-OutGroup* treatment arms. The *Unfair-InGroup* treatment arm ($N = 1,061$) is analyzed separately. More participants were recruited to the treatment groups than the control group to account for heterogeneity in perceived discrimination and to preserve power when conducting subsample analyses.

3.2 Experimental Design

3.2.1 Wave 1: ***Dictator*** Sample

In Wave 1, participants were recruited to the *Dictator* condition. Participants were given a choice between two actions: 1) Perform a completion task requiring 200 units

[2]The *Dictator* sample was collected in June 2024. The *Control* and *Unfair-OutGroup* samples were collected in August 2024. The *Fair-OutGroup* was conducted in January 2025. A fourth wave for the *Unfair-InGroup* condition was collected in June 2025 and not included in the main sample.

of real-effort for a \$0.75 payment (*Not-Take*); 2) Take \$0.75 from a participant and skip completing the real effort task (*Take*). Participants received instructions about the real-effort task, the possibility of taking the earnings of another participant, and the subsequent attention tasks. Participants received detailed instructions explaining the choice they would make and the information set of their paired participant as well as their own before they made their choice to take or not.[3] Truthful information about the *Dictator* demographic information was shared with the other participant they were paired with, and the same demographics information of the participant they were paired with was shared with them (age group, political affiliation, and gender). The information was shown twice, once in the instructions before they made their decision and on the same screen where they made their choice to take. If a *Dictator* made the *Not-Take* choice, they immediately began the real-effort task. If they did not complete 200 units of effort (keyboard presses) in five minutes, they could not continue with the study.

3.2.2 Wave 2: Unfair-OutGroup Treatment

In the second wave, participants were randomly assigned to either a treatment (*Unfair-OutGroup*) or control (*Control*) condition. In the treatment condition, participants were informed that they would complete the same real-effort task as in Wave 1 and be paired with a participant from the *Dictator* sample. The probability of being matched with a Dictator who chose *Take* was 100%. As in Wave 1, participants were told that their demographic characteristics (age group, political affiliation, and gender) would be shared with their partner, and that they would also receive the same information about their partner. This setup created the potential for perceived discrimination: treated participants could attribute the Dictator's decision to take their bonus to the demographic information that had been exchanged.

Participants in the *Unfair-OutGroup* condition received instructions about the real-effort task, the possibility of having their earnings taken by a Wave 1 participant,

[3] Participants in the Dictator condition were told: Before the keyboard pressing task, you will be paired with another person in the study. You will not be told who you are matched with during or after the study, and he or she will not be told who you are either during or after the study. Your decisions will be strictly anonymous and cannot be linked to you in any way.

and the subsequent attention tasks. They were then shown the demographics of the Wave 1 participant with whom they were paired. All participants performed the same real-effort task as in Wave 1, with the requirement of completing at least 200 units of effort (keyboard presses) within five minutes in order to proceed. Finally, participants were informed whether their earnings had been taken by their paired partner.

Participants in the control group performed the same real-effort task for a $0.00 payment (i.e., no financial bonus). This ensured that final wealth levels were the same in both the *Unfair-OutGroup* group and the control group. Instead of being paired with a participant from the *Dictator* sample, the control observed a draw from a mystery urn filled with green and black colored balls before completing the attention tasks. This was to address the fact that any information exerts cognitive load. This setup ensured that any observed differences in attention task performance could be attributed to the content of the information in the *Unfair-OutGroup* setting itself, rather than differences in financial incentives and informational provision.

3.2.3 Wave 3: Fairness Treatment

In the third wave, recruited participants were assigned to a second treatment group, *Fair-OutGroup*. In the treatment condition, participants were informed that they would perform a real-effort task (same task as in Waves 1 and 2) and be matched with a participant from the *Dictator* sample. The probability that the participant was matched with a Dictator who chose *Not Take* was 100%. As in Waves 1 and 2, they were informed that their demographics (age group, political affiliation, and gender) would be shared with the participant they are paired with and would receive the same demographics information of the participant paired with them.

Fair-OutGroup participants received instructions about the real-effort task, the possibility of having their earnings taken by a participant from Wave 1, and the subsequent attention tasks. They were then given information about the demographics (age group, political affiliation, gender) of the participant in Wave 1 they were paired with. If they did not complete 200 units of effort (keyboard presses) in five minutes, they could not

Dictator	Recipient
Male Republican	Female Democrat
Female Republican	Male Democrat
Male Democrat	Female Republican
Female Democrat	Male Republican

Table 1: Dictator-Recipient Pairings (*Unfair-OutGroup* and *Fair-OutGroup* Treatments)

continue with the study. Participants were then informed whether their earnings were taken by the Wave 1 participant they were paired with.

3.3 Matching

For the *Unfair–Outgroup* and *Fair–Outgroup* treatments, participants were paired across both gender and political affiliation—for example, a female Democrat was always matched with a male Republican *Dictator*, and vice versa. Table 1 lists all *Dictator–Recipient* pairings. This cross-group matching design was intended to heighten the likelihood that unfair treatment would be perceived as discriminatory by invoking out-group dynamics. Post-experimental survey responses measure whether participants attributed unfairness to discrimination based on race, gender, political affiliation, or age. In addition to serving as manipulation checks, these responses allow me to test whether unfair treatment exerts a stronger effect on attention when it originates from an in-group versus an out-group member, as I show in the exploratory analysis section.

3.4 Real-effort Task

The real-effort task, adapted from DellaVigna and Pope (2018), was designed to be purely effortful, ensuring a minimum level of effort was required to earn a financial bonus. This task was chosen to ensure that participants had to exert a small amount of effort to earn their payment. This meant that participants who lost the financial bonus would feel that their effort was unfairly allocated towards compensating another participant. Participants pressed two keys in sequence on their keyboard ("Z" and "X"). If they did not complete 200 of these sequences correctly in the span of 300 seconds (5 minutes), they were automatically redirected to the end of the survey. As soon as each participant completed 200 keyboard presses, they were each automatically advanced to the next stage of the experiment. This ensured that all participants who advanced had performed the same number of tasks and were equal in effort cost up to that point.

3.5 Attention Task 1: Credit Card Choice Task

In the first attention task, participants were asked to compare two financial contracts and choose the one that was less expensive. This task had three multiple choice responses, and only one response was correct and scored with a value of 1. This is a novel incentivized attention task based on attention checks designed to measure attention in forms typical of an office work environment. Scored as a binary (correct/incorrect) measure and agreegated into a composite score. The key to answering the question correctly is to notice that the contract which has a seemingly lower fee (Credit Card B) has a more frequent fee cycle, so it is more expensive than Credit Card A on an annual basis.

Credit Card Task

Your task: Decide which credit card below the firm should adopt to minimize cost. You get $0.25 cents if your answer is correct.

Option ID	A	B
APR for Purchases	19.99%	19.99%
APR for Transfers	11.99%	11.99%
Fee	$99	$39
Grace Period	30 days	30 days
Due Date	14th day of each month	21st day of each month
Fee Cycle	Annually	Monthly
Minimum Payment	$15	$15

Based on the information shown in the table, which card should the firm adopt?

- ○ Card A
- ○ Card B
- ○ Cards A and B are equal in terms of cost

Figure 1: Attention Task 1: Credit Card Choice Task

3.6 Attention Task 2: Workplace Email Task

In the second attention task, participants were presented with a workplace email scenario. They had to choose the correct response to an email thread, where the initial email inaccurately summarizes the prior conversation. The correct answer was evident if the participant thoroughly reads the forwarded conversation. The order of the two attention tasks was counterbalanced so that half of the participants saw the workplace email task first.

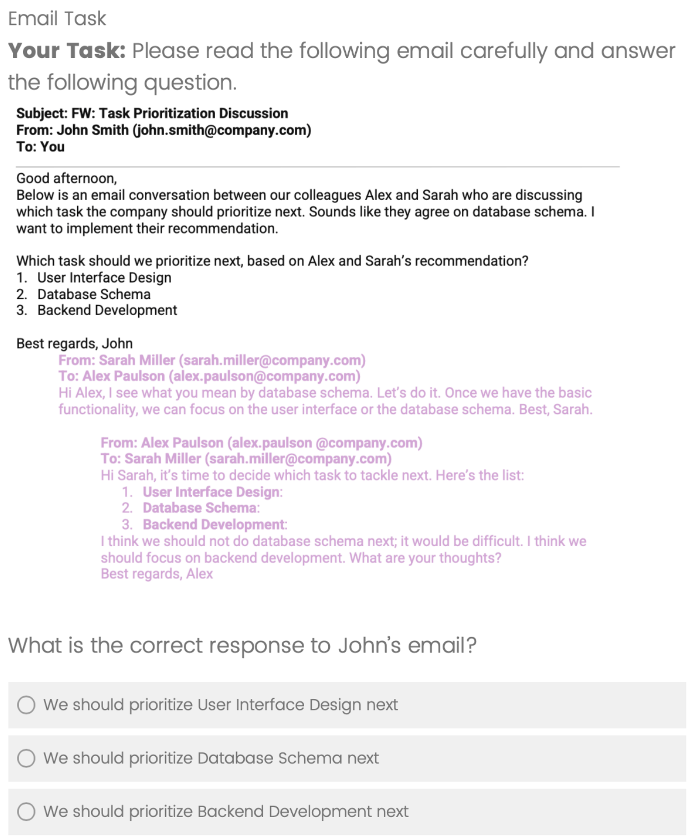

Figure 2: Attention Task 2: Workplace Email Task

Neither of the attention tasks were perceived as difficult by the participants. Median score was 2 out of 5 (somewhat easy).

3.7 Empirical Analysis

To understand the effects of unfair treatment on attention, I employ a specification that compares outcomes between workers assigned to the treatment arms (*Unfair-OutGroup* and *Fair-OutGroup*) and those assigned to the *Control* arm.

The primary analysis tests the effect of unfair treatment on participants' attention task performance. The main model specification is as follows:

$$Y_i = \alpha + \beta_1 UnfairOutGroup_i + \beta_2 FairOutGroup_i + \gamma X_i + \epsilon_i$$

Y_i is the composite attention score for participant i, $UnfairOutGroup_i$ is a binary indicator for assignment to the *Unfair-Outgroup* treatment, $FairOutGroup_i$ is a binary indicator for assignment to the *Fair-OutGroup* treatment, X_i is a vector of control variables for baseline characteristics (age, gender, political affiliation, race, and education), and ϵ_i is the participant-level error term (no clustering).

In the unadjusted model (without covariates), the model is simply

$$Y_i = \alpha + \beta_1 UnfairOutGroup_i + \beta_2 FairOutGroup_i + \epsilon_i$$

without the X_i of control variables.

Based on the theoretical framework in section 2, there are two primary hypotheses and testable predictions in this empirical design:

H1 (Unfair treatment lowers attention). Because the attention cost parameter is $\mu > 0$ only under the *Unfair–Outgroup* condition, participants in this group should exhibit lower average attention and performance compared to both the control and *Fair-OutGroup* conditions. Formally, the coefficient on the unfair treatment indicator should be negative:

$$H1: \ \beta_1 < 0$$

and performance under unfair treatment should be significantly lower than under fair treatment:

$$\beta_1 - \beta_2 < 0.$$

H2 (Fair treatment does not reduce attention). Under the *Fair–Outgroup* condition, no attention cost is incurred ($\mu = 0$), so fair treatment should not reduce attention relative to the control group and should result in higher performance relative to unfair treatment:

$$H2: \ \beta_2 = 0$$

and

$$\beta_2 - \beta_1 > 0.$$

The data collected from the experiment was analyzed with OLS, Tobit, and negative Binomial regressions to assess the impact of unfair treatment on attention-related performance. The primary independent variable was the treatment condition. Task performance on the real-effort task was controlled through the design of the experiment as all participants completed the same number of real-effort units.

4 Results

4.1 Summary Statistics

Appendix Tables 1-3 show that after random assignment of participants, there are no significant differences between the control and treatment groups based on observable demographic characteristics, so I compare the groups without any correction or reweighing procedures. Participants received an average payment of $2.00. The sample is representative of the U.S. population. In the full sample, 50% of participants are female, 32% are a non-white minority, 44% are politically conservative, 49% are politically liberal, 59% are college educated or higher, and 62% are 35 or older. Average duration of the experiment was $13 - 14$ minutes.

4.2 Main Findings

I begin with analyzing the treatment effect on the key dependent variable, *Composite Attention Score*. This outcome is a sum of binary (0/1) scores measuring the number of correct answers on each of the two attention tasks for each participant. Thus, *Composite Attention Score* is a discrete integer variable that takes values $\{0, 1, 2\}$. Next I analyze the effects on two measures of cognitive performance, Raven's Progressive Matrices (RPM) and the Cognitive Reflection Test (CRT).

4.2.1 Performance on Attention Tasks By Treatment

Using an OLS model (Columns 1 and 2 of Table 2), the unconditional average composite score on the two attention tasks is 1.186 for the baseline control group (95% CI $[1.12, 1.25]$) and 1.102 for the *Unfair-OutGroup* group (95% CI $[1.12, 1.25]$). The difference (0.084) is statistically significant at the 0.05 level ($\rho = 0.0395$). This difference is equivalent to a 7% reduction in performance when compared to the control group mean. After including controls for age, gender, political affiliation, race, and education, the treatment effect remains robust, with a magnitude of $\beta_1 = -0.08$. The inclusion of covariates does not alter the primary finding, suggesting that the effect of unfair treatment on attention performance is not driven by confounding demographic factors.

The composite attention score is evaluated discretely from 0 to 2. Thus, I also employ a Tobit model specification with upper and lower bounds imposed at 0 and 2 respectively, as a more accurate estimate than OLS. The treatment effect is even larger (Columns 3-4 of Table 9). Because the outcome measure is a count variable, the most preferred specification is a Negative Binomial regression and I find similar significant results (Columns 5-6 of Table 9). The negative binomial regression is shown in terms of incident rate ratios. The *Unfair-OutGroup* condition had an average composite attention score of 1.10, whereas the control group had an average score of 1.19. The difference in means is statistically significant ($\rho = 0.04$).

Importantly, when comparing the *Fair-OutGroup* and *Control* groups, the difference is not statistically significant. The unconditional average composite score for the *Fair-*

Table 2: Composite Attention Score as Function of Treatments

DV: Composite Attention Score

	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	1.19***	1.18***	1.42***	1.42***	1.19***	1.18**
	(0.03)	(0.08)	(0.08)	(0.18)	(0.03)	(0.09)
Unfair-OutGroup	-0.08**	-0.08*	-0.19**	-0.18*	0.93**	0.93*
	(0.04)	(0.04)	(0.09)	(0.09)	(0.03)	(0.03)
Fair-OutGroup	-0.01	-0.02	-0.02	-0.04	0.99	0.98
	(0.04)	(0.04)	(0.10)	(0.10)	(0.04)	(0.03)
Demographic Controls	NO	YES	NO	YES	NO	YES
Estimation Method	OLS	OLS	Tobit	Tobit	N.B.Reg	N.B.Reg
Participants	2303	2303	2303	2303	2303	2303

*** p<.01, ** p<.05, * p<.1. Robust standard errors shown in parentheses. Dependent variable is composite attention score measured at the individual level as a discrete integer value (0, 1, or 2). Unfair-OutGroup and Fair-OutGroup are indicator variables for treatment assignment. Omitted group is Control. Demographic controls are age, gender, political affiliation, race, and college education. Estimation methods are Ordinary Least Squares (OLS), Tobit, and Negative Binomial Regression (N.B.Reg). Negative binomial regression is shown in terms of incident rate ratios.

OutGroup group is 1.174 (95% CI $[1.12, 1.22]$). Compared to the *Control* group mean of 1.186, the difference (0.01) is not statistically significant at the 0.05 level ($\rho = 0.678$). This difference is equivalent to a 1% reduction in performance when compared to the *Control* group mean. Compared to the *Unfair-OutGroup*, the difference (0.07) is also statistically significant at the 0.05 level ($\rho = 0.0361$).

63.4% of all participants answered the credit card attention task correctly. The email attention task was slightly more difficult: 52.15% of participants answered it correctly. The pairwise Spearman correlation between the credit card task and email task performance is weakly positive and significant, with a Spearman correlation of 0.1016 ($\rho < 0.01$). Because there were three possible responses for each of the two tasks, the chance of getting both right at random is 11.12%.

4.2.2 Cognitive Reflection Test (CRT) Performance

I then examined performance on secondary cognitive tasks, such as the Cognitive Reflection Test (CRT). This task was moderately correlated with the main attention score ($\rho = 0.13$).

The difference in CRT scores between the control group and the Unfair-OutGroup participants is relatively small (Table 3) and not significant, suggesting the cognitive impact of unfair treatment is more task-specific than reflective across all cognitive domains.

The unconditional average composite score on the four CRT questions is 1.872 for the baseline control group and 1.879 for the *Unfair-OutGroup* condition (95% CI $[1.740, 2.018]$). The OLS difference ($\beta_1 = 0.13$) is not statistically significant at the 0.05 level. After including controls for age, gender, political affiliation, race, and education, the treatment effect remains insignificant, with a magnitude of $\beta_2 = -0.19$. The inclusion of covariates does not alter the findings. For the Tobit and Negative Binomial regressions, the coefficients are also not significant ($\rho > 0.05$) with conditional coefficients of $\beta_1 = 0.08$ and $\beta_2 = 1.04$ respectively, in line with the smaller magnitudes of the OLS regression results.

Table 3: Cognitive Reflection Test Score as Function of Treatments

DV: Cognitive Reflection Test Score

	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	1.87***	1.83***	2.64***	2.45***	1.87***	1.80***
	(0.06)	(0.13)	(0.16)	(0.33)	(0.06)	(0.13)
Treatment						
Unfair-OutGroup	0.01	0.01	-0.08	-0.08	1.00	1.00
	(0.07)	(0.07)	(0.18)	(0.18)	(0.04)	(0.04)
Fair-OutGroup	0.08	0.08	0.19	0.20	1.04	1.04
	(0.07)	(0.07)	(0.18)	(0.18)	(0.04)	(0.04)
Demographic Controls	NO	YES	NO	YES	NO	YES
Estimation method	OLS	OLS	Tobit	Tobit	N.B.Reg	N.B.Reg
Participants	2303	2303	2303	2303	2303	2303

*** p<.01, ** p<.05, * p<.1. Robust standard errors shown in parentheses. Dependent variable is Cognitive Reflection Test score measured at the individual level as a discrete integer value (0, 1, 2, 3, or 4). Unfair-OutGroup and Fair-OutGroup are indicator variables for treatment assignment. Omitted group is Control. Demographic controls are age, gender, political affiliation, race, and college education. Estimation methods are Ordinary Least Squares (OLS), Tobit, and Negative Binomial Regression (N.B.Reg). Negative binomial regression is shown in terms of incident rate ratios.

4.2.3 Raven's Progressive Matrices Test (RPM) Performance

I next present the results to estimate the effect of unfair treatment on Raven's Progessive Matrices (RPM) score. The first OLS model without demographic controls shows a statistically significant negative effect of treatment on RPM scores ($\beta = -0.20$, $p < 0.01$), indicating that participants who experienced unfair treatment had, on average, lower RPM scores by 0.10 points compared to the control group (Table 4). When demographic controls are included, the effect of treatment remains statistically significant ($\beta = -0.16$, $\rho < 0.05$).

Table 4: Raven's Progressive Matrices Score as Function of Treatments

DV: Raven's Progressive Matrices Score

	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	2.46***	2.60***	3.74***	4.08***	2.46***	2.59***
	(0.06)	(0.13)	(0.18)	(0.34)	(0.06)	(0.14)
Treatment						
Unfair-OutGroup	-0.18***	-0.17**	-0.40**	-0.38**	0.93***	0.93**
	(0.07)	(0.07)	(0.19)	(0.18)	(0.03)	(0.03)
Fair-OutGroup	-0.05	-0.08	-0.06	-0.19	0.98	0.97
	(0.07)	(0.07)	(0.19)	(0.19)	(0.03)	(0.03)
Demographic Controls	NO	YES	NO	YES	NO	YES
Estimation Method	OLS	OLS	Tobit	Tobit	N.B.Reg	N.B.Reg
Participants	2303	2303	2303	2303	2303	2303

$***p < .01, **p < .05, *p < .1$. Robust standard errors shown in parentheses. Dependent variable is Raven's Progressive Matrices score measured at the individual level as a discrete integer value (0, 1, 2, 3, or 4). Unfair-OutGroup Fair-OutGroup are indicator variables for treatment assignment. Omitted group is Control. Demographic controls are age, gender, political affiliation, race, and college education. Estimation methods are Ordinary Least Squares (OLS), Tobit, and Negative Binomial Regression (N.B.Reg). Negative binomial regression is shown in terms of incident rate ratios.

The Tobit model also shows a negative and statistically significant treatment effect. Without demographic controls, the estimated coefficient on *Unfair-OutGroup* is $\beta = -0.35$ ($\rho < 0.05$) . This indicates a larger negative effect compared to the OLS model. With demographic controls, the treatment effect loses significance ($\beta = -0.28$) but is still a large negative effect compared to the OLS results.

In the Negative Binomial regression model, which is shown in terms of incident rate ratios, the treatment effect is less than 1 and significant. Without demographic controls, the treatment effect is $\beta_1 = 0.92$ ($\rho < 0.01$), indicating that participants in the treatment group had significantly lower predicted RPM scores. When demographic controls are included, the treatment effect remains at $\beta_1 = 0.93$ ($\rho < 0.05$).

Across all model specifications (OLS, Tobit, and Negative Binomial), the treatment effect is consistently negative and statistically significant, suggesting that unfair treatment has a detrimental impact on participants' attention as measured by RPM.

4.3 Mediation Analysis

4.3.1 Perceived Discrimination

The proportion of *Fair-OutGroup* participants that perceive any discrimination is 12% (n=109) compared to 44% (n=414) for the *Unfair-OutGroup*. This proportion drops to 2% versus 15% when for participants that strongly agree that discrimination occurred.

Table 5 shows that when controlling for perceived discrimination, the effect of *Unfair-OutGroup* on the composite attention score becomes insignificant (Columns 2, 4, and 6), and the coefficient on *High Perceived Discrimination* is negative and significant, suggesting that the treatment is more pronounced for those who perceive discrimination. This implies a plausible mechanism for the treatment effect, and is in line with predictions that perceiving unfair treatment (of which perceived discrimination is an extreme form) could be causing reduced performance on the attention tasks. The interaction term between treatment status and perceived discrimination is statistically significant suggesting that those who are more sensitive to discrimination are more vulnerable to attention lapses in the face of unfair treatment. Among racial minorities, the treatment effect was more pronounced, especially for participants identifying as Black or African American, whose composite score dropped significantly ($\beta_1 = -0.34$ compared to $\beta_1 = -0.08$).

Table 5: Composite Attention Score as Function of Treatments
Interaction with Perceived Discrimination

	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	1.19***	1.19***	1.41***	1.41***	1.19***	1.19***
	(0.03)	(0.03)	(0.08)	(0.08)	(0.03)	(0.03)
Unfair-OutGroup	-0.00	-0.00	-0.01	-0.01	1.00	1.00
	(0.05)	(0.05)	(0.12)	(0.12)	(0.05)	(0.05)
Unfair-OutGroup #	-0.12**	-0.12**	-0.26**	-0.26**	0.90**	0.90**
High Perceived Discrimination	(0.05)	(0.05)	(0.11)	(0.11)	(0.04)	(0.04)
Demographic Controls	NO	YES	NO	YES	NO	YES
Participants	3362	3362	3362	3362	3362	3362
Estimation Method	OLS	OLS	Tobit	Tobit	N.B.Reg	N.B.Reg

*** p<.01, ** p<.05, * p<.1. Robust standard errors shown in parentheses. Dependent variable is composite attention score measured at the individual level as a discrete integer value (0, 1, or 2). Unfair-OutGroup and Unfair-InGroup are indicator variables for treatment assignment. Omitted group is Control. Demographic controls are age, gender, political affiliation, race, and college education. Estimation methods are Ordinary Least Squares (OLS), Tobit, and Negative Binomial Regression (N.B.Reg). Negative binomial regression is shown in terms of incident rate ratios.

4.3.2 Beliefs

One possibility is that the observed effects of unfair treatment on attention could be driven by participants' prior beliefs about how common unfair behavior is in the population, rather than by the experience of being unfairly treated itself. Thus, the degree to which the outcome is surprising could be the source of distraction. To address this concern, I elicited participants' beliefs about the prevalence of unfair behavior and re-estimated the main results controlling for this belief measure. Because this belief is not elicited for the control group, I compare the unfairly treated groups to the fairly treated group, using the *Fair-OutGroup* as the omitted baseline group. The inclusion of beliefs does not substantively change the results: participants who experienced fair treatment continued to exhibit significantly higher attention relative to the unfair treatments, and the effect was larger when the unfair treatment came from an in-group partner. This pattern suggests that the results are not explained by differences in expectations about the likelihood of unfair behavior, but rather by the direct attentional cost of experiencing unfair treatment.

Performance is not correlated with beliefs about how often an individual is treated unfairly. The (ex-post) belief about the likelihood that another participant would take is

Table 6: Composite Attention Score as Function of Treatments
Controlling for Beliefs

	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	1.41***	1.33***	1.94***	1.75***	1.43***	1.32***
	(0.05)	(0.10)	(0.13)	(0.23)	(0.06)	(0.12)
Unfair-OutGroup	-0.17*	-0.16*	-0.40*	-0.38*	0.87*	0.88*
	(0.09)	(0.09)	(0.21)	(0.21)	(0.06)	(0.07)
Controlling for Beliefs	YES	YES	YES	YES	YES	YES
Demographic Controls	NO	YES	NO	YES	NO	YES
Estimation Method	OLS	OLS	Tobit	Tobit	N.B.Reg	N.B.Reg
Participants	1823	1823	1823	1823	1823	1823

*** p < .01, ** p < .05, * p <.1. Robust standard errors shown in parentheses. Dependent variable is composite attention score measured at the individual level as a discrete integer value (0, 1, or 2).Unfair-OutGroup is an indicator variable for treatment assignment. Omitted group is Fair-OutGroup. Demographic controls are age, gender, political affiliation, race, college education, and beliefs about baseline likelihood of encountering unfairness. Estimation methods are Ordinary Least Squares (OLS), Tobit, and Negative Binomial Regression (N.B.Reg).

greater for the *Unfair-Outgroup* condition than the *Fair-Outgroup* condition (69% versus 54 %, $t = 12.357, \rho < 0.01$).

Table 6 reports the results. Columns (1) (3) and (5) reproduces the main specification without beliefs, while Columns (2) (4) and (6) add the belief control. Across specifications, the coefficients on the fair treatment condition remains stable in magnitude and significance, confirming that the effect of unfair treatment on attention is robust to controlling for participants' beliefs.

4.3.3 Affect

Next, I consider the possibility that the observed decline in task performance is not due to cognitive distraction per se, but rather to participants becoming emotionally upset by unfair treatment. To test this, I exclude participants who reported high levels of emotional upset immediately after treatment. As shown in Table 7, the negative effect of unfair treatment on attention persists even in this restricted sample, with effect sizes comparable to those in the full analysis. This finding suggests that emotional valence alone cannot account for the treatment effect. Instead, the evidence is more consistent with the interpretation that unfair treatment induces a sustained cognitive load, diverting

Table 7: Composite Attention Score as Function of Treatments
Interaction with Affect

	OLS	OLS	Tobit	Tobit	N.B.Reg	N.B.Reg
Intercept	1.19***	1.19***	1.41***	1.41***	0.17***	0.17***
	(0.03)	(0.03)	(0.08)	(0.08)	(0.03)	(0.03)
treated						
Unfair-OutGroup	-0.08*	-0.08*	-0.18**	-0.18**	-0.07**	-0.07**
	(0.04)	(0.04)	(0.09)	(0.09)	(0.04)	(0.04)
Unfair-OutGroup # High Upset	-0.02	-0.02	-0.04	-0.04	-0.02	-0.02
	(0.07)	(0.07)	(0.15)	(0.15)	(0.06)	(0.06)
treated						
Unfair-InGroup	-0.17***	-0.17***	-0.39***	-0.39***	-0.16***	-0.16***
	(0.04)	(0.04)	(0.09)	(0.09)	(0.04)	(0.04)
Unfair-InGroup # High Upset	-0.05	-0.05	-0.10	-0.10	-0.05	-0.05
	(0.06)	(0.06)	(0.12)	(0.12)	(0.06)	(0.06)
Demographic Controls	NO	YES	NO	YES	NO	YES
Emotional Upset	YES	YES	YES	YES	YES	YES
Participants	3364	3364	3364	3364	3364	3364

*** p < .01, ** p < .05, * p <.1. Robust standard errors shown in parentheses. Dependent variable is composite attention score measured at the individual level as a discrete integer value (0, 1, or 2). Unfair-OutGroup and Unfair-InGroup are indicator variables for treatment assignment. Omitted baseline is the control group. Demographic controls are age, gender, political affiliation, race, college education, and beliefs about baseline likelihood of encountering unfairness. High Upset is a binary indicator for individual score equal to the highest value on a 4-point likert scale. Estimation methods are Ordinary Least Squares (OLS), Tobit, and Negative Binomial Regression (N.B.Reg).

scarce attentional resources away from task-relevant processing and thereby lowering workplace performance.

4.4 Heterogeneity Analysis

Next, I conduct subgroup analysis based on participants' demographic characteristics, stratifying by race, gender, and education level. The purpose of these analyses was to investigate whether certain subpopulations were more susceptible to the cognitive costs of unfair treatment.

Race and Perceived Discrimination. There are significant differences across racial groups. For example, Black or African American participants had lower average scores (mean = 0.87) while White and Asian participants performed better (mean = 1.18, mean = 1.29 respectively). This may reflect broader socio-cognitive factors, such as the cognitive toll of discrimination. One plausible scenario is that individuals who reported higher levels of perceived discrimination would exhibit stronger negative effects on attention performance following unfair treatment.

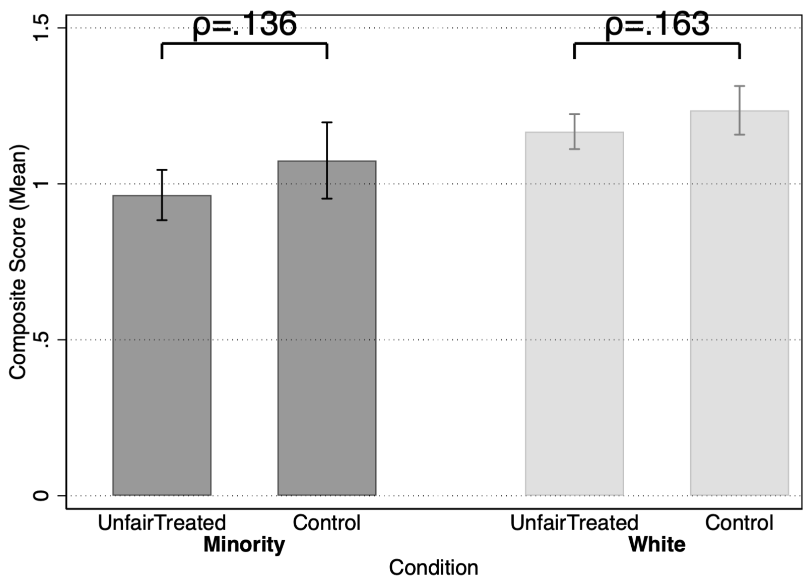

Figure 3: Mean Composite Score By Minority Status

Gender. There are minimal differences between male and female participants in terms of attention performance as shown in figure 4. On average, females had a performance score of 1.16, while males had a slightly lower average score of 1.13. In summary, there are small differences between male and female participants, with females slightly outperforming males on the attention tasks, though the gender differences in performance are minor.

Age. Younger participants tended to perform better on the attention tasks, with the 18-25 age group having the highest average score (mean = 1.28, SD = 0.75). Scores generally declined with age, with participants aged 65 and above showing the lowest average score (mean = 0.93, SD = 0.74). Participants in the 26-35 age group had the next lowest average score (mean = 1.09, SD = 0.75). Participants in the 36-55 age groups had relatively consistent scores, around 1.19 on average. There is a weak negative correlation between age and composite attention score (correlation = -0.054). This suggests that,

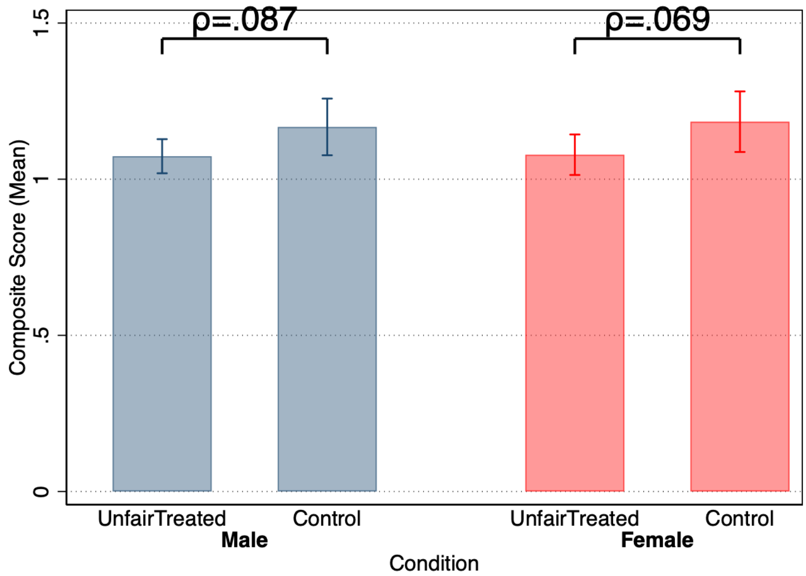

Figure 4: Mean Composite Score (By Gender)

overall, older participants have slightly lower attention scores, but the effect is relatively minor.

Political affiliation. While the differences are modest, Democrats performed slightly better than Republicans and Moderates in terms of attention scores. However, these differences are small and may not reflect substantial differences tied to political beliefs. Democrats had an average composite attention score = 1.22 (SD = 0.72), with 560 participants. This group performed the highest on average among the political affiliations. Moderates had average composite attention score = 1.13 (SD = 0.73), with 68 participants. Moderates scored slightly lower than Democrats but were similar to the overall mean. Republicans had average composite attention score = 1.09 (SD = 0.74), with 488 participants. Republicans had the lowest average score compared to Democrats and Moderates.

Political affiliation does appear to have a small impact on attention scores, with Democrats performing slightly better than Moderates and Republicans. However, the differences are relatively modest, and while they may reflect underlying ties to political beliefs, they are not large enough to suggest major disparities in attention performance based on political affiliation alone.

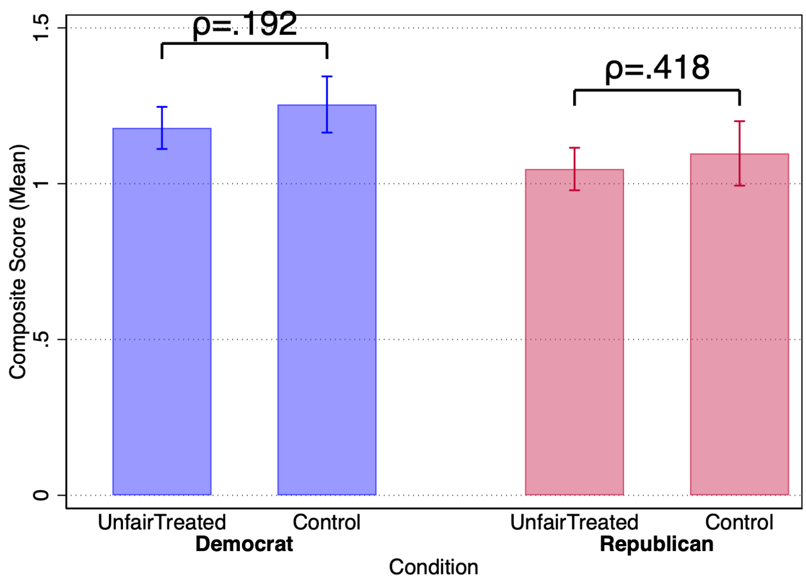

Figure 5: Mean Composite Score (By Political Affiliation)

4.5 Survey Results

Comparing the sources of possible discrimination, recipients ranked discrimination on the basis of their race the least likely to have occurred (race was not shared across pairings) and perceived discrimination on the basis of political affiliation the most likely. Age and sex discrimination were seen as more likely than racial discrimination but less likely than political discrimination.

Dictators. *Dictators* reported an average belief that 73% of all *Dictators* would take the financial bonus. 64.71% of *Dictators* chose to take the $0.75 financial bonus, less than the average reported belief of *Dictators* that 73% overall would take. Interestingly, this is also less than the average belief of *Recipients* (*Recipients* reported 69% of *Dictators* would take). Dictators over-predicted the likelihood that *Dictators* would take and over-predicted the proportion of recipients who believed they were discriminated against. 64.71% of dictators reported that the other participant would be "Somewhat likely" or "Extremely likely" to feel they were discriminated against, compared to the actual proportion of 43.66%. 70.58% of Dictators reported that *Recipients* would be "a little upset" or "not at all upset" by the unfair treatment, although the true percent was 62.24%. Thus, dictators underestimated how upset recipients would be by the unfair treatment.

Recipients 18.64% reported having difficulty with the Email Task, while 14.7% reported having difficulty with the Card Task. This proportion is significantly smaller than the percentage of participants who did not answer each of the tasks correctly. The correlation between subjective difficulty of these two tasks is 0.2628 ($\rho < 0.01$). 62.9% reported being at least somewhat upset by the treatment, suggesting multiple pathways in which attention is affected. The correlation between perceived discrimination and being upset is 0.2770 ($\rho < 0.01$). There is no significant difference in perceived difficulty by treatment.

4.6 Robustness Checks

I conduct several additional analyses on the duration and comprehension of participants in completing the attention tasks and understanding the task instructions. The full set of instructions and screenshots for each condition is available in the Appendix.

4.6.1 Timing Results

I first analyze whether the time participants spent on the attention tasks was correlated with performance. Participants who completed the tasks within a moderate time frame (500 to 1000 seconds) performed better than those who completed the tasks either too quickly or too slowly. This suggests that pacing plays a role in cognitive performance, where rushing or excessive time-taking may reduce focus. Participants who completed the tasks within 500 to 1000 seconds had the highest average composite attention score (mean = 1.20, SD = 0.73). Those who completed the tasks faster (0-500 seconds) had a lower mean score (mean = 0.95, SD = 0.75). Participants who took longer (1501-2000 seconds) had similarly lower average composite scores (mean = 0.96, SD = 0.72).

The analysis of timing data reveals that there is a very weak negative correlation between the time spent on the tasks (duration in seconds) and the composite attention score (correlation = -0.028). This suggests that, overall, time spent on the tasks does not strongly affect attention performance.

The *Unfair-Outgroup* treatment participants completed the task significantly faster than the control group participants 6. This shorter time spent on the attention tasks (even though equally incentivized as the control group) provides further evidence that the treatment is operating through less attention allocated to the tasks.

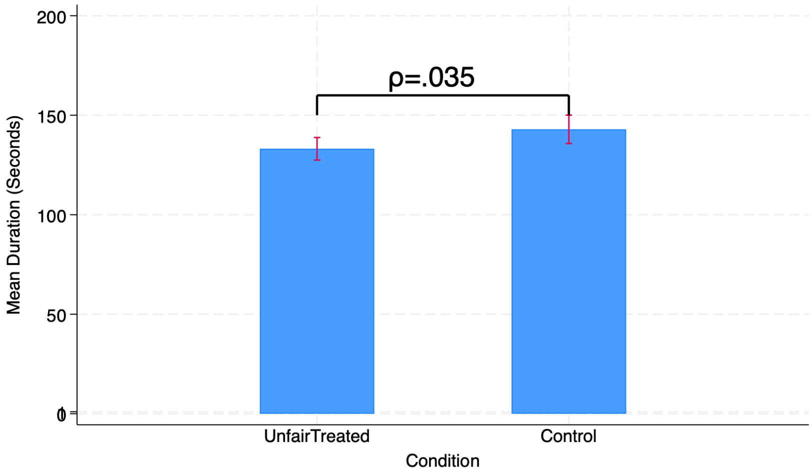

Figure 6: Mean Total Seconds on Attention Task

4.7 Exploratory Analysis

4.7.1 Wave 4: Unfair-InGroup Treatment

In a fourth wave of data collection, where all participants were assigned to the *Unfair-InGroup* condition, participants were informed that they would complete the same real-effort task as in Wave 1 and be paired with a participant from the *Dictator* sample. As in the Unfair-OutGroup condition, the probability of being matched with a Dictator who chose *Take* was 100%. Participants were told that their demographic characteristics (age group, political affiliation, and gender) would be shared with their partner, and that they would receive the same information about their partner. However, in this condition, the paired Dictator always shared the same demographic group membership on at least two salient dimensions (e.g., gender and political affiliation). This design created the potential for perceived discrimination within an in-group relationship, where unfair treatment may be experienced as a stronger violation of fairness norms.

Dictator	Recipient
Male Republican	Male Republican
Female Republican	Female Republican
Male Democrat	Male Democrat
Female Democrat	Female Democrat

Table 8: Dictator-Recipient Pairings (*Unfair-inGroup* Only)

Participants in the *Unfair-InGroup* condition then received instructions about the real-effort task, the possibility of having their earnings taken by a *Dictator* sample participant, and the subsequent attention tasks. They were shown the demographics of their paired partner to reinforce the in-group match. All participants performed the same real-effort task as in Wave 1, with the requirement of completing at least 200 units of effort (keyboard presses) within five minutes to proceed. Finally, participants were informed whether their earnings had been taken by their paired participant.

For the *Unfair-InGroup* treatments, the pairings were different. Table 8 shows the complete list of pairings, where now the recipients share the same political affiliation and gender as the dictator. The *Unfair-InGroup* treatment has 1,061 participants, and the sample size was determined by the effect sizes from the August 2024 data collection of the main study. It was pre-registered on aspredicted.org.[4] The average composite attention score for the *Unfair-InGroup* treatment group is 1.26 with a standard deviation of 0.73.

The *Unfair-InGroup* group has significantly stronger treatment effects across all specifications compared to the *Unfair-OutGroup* (Table 9). The unconditional average composite score on the two attention tasks is 1.00 for the for the *Unfair-OutGroup* group (95% CI $[0.924, 1.081]$). The OLS difference ($\beta_2 = 0.08$) is statistically significant at the 0.01 level ($\rho < 0.01$). This difference is equivalent to a 16% reduction in performance when compared to the control group mean. After including controls for age, gender, political affiliation, race, and education, the treatment effect remains robust, with a magnitude of $\beta_2 = -0.19$. The inclusion of covariates does not alter the findings. For the Tobit and Negative Binomial regressions, the coefficients are statistically significant at the 0.01 level

[4] Analysis Plan #209645. Link: https://aspredicted.org/3pff-z6bm.pdf

Table 9: Composite Attention Score as Function of Treatments

DV: Composite Attention Score

	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	1.19***	1.19***	1.41***	1.42***	1.19***	1.18***
	(0.03)	(0.07)	(0.08)	(0.15)	(0.03)	(0.07)
Unfair-OutGroup	-0.10**	-0.08**	-0.22**	-0.20*	0.92**	0.93**
	(0.04)	(0.04)	(0.10)	(0.10)	(0.04)	(0.04)
Unfair-InGroup	-0.18***	-0.19***	-0.41***	-0.41***	0.84***	0.84***
	(0.04)	(0.04)	(0.09)	(0.09)	(0.03)	(0.03)
FairTreated	0.01	0.01	0.03	0.03	1.01	1.01
	(0.04)	(0.04)	(0.09)	(0.09)	(0.03)	(0.03)
Demographic Controls	NO	YES	NO	YES	NO	YES
Estimation Method	OLS	OLS	Tobit	Tobit	N.B.Reg	N.B.Reg
Participants	3364	3364	3364	3364	3364	3364

*** p<.01, ** p<.05, * p<.1. Robust standard errors shown in parentheses. Dependent variable is composite attention score measured at the individual level as a discrete integer value (0, 1, or 2). Unfair-OutGroup, Unfair-InGroup, and FairTreated are indicator variables for treatment assignment. Omitted group is Control. Demographic controls are age, gender, political affiliation, race, and college education. Estimation methods are Ordinary Least Squares (OLS), Tobit, and Negative Binomial Regression (N.B.Reg). Negative binomial regression is shown in terms of incident rate ratios.

($\rho < 0.01$) with coefficients of $\beta_2 = -0.41$ and $\beta_2 = 0.84$ respectively, in line with the larger magnitudes of the OLS regression results.

5 Conclusion

The results have important implications for the design of workplace incentives. Standard economic models often assume that worker effort responds primarily to pecuniary rewards, such as wages, bonuses, or promotion opportunities. I suggest that perceptions of fairness can undermine the effectiveness of these incentives by draining attention and impairing performance. Even when monetary incentives are present, unfair treatment creates a cognitive tax that reduces the ability of workers to respond to those incentives effectively.

One implication is that organizations relying heavily on performance-based pay may inadvertently erode its effectiveness if employees perceive the allocation process as unfair. For example, if workers suspect bias in the distribution of bonuses or assignments, they may be less able to concentrate on the tasks that determine those outcomes. The result is lower productivity and weaker alignment between effort and reward. This dynamic helps explain why two firms offering similar incentive pay structures may observe different performance outcomes depending on whether employees view the environment as procedurally fair.

A second implication is that non-pecuniary aspects of the work environment can complement or even substitute for financial incentives. Transparent evaluation procedures, equitable workload distribution, and clear communication channels may reduce perceptions of unfairness and thereby mitigate the attentional costs that I document here. In such settings, workers are better able to channel effort toward productive tasks, making pecuniary incentives more effective. By contrast, environments rife with perceived unfairness may require higher financial rewards to achieve the same level of output, leading to inefficiency and greater wage dispersion without corresponding productivity gains.

Finally, these results underscore that fairness is not merely a moral or legal imperative but an economic one. Employers concerned with maximizing productivity should treat perceptions of fairness as a core component of their incentive design. By reducing cognitive distraction, equitable workplace practices can amplify the effectiveness of both monetary and non-monetary incentives, leading to more efficient and inclusive organizations.

This paper provides causal evidence that perceived unfair treatment reduces attention, leading to measurable declines in workplace-relevant performance. Effects are especially pronounced when unfairness comes from in-group members or is attributed to discrimination, and they persist even after excluding participants who reported feeling upset. The pattern across tasks suggests that unfair treatment specifically impairs attention and abstract reasoning rather than general cognitive ability.

These findings highlight an important but overlooked mechanism: unfairness diverts scarce attentional resources, imposing hidden productivity costs on workers and firms. Because fair treatment establishes a baseline rather than a performance boost, organizations that tolerate unfair dynamics may underestimate the true economic costs of these environments.

6 Appendix

Table 1: Balance Table - Control and Unfair-OutGroup

	Total	Control	Unfair-OutGroup	p-value
Female	0.50 (0.50)	0.50 (0.50)	0.50 (0.50)	0.985
Nonwhite	0.32 (0.47)	0.31 (0.46)	0.32 (0.47)	0.534
Hispanic	0.07 (0.26)	0.06 (0.24)	0.08 (0.27)	0.220
Native American	0.02 (0.12)	0.02 (0.13)	0.01 (0.12)	0.782
Politically conservative	0.44 (0.50)	0.43 (0.50)	0.45 (0.50)	0.501
Politically liberal	0.49 (0.50)	0.51 (0.50)	0.47 (0.50)	0.182
College educated or higher	0.59 (0.49)	0.56 (0.50)	0.60 (0.49)	0.154
Aged 35 or older	0.62 (0.48)	0.62 (0.49)	0.63 (0.48)	0.762

Mean (Standard deviation): ρ-value from a two-sample t test with unequal variances.

Table 2: Balance Table - Control and Fair-OutGroup

	Total	Control	FairTreat	p-value
Female	0.50 (0.50)	0.50 (0.50)	0.50 (0.50)	0.873
Hispanic	0.06 (0.24)	0.06 (0.24)	0.07 (0.25)	0.597
Native American	0.01 (0.12)	0.02 (0.13)	0.01 (0.12)	0.651
Politically conservative	0.44 (0.50)	0.43 (0.50)	0.45 (0.50)	0.608
Politically liberal	0.49 (0.50)	0.51 (0.50)	0.48 (0.50)	0.314
College educated or higher	0.59 (0.49)	0.56 (0.50)	0.61 (0.49)	0.144
Aged 35 or older	0.63 (0.48)	0.62 (0.49)	0.64 (0.48)	0.459

Mean (Standard deviation): ρ-value from a two-sample t test with unequal variances.

Table 3: Balance Table - Fair-OutGroup and Unfair-OutGroup

	Total	UnfairTreat	FairTreat	p-value
Female	0.50 (0.50)	0.50 (0.50)	0.50 (0.50)	0.829
Hispanic	0.07 (0.26)	0.08 (0.27)	0.07 (0.25)	0.394
Native American	0.01 (0.12)	0.01 (0.12)	0.01 (0.12)	0.832
Politically conservative	0.45 (0.50)	0.45 (0.50)	0.45 (0.50)	0.854
Politically liberal	0.48 (0.50)	0.47 (0.50)	0.48 (0.50)	0.707
College educated or higher	0.60 (0.49)	0.60 (0.49)	0.61 (0.49)	0.950
Aged 35 or older	0.63 (0.48)	0.63 (0.48)	0.64 (0.48)	0.593

Mean (Standard deviation): ρ-value from a two-sample t test with unequal variances.

Table 4: Variation in instructions by treatment. Instructions were shown in an HTML webpage on Qualtrics.

Condition	Instruction
Dictator	Before the keyboard pressing task, you will be paired with another person in the study. You will not be told who you are matched with during or after the study, and he or she will not be told who you are either during or after the study. Your decisions will be strictly anonymous and cannot be linked to you in any way. You must choose between the following two options: "Take" Option: Take $0.75 cents from the other participant, do not do any pairs and continue automatically. "Not Take" Option: Do 200 keyboard pairs in 5 minutes, earn $0.75 cents, and continue. The other person you are paired with: Has to do 200 keyboard pairs in 5 minutes to potentially earn $0.75 cents regardless of your decision, knows that whether they get the $0.75 cents or not depends on your decision, and learn if their $0.75 cents was taken or not after 5 minutes. Before you make your decision, you will be given some information about the other participant. Likewise, the other participant will be given some information about you. The other participant knows that this information has been shared with you and that they will receive the same information about you. We will now ask you some questions to make sure you understand these instructions.
Control	After the keyboard pressing task, you will learn if a green ball or a black ball is drawn from a urn. The urn is filled with 10 balls total. The number of green balls is unknown and can be zero, 10, or any whole number in between. The number of black balls is unknown and can be zero, 10 or any whole number in between. One ball is drawn from the urn. You will learn what color of ball is drawn from the urn after completing the keyboard pressing task. The color of the ball has no impact on your earnings or opportunities in this study.
UnfairTreated	Before the keyboard pressing task, you will be paired with another person in the study. You will not be told who you are matched with during or after the study, and he or she will not be told who you are either during or after the study. Your decisions will be strictly anonymous and cannot be linked to you in any way. The other person you are paired with has been asked to do the same keyboard pressing task as you. However, they can decide to take your financial bonus and not do the work task for themself. You will only learn their decision after you complete the task. After you learn the other participant's decision, you will immediately continue with the study.
FairTreated	Before the keyboard pressing task, you will be paired with another person in the study. You will not be told who you are matched with during or after the study, and he or she will not be told who you are either during or after the study. Your decisions will be strictly anonymous and cannot be linked to you in any way.

Table 5: Table of comprehension questions (Control group)

Label	Question Text	Responses
Comprehension Question 1	How many different types of color balls are there in the urn?	1 / 2 / 3 / 4
Comprehension Question 2	The total number of green and black balls in the urn is...	5 / 10 / 100 / Unknown
Comprehension Question 3	True or False: I will learn the color of the ball from the urn before completing the keyboard pressing task.	True / False

Table 6: Table of comprehension questions (Dictator group)

Label	Question Text	Responses
Comprehension Question 1	If you take the $0.75 from the other person, what will happen?	The other participant does not get $0.75 and I skip the keyboard pressing task / The other participant gets $0.75 and I do 5 minutes of the keyboard pressing task / The other participant gets $0.75 and I skip the keyboard pressing task / The other participant does not get $0.75 and I do 5 minutes of the keyboard pressing task
Comprehension Question 2	If you DO NOT take the $0.75 from the other person, what will happen?	I do 5 minutes of the keyboard pressing task and will continue with the study even if I do not complete at least 200 pairs. / I skip the keyboard pressing task and earn $0.00 / I do 5 minutes of the keyboard pressing task and can only continue with the study if I complete at least 200 pairs / I skip the keyboard pressing task and earn $0.75
Comprehension Question 3	True or False: The other person you are paired with must do 200 keyboard pairs in 5 minutes no matter what decision you make.	True /False
Comprehension Question 4	True or False: If you do not take the $0.75 for yourself and do the keyboard pressing task, you can choose to end the keyboard pressing task without completing 200 keyboard pairs before 5 minutes are up.	True / False

Table 7: Table of comprehension questions (Treatment groups)

Label	Question Text	Responses
Comprehension Question 1	If the other person takes the $0.75 bonus, what will happen?	I skip the keyboard pressing task and I do not get the $0.75 cent bonus / I skip the keyboard pressing task and I get the $0.75 cent bonus / I do the keyboard pressing task and I do not get the $0.75 cent bonus / I do the keyboard pressing task and I get the $0.75 cent bonus
Comprehension Question 2	If the other person DOES NOT take $0.75 cents, what will happen?	I do the keyboard pressing task and earn the $0.75 if I complete at least 200 pairs / I do the keyboard pressing task and do not earn the $0.75 cents if I complete at least 200 pairs / I skip the keyboard pressing task and earn the $0.75 cents / I skip the keyboard pressing task and do not earn the $0.75 cents
Comprehension Question 3	True or False: You must do the keyboard pressing task no matter what decision the other person makes	True/False
Comprehension Question 4	True or False: The other person must do the keyboard pressing task no matter what decision they make	True/False

References

Bartling, Björn, Ernst Fehr, Michel André Maréchal, and Daniel Schunk. 2009. "Egalitarianism and competitiveness." *American Economic Review*, 99(2): 93–98.

Becker, Gary S. 2010. *The economics of discrimination.* University of Chicago press.

Bertrand, Marianne, and Esther Duflo. 2017. "Field experiments on discrimination." *Handbook of economic field experiments*, 1: 309–393.

Blau, Francine D, and Lawrence M Kahn. 2017. "The gender wage gap: Extent, trends, and explanations." *Journal of economic literature*, 55(3): 789–865.

Brañas-Garza, Pablo, Praveen Kujal, and Balint Lenkei. 2019. "Cognitive reflection test: Whom, how, when." *Journal of Behavioral and Experimental Economics*, 82: 101455.

Burfurd, Ingrid, and Tom Wilkening. 2022. "Cognitive heterogeneity and complex belief elicitation." *Experimental economics*, 25(2): 557–592.

Cappelen, Alexander W, Johanna Mollerstrom, Bjørn-Atle Reme, and Bertil Tungodden. 2022. "A meritocratic origin of egalitarian behaviour." *The Economic Journal*, 132(646): 2101–2117.

Carpenter, Patricia A, Marcel A Just, and Peter Shell. 1990. "What one intelligence test measures: a theoretical account of the processing in the Raven Progressive Matrices Test." *Psychological review*, 97(3): 404.

Cohn, Alain, Ernst Fehr, Benedikt Herrmann, and Frédéric Schneider. 2011. "Social comparison in the workplace: evidence from a field experiment." *Working paper series/Department of Economics*, , (07).

DellaVigna, Stefano, and Devin Pope. 2018. "What motivates effort? Evidence and expert forecasts." *The Review of Economic Studies*, 85(2): 1029–1069.

Eysenck, Michael W, Nazanin Derakshan, Rita Santos, and Manuel G Calvo. 2007. "Anxiety and cognitive performance: attentional control theory." *Emotion*, 7(2): 336.

Falk, Armin, and Urs Fischbacher. 2002. "The economics of reciprocity: evidence and theory." *Inequality around the World*, 207–233.

Fehr, Ernst, and Klaus M Schmidt. 1999. "A theory of fairness, competition, and cooperation." *The quarterly journal of economics*, 114(3): 817–868.

Fershtman, Chaim, and Uri Gneezy. 2001. "Discrimination in a segmented society: An experimental approach." *The Quarterly Journal of Economics*, 116(1): 351–377.

Frederick, Shane. 2005. "Cognitive reflection and decision making." *Journal of Economic perspectives*, 19(4): 25–42.

Gabaix, Xavier, and David Laibson. 2005. "Bounded rationality and directed cognition." *Harvard University*, 20.

Gächter, Simon, Daniele Nosenzo, and Martin Sefton. 2012. "The impact of social comparisons on reciprocity." *The Scandinavian Journal of Economics*, 114(4): 1346–1367.

Houser, Daniel, and Robert Kurzban. 2002. "Revisiting kindness and confusion in public goods experiments." *American Economic Review*, 92(4): 1062–1069.

John, and Jean Raven. 2003. "Raven progressive matrices." In *Handbook of nonverbal assessment.* 223–237. Springer.

Kahneman, Daniel, Jack L Knetsch, and Richard H Thaler. 1986. "Fairness and the assumptions of economics." *Journal of Business*, S285–S300.

Kline, Patrick, Evan K Rose, and Christopher R Walters. 2022. "Systemic discrimination among large US employers." *The Quarterly Journal of Economics*, 137(4): 1963–2036.

Krueger, Alan B, and Alexandre Mas. 2004. "Strikes, scabs, and tread separations: Labor strife and the production of defective Bridgestone/Firestone tires." *Journal of Political Economy*, 112(2): 253–289.

Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt. 2023. "Rational inattention: A review." *Journal of Economic Literature*, 61(1): 226–273.

Mullainathan, Sendhil, and Eldar Shafir. 2013. *Scarcity: Why having too little means so much.* Macmillan.

Neumark, David. 2018. "Experimental research on labor market discrimination." *Journal of Economic Literature*, 56(3): 799–866.

Peer, Eyal, David Rothschild, Andrew Gordon, Zak Evernden, and Ekaterina Damer. 2022. "Data quality of platforms and panels for online behavioral research." *Behavior research methods*, 54(4): 1643–1662.

Pennycook, Gordon, and David G Rand. 2021. "The psychology of fake news." *Trends in cognitive sciences*, 25(5): 388–402.

Ruebeck, Hannah. 2025. "Causes and Consequences of Perceived Workplace Discrimination." *Available at SSRN 4799864.*

Salvatore, Jessica, and J Nicole Shelton. 2007. "Cognitive costs of exposure to racial prejudice." *Psychological science*, 18(9): 810–815.

Sloan, Melissa M. 2012. "Unfair treatment in the workplace and worker well-being: The role of coworker support in a service work environment." *Work and Occupations*, 39(1): 3–34.

Whitmer, Anson J, and Ian H Gotlib. 2013. "An attentional scope model of rumination." *Psychological bulletin*, 139(5): 1036.

Xiao, Erte, and Daniel Houser. 2005. "Emotion expression in human punishment behavior." *Proceedings of the National Academy of Sciences*, 102(20): 7398–7401.

Zahodne, Laura B, Emily P Morris, Neika Sharifian, Afsara B Zaheed, A Zarina Kraal, and Ketlyne Sol. 2020. "Everyday discrimination and subsequent cognitive abilities across five domains." *Neuropsychology*, 34(7): 783.