

Discriminatory Enforcement of Consumer Contracts: Evidence from the Retail Market

Meirav Furth-Matzkin*

October 2022

Abstract

Sellers often authorize their employees to deviate from the exact terms of their form contracts when interacting with consumers. This study tests whether, when departing from the letter of a seller’s form contract, salespeople treat white consumers disproportionately more favorably than similarly situated Black consumers. Nineteen testers were recruited and trained to return more than 200 unworn clothing items to 59 retail stores in Chicago. The testers tried to return the items without presenting a receipt, although the stores had formal policies requiring a receipt as a condition to making a return. The findings reveal significant racial disparities in the salespeople’s willingness to accommodate the tester. In this study, Black testers were 7 percentage points less likely than comparable white testers to obtain an immediate refund. Moreover, white testers were 15 percentage points more likely to obtain any form of concession—refund, store credit, or exchange—than their Black counterparts. These racial disparities in treatment persisted, and even widened, after the testers who were initially denied a refund then demanded to speak with management. When tested, these results proved to be robust to various specifications and restrictions. These findings suggest that there is systematic race discrimination in the enforcement of consumer contracts.

*Assistant Professor, UCLA School of Law; Associate Professor, Tel Aviv University School of Law. This research was generously supported by the Coase-Sandor Institute at the University of Chicago and the Institute for Quantitative Social Science at Harvard University. [Acknowledgements to be added.] I also wish to thank the participants at the Annual Law & Economics Association Conference at Columbia University, the Harvard Law & Economics Workshop, the NYU Law & Economics Workshop, the University of Chicago’s Junior Scholars Forum in Law and Social Science, the Berkely Consumer Scholars Conference in Boston University, the Conference of Empirical Legal Studies (CELS) in Toronto, the UCLA Junior Scholars Forum, the Tel Aviv University Law and Economics Workshop, and the Experimental Methods in Law Forum in Bonn. The methods used in this study have been approved by the University of Chicago Institutional Review Board prior to data collection (IRB Study Number: IRB 18–1529). The study was pre-registered in AsPredicted (see: “Post-Contract Discrimination in the Retail Market,” #16928, created 11/23/2018). Alona Zafirir, Laura Ash Smith, Nuri Albayrak, Raz Halfon, and Rhiannon Jones provided excellent research assistance. I also thank Henry Kim and Ben Nyblade for their statistical advice.

Contents

| | |
|---|-----------|
| Introduction | 3 |
| 1 Background and Literature Review | 4 |
| 1.1 Selective Enforcement of Form Contracts | 4 |
| 1.2 Discretion and Discrimination | 6 |
| 2 Sample and Methodology | 7 |
| 2.1 The Test-Case: Product Returns | 7 |
| 2.2 The Sample of Stores | 8 |
| 2.3 The Design of the Audits | 9 |
| 3 Results | 11 |
| 3.1 The Effect of Race on Immediate Return Outcomes | 12 |
| 3.1.1 Raw Outcomes | 12 |
| 3.1.2 Regression Analysis | 14 |
| 3.2 Robustness: Eventual Return Outcome as the Dependent Variable | 18 |
| 3.3 Racial Disparities within Gender Sub-Groups | 20 |
| 4 Discussion | 23 |
| 4.1 Summary and Implications | 23 |
| 4.2 Future Research | 24 |
| 4.3 Potential Explanations | 25 |
| 5 Conclusion | 27 |
| 6 Appendices | 28 |
| 6.1 Study Materials | 28 |
| 6.1.1 Instructions to Buyers | 28 |
| 6.1.2 Instructions to Testers | 29 |
| 6.2 Additional Robustness Checks | 31 |
| 6.2.1 Likelihood of Speaking with Management as the Dependent Variable | 31 |
| 6.2.2 Likelihood of Obtaining an Improved Outcome (Conditional on Speak- | |
| ing with Management) as the Dependent Variable | 33 |
| 6.2.3 Regressions of Return Outcomes on all Race-Gender Tester Sub-groups | 35 |

Introduction

To date, efforts to protect consumers from race discrimination have focused on curbing discrimination that occurs *ex ante*, before the consumer makes the transaction, whether in the form of differential pricing, heightened surveillance in stores, lower quality service, or complete refusals to transact. Considerably less attention, however, has been given to discrimination in how contracts are implemented and performed *ex post*, after a consumer completes a transaction.

This dearth of research may be attributed, at least in part, to the long-held assumption that the written text of “standard” form consumer contracts reflects the real terms of the transaction, and that terms in form contracts apply uniformly to all consumers entering the same transaction (for example, Rakoff, 1983, p. 1223; Becher and Unger-Aviram, 2009, p. 201). More recent scholarship has called this assumption into question. Some scholars have speculated that even contract terms that seem rigid and unconditional on paper may be relaxed in practice (see Part 1.1). Others have provided anecdotal evidence that sellers often grant their representatives discretion to exercise leniency in their interactions with consumers (see Part 1.1). Yet there is almost no empirical research regarding how salespeople exercise that discretion or whether their decisions reveal racial disparities.¹

This Article begins to fill this gap. Using an original field study of product returns to retail stores, it asks whether, when authorized to deviate from a retailer’s return policy, salespeople disproportionately favor white customers compared to similarly situated Black customers. Answering that question has important implications for anti-discrimination and consumer protection efforts and may help policymakers, scholars, and courts decide whether they need to revisit current laws and regulations to ensure that minority consumers are adequately protected.

To study this issue, the research team recruited and trained 19 testers—5 Black males, 4 Black females, 5 white males, and 5 white females.² All testers were roughly the same age and were similarly attired in an attempt to control for socioeconomic status. The testers were then sent to return 200 items of unworn clothing that had been purchased in advance (by other members of the research team) to fifty-nine retail stores in Chicago.³ All stores officially required receipts as a condition to return. Using a uniform script,⁴ the testers attempted to return the items, all of which were unused and in their original packaging and all of which were unaccompanied by receipts. The testers were instructed to ask to speak to a manager if they were not immediately offered a refund.

The study measured the success of the two racial groups using two main dependent variables: (i) whether the tester was able to obtain an immediate refund; and (ii) whether

¹This deficit in the literature has been largely attributed to the difficulty in obtaining data on sellers’ on-the-ground behavior (Gabbidon and Higgins, 2020, p. xii), but it may also be due to the traditional assumption that standardized contracts are applied uniformly to all the consumers.

²The audits were conducted by testers of both races and genders to verify whether the results were robust across gender groups. See Part 3.3

³Chicago was chosen because it is a mixed city with a relatively large Black population. According to the U.S. Census Bureau data, Chicago is almost 30 percent Black (U.S Census Bureau, 2021)(U.S. Census Bureau 2021). For an elaborate description of the stores’ selection criteria and details on the data collection process, see Part 2.2.

⁴For details about the script and the design of the audit study more generally, see Part 2.3.

the tester was able to obtain an immediate “concession,” defined as any deviation from the store’s formal policy (including a refund, a store credit, or an exchange), as compared to having their return request completely rejected.

The study showed that, while 9 percent of white testers were offered an immediate refund, only 2 percent of their Black counterparts were offered the same, revealing a seven percentage-point gap. In terms of obtaining an immediate concession (refund, store credit, or exchange), Black testers fared worse than white testers as well: while 74 percent of white testers received an immediate concession, only 59 percent of their Black counterparts were offered the same, revealing a fifteen percentage-point gap.

These results passed various robustness checks and held even when the analysis included the eventual return outcomes of the testers who asked to speak to management because their requests had been denied (see Part 3.2). These findings suggest that existing antidiscrimination statutes may not be sufficient and that more research is necessary to determine how to address this phenomenon.

This Article proceeds as follows. Part 1 situates this research in prior literature, including existing evidence that suggests that, in non-consumer transactions, decision-makers often exercise discretion along racial lines and that racial disparities in treatment typically increase as discretion increases. Part 2 describes the study, including the data collection methods and the measures taken to control for variables other than the tester’s race. Part 3 presents the results, revealing significant racial disparities in the enforcement of retailers’ return policies. Part 4 then discusses the implications of the findings and points to potential explanations and to areas where future research is warranted. Part 5 concludes.

1 Background and Literature Review

1.1 Selective Enforcement of Form Contracts

Sellers often use standard form contracts when transacting with consumers (Bakos, Marotta-Wurgler and Trossen, 2014; Wilkinson-Ryan, 2013). Standardization saves costs and facilitates transactions (Butler, Drahozal and Shepherd, 1998, p. 183), while also allowing sellers to keep salespeople in check by ensuring adherence to the terms of the form contract (Rakoff, 1983, p. 1223.). As legal sociologist Stewart Macaulay (1965, p. 1059) has observed, form contracts allow sellers to efficiently control salespeople by putting customers on notice of salespeople’s “limited authority.” As Macaulay further notes, firms typically use standardized agreements or policies “to avoid being legally bound to expectations” set by salespeople who might hold ideas “inconsistent with company policy” (Macaulay, 1965, p. 1059). As a result, scholars have traditionally assumed that sellers have little interest in allowing sales personnel to deviate from the text of their form agreements. For example, Shmuel Becher and Esther Unger-Aviram (2009) have suggested that “sellers are not likely to allow . . . deviations from pre-printed forms” and that “the typical seller does not empower its representatives (salespeople) to make changes in a standard form contract” (p. 201). Indeed, the widespread use of standard-form contracts has led experts to conclude that there is no longer any bargaining in consumer contracts (Johnston, 2005, p. 858).

Several scholars, however, have recently begun to question that common assumption,

suggesting that, in competitive markets, sellers’ enforcement of their contractual provisions may be more lenient than contractually required (Bebchuk and Posner, 2005; Becher and Zarsky, 2019; Gillete, 2004; Johnston, 2005; Katz, 1990, p. 281). As these commentators explain, sellers may include ostensibly rigid, non-negotiable terms in their contracts as a way to protect themselves from consumer opportunism, while still allowing their representatives to deviate from standard terms on a case-by-case basis.

Credit card issuers, for example, may attempt to discourage borrowers from regularly paying their bills late by including a “late fee” provision in their credit card agreements. But they may still decide to waive the fee if a certain borrower accidentally misses one payment (Bond, 2019). Similarly, airlines, whose “contracts of carriage” may require that passengers pay a fee if they miss their flight, may waive the fee for certain passengers who miss a flight due to unforeseen circumstances (Elliott, 2018). Or mortgage service providers, who have the contractual right to foreclose, may choose not to exercise their rights, depending, for example, on a borrower’s credit risk (Avery et al., 1996).

Some scholars have argued that the practice of complementing rigid written contract terms with authority to deviate from them *ex post*—a practice by sellers which this article dubs “selective enforcement of consumer contracts”—may be socially desirable (Bebchuk and Posner, 2005; Johnston, 2005). They have reasoned that, to the extent that consumer misbehavior is observable to the seller but non-verifiable (or verifiable at a high cost) to third parties (such as arbitrators or courts), sellers will benefit from having a protective provision in writing, but—at the same time—market competition will pressure sellers to behave fairly (Bebchuk and Posner, 2005; Gillete, 2004). And because standard form terms may be altered after the fact and “implemented in a balanced way,” some of these scholars have also maintained that courts should “take a hard line” in enforcing the terms of form contracts in the absence of evidence of fraud (Bebchuk and Posner, 2005, p. 823-830).

Other commentators have questioned the actual ability of competitive forces to deter sellers from adhering to the four corners of their agreements (Arbel, 2019, p. 1303; Becher and Zarsky, 2019, p. 82; Hoffman, 2016, p. 1641). Yonathan Arbel (2019), for example, has argued that reputation may not sufficiently constrain sellers from strictly observing the terms of their agreements. He has therefore called for “greater skepticism towards current trends to deregulate consumer transactions on the basis of faith in the internal regulatory power of market forces” (p. 1303).

While scholars continue to debate the normative implications of selective enforcement strategies in consumer markets, there has been almost no empirical investigation into sellers’ actual implementation of these strategies on the ground. This paucity in research is surprising since empirical data could shed light on the factual premises on which these normative debates hinge. Particularly missing is systematic research exploring whether consumer contracts are enforced in ways that disproportionately favor certain consumer groups, for example based on immutable characteristics such as gender or race.

While not specifically mentioning gender or race, scholars have speculated that, in theory, sellers might selectively enforce their contract terms disproportionately to the disadvantage of some consumer groups, in particular their “less valuable” or “less sophisticated” consumers. For example, Becher and Zarsky (2011, p. 91) have hypothesized that “uninformed and weak groups of consumers” will be disadvantaged, as “sophisticated and informed” groups will be treated more forgivingly or generously. Eyal Zamir (2014, p. 2100) has similarly speculated

that reputational forces “are much more likely to work in favor of large, recurring, and sophisticated customers—whose goodwill the supplier values highly—than in favor of the weak, occasional, and unsophisticated customer, whose goodwill is valued less.” But these theoretical accounts have not been empirically tested.

One notable exception is a study conducted by Manisha Padi (2021). Padi’s study explores the socio-economic implications of selective contract enforcement in the context of mortgage agreements. Using a detailed commercial dataset of mortgage performance from 2000 to 2008, Padi showed that 40 percent of borrowers who fell behind on their mortgage payments during this period avoided foreclosure due to the service providers’ exercise of discretion. Perhaps more important, the study also revealed significant disparities in treatment between borrowers residing in wealthy neighborhoods and those residing in poorer neighborhoods (Padi, 2021, p. 857-860). Specifically, the analysis revealed a significant positive correlation between forbearance decisions and the mean income in the borrower’s neighborhood: high-income neighborhoods had a 0.33 higher probability of avoiding foreclosure (p. 858).

Since Padi’s analysis was done by neighborhood, rather than at the individual level, she could not untangle income and race. It may well be, however, that disparities in treatment are driven, at least in part, by the different racial characteristics of borrowers’ neighborhoods (p. 862). Indeed, as the next section shows, the literature in domains outside consumer contracts suggests that discretionary judgments are often exercised discriminatorily across racial lines.

1.2 Discretion and Discrimination

There is ample evidence from a wide range of disciplines—from policing and adjudication to housing and employment—that suggests that discretion is often exercised in ways that disproportionately harm minorities, and racial minorities in particular (Bertrand and Duflo, 2017).

In the criminal justice system, for example, decision makers—including police officers, prosecutors, jurors, and judges—have been found to exercise their discretion discriminatorily along racial lines in variety of contexts, ranging from jury selection to sentencing and requests for the imposition of the death penalty (Bushway and Piehl, 2001; Paternoster, 1984; Price, 2009). In the employment context, studies have found that employers are significantly less likely to call back similarly situated minority-group job applicants, particularly Black and female candidates (Neumark, Bank and Van Nort, 1996). And in the housing context, researchers have found that homeowners are significantly more likely to refuse to rent apartments to minority tenants than to comparable tenants belonging to the majority (Carpusor and Loges, 2006; Ewens, Tomlin and Wang, 2014; Hanson and Hawley, 2011).

Moreover, there is a vast body of research that demonstrates that that increased discretion in decision-making widens disparities in treatment, and that “unstructured decision-making is exactly the sort of environment in which implicit biases can have their biggest impact” (Kleinberg et al., 2018, p.120).⁵ In the context of criminal sentencing, for example, Crystal

⁵The substantiated claim that broader discretion results in greater discrimination has also been raised and discussed in the context of the “rules versus standards” conundrum (see, for example, Covey, 2016;

Yang (2015) has studied the effect of judicial discretion on racial disparities in sentencing in the aftermath of the Supreme Court’s ruling in *United States v. Booker* (543 U.S. 220 [2005]), which removed mandatory sentencing guidelines and granted judges greater discretion. Using data on sentences imposed between 1994 and 2010, Yang found that the Black-white sentencing gap grew significantly post *Booker*.⁶

While considerably less is known about the relationship between discretion and discrimination in consumer markets, there is preliminary evidence that both consumers and sellers exercise their discretion discriminatorily against racial and gender minorities. For example, researchers have found that non-white taxi drivers receive, on average, significantly lower tips than do white drivers Ayres, Vars and Zakariya, 2004 and that female online sellers receive, on average, significantly lower price offers than do male online sellers offering the same product (Kricheli-Katz and Regev, 2016).

The research on the seller side is equally revealing. Researchers have found, for example, that sellers exercise their discretion discriminatorily against minority consumers in various ways, including by quoting significantly higher prices for the same products when facing Black or female customers (compared to similarly situated white or male customers) (Ayres, 1991; Ayres and Siegelman, 1995) or by refusing to transact with Black customers altogether (Edelman, Luca and Svirsky, 2017).

While very few field studies examine the discriminatory exercise of discretion in the context of retail shopping, one study has found that, when given the opportunity to deviate from a random-selection surveillance protocol aimed at preventing shoplifting, trained security guards at a department store in Atlanta oversampled non-white consumers (Dabney et al., 2006).

These findings suggest that, if given the authority to exercise discretion in the enforcement of standard store policies and contract terms, store clerks and managers might exercise their discretion in ways that discriminate against minority consumers. The field study described in this Article was designed to test that hypothesis.

2 Sample and Methodology

2.1 The Test-Case: Product Returns

This Article studies whether retail clerks’ exercise of discretion in the enforcement of their employers’ return policies disproportionately disadvantages Black consumers, compared to similarly situated white consumers.

The area of product returns is a fruitful ground for studying retail discrimination because store clerks and managers are typically granted a significant degree of discretion about whether to accept or deny returns (Bebchuk and Posner, 2005, p. 827; Becher and Zarsky, 2019, p. 91; Johnston, 2005, p. 873).⁷

Kaplow, 1992, p. 609; Sunstein, 1995, p. 974)

⁶However, some studies have yielded mixed or conflicting results. For example, Fischman and Schanzenbach (2012) have found that racial disparities were either reduced or little changed when the sentencing guidelines became advisory.

⁷In informal interviews conducted with store clerks in preparation for this study, several interviewees also mentioned being granted a considerable degree of discretion in deciding whether to accept a return

Product returns are also an important test case because consumers greatly value their ability to return unused products to retail stores (Zamir and Teichman, 2018, p. 290-291). According to recent data, more than fifteen percent of all purchases made in the United States are ultimately returned to the stores, adding up to hundreds of billions of dollars each year (Repko 2022).

2.2 The Sample of Stores

The sample initially consisted of all 192 retail stores with a Chicago location appearing in the ReferenceUSA and Hoover’s Company Directories’ databases. Dollar stores, high-end stores not offering items for 30 dollars or less, or stores that did not have a downtown Chicago location, were excluded from the sample to make the study more manageable. Stores whose formal return policy did not explicitly require a receipt for a refund were also excluded from the sample. The final sample included 59 retail stores located in downtown Chicago.⁸

Table 1 presents the sampled stores’ descriptive statistics. This data was obtained from Bloomberg and Hoover’s Company Directories. The sampled stores’ mean annual revenue (for the year of 2018) was 177 million dollars. Their median revenue, however, was only 3.1 million dollars, suggesting that the mean revenue was driven by very large companies. The average company age was 63 years, while the median age was 46.5 years. Publicly traded companies constituted 66 percent of the sample.

Although dollar stores and high-end stores not offering items for 30 dollars or less were excluded from the sample, the database included stores of various price ranges. To control for this variation in prices, store websites were scraped using Python, and the median prices of the clothing items offered at each website were computed.⁹ Median prices ranged between 5 dollars and 350 dollars, with a mean of 60 dollars and a median of 45 dollars.

Finally, while all stores formally required a receipt for a refund, some stores allowed for non-refund concessions in the form of store credit or exchange subject to the store representative’s discretion, while others did not mention non-refund concessions in their return policies. The stores that explicitly allowed for non-refund concessions subject to their employees’ discretion, which were classified as stores with a more lenient formal policy, constituted 61 percent of the sample. The store’s policy type (lenient or not) was subsequently controlled for in the regression analysis (see Part 3).

(interviews on file with the author).

⁸Due to discarded tests because of failure to adequately follow the script and scheduling difficulties, the final dataset includes 203 audits of the sampled fifty-nine stores (59 by white female testers, 51 by white male testers, 51 by Black male testers, and 42 by Black female testers). The analysis was conducted with and without the unmatched or only partially matched observations, and the results are not sensitive to their inclusion. I therefore report the results from the analysis of the entire dataset.

⁹Median prices were chosen instead of mean prices, as mean prices—unlike median prices—are affected by outliers (that is, extremely expensive or very cheap products). Some stores blocked access to their websites, so these websites were manually coded. If a store did not offer items for sale online, members of the research team were instructed to randomly choose 25 items at the store and compute their median price.

Table 1: Sampled Stores’ Descriptive Statistics

| | Mean | Median |
|-----------------------------|---------------|---------------|
| Annual Revenues | \$177 million | \$3.1 million |
| Age | 63 years | 46.5 years |
| Publicly Traded | 0.66 | |
| Median Price of Store Items | \$60 | \$45 |
| Lenient Formal Policy | 0.61 | |

Table 1 presents the sampled stores’ summary statistics. Annual revenues are in dollars. Age is defined as the years passed since the company’s incorporation. The median price of the store items refers to the median price of the items offered for sale at the store’s website. Lenient formal policy is defined as a policy that formally allows store clerks to offer consumers non-refund concessions (store credit or exchange) even when they fail to show a receipt.

2.3 The Design of the Audits

This study used an audit technique to test for discriminatory enforcement of retail return policies. While audit studies—in which testers are matched for all relevant observable characteristics other than the suspect one (for example, race)—are commonly used to test for discrimination in the housing, mortgage, and labor markets (see, for example Bertrand and Duflo, 2017, p. 310), they are rarely used to study discrimination in consumer markets.¹⁰

This study uses audits, rather than observational, regression-based methods (which are typically subject to omitted variable bias) or correspondence methods (which are typically only able to track the preliminary or *ex ante* stages of the contractual relationship), to test for disparities in *ex post* contractual performance based on race.

Before the audits began, members of the research team (the “buyers”) purchased more than 200 items from the stores participating in the study. To minimize differences across stores, the buyers were instructed to purchase a gender-neutral accessory (that is a hat, gloves, socks, scarves, purses, or bags). If no accessories were available, they were instructed to purchase a shirt, pair of pants, or another clothing item. They were specifically instructed to refrain from purchasing underwear, swimwear, jewelry, electronic devices, clearance or sale items, or any item that was specifically not eligible for returns according to each store’s formal return policy. To eliminate concerns that any observed disparities would be driven by the price of the returned item, purchased items’ prices were kept constant at between 20 to 30 dollars. Buyers paid in cash so that sellers would not obtain any personal information from the purchase.

The testers—five white females, five white males, four Black females,¹¹ and five Black

¹⁰For a notable exception, see Ayres and Siegelman (1995).

¹¹I initially hired five Black female testers. However, one tester dropped out during training after expressing safety concerns.

males—were randomly assigned to stores, and the order in which testers audited each store was also chosen randomly. The items were then returned in their original packaging and condition and with tags attached—but without a receipt—by different members of the research team (the “testers”) during the store’s specified return period window.

This design, in which the testers returned items that had been purchased in advance by the buyers, was chosen for several reasons. First, it allowed for the purchase of almost identical items from each store. Second, it mitigated the concern that store clerks might identify the person making the return.¹²

As Heckman and Siegelman (1993) observe, although audit studies are often referred to as “experiments” or “matched pair studies,” race or ethnicity are personal characteristics that cannot be assigned by randomization as in a classical experiment. Therefore, “adjustments must be made instead on ‘relevant’ observed characteristics to ‘align’ audit pair members” (p. 188).

Because accepting a returned clothing item, as opposed to renting an apartment or hiring an employee, is not a high-stakes transaction and does not involve a protracted negotiation or interchange, the task of matching testers is considerably less complicated in the context of this study. Still, to reinforce confidence that any differences in outcomes were not driven by unobserved differences between testers, measures were taken to reduce inter-tester variation.

To that end, testers were recruited according to uniform criteria. All were undergraduate students residing in Chicago, between 18–25 years of age. In addition, all testers were instructed to wear similar attire (casual clothing) and were told to audit the stores on weekday afternoons.

A significant challenge in audit studies of the one used here is to ensure that testers behave uniformly at the store (Bielen, Marneffe and Mocan, 2021, p. 269; Heckman and Siegelman, 1993; Heckman, 1998). To ensure uniformity, all testers were required to memorize an identical script and then follow it exactly in their interactions with store clerks and managers. Testers were provided with uniform answers about the reason for the return and for not having the receipt. They also received a list of contingent responses to additional questions they were likely to encounter. For example, if the store clerk asked why they wanted to return the item, testers were instructed to say that they realized they did not need it after purchasing it.¹³ Before auditing the stores, the testers attended training sessions at the University of Chicago, where they practiced their scripts and participated in numerous mock negotiations designed to help them behave uniformly during the audits.

¹²Still, this design raises the concern that in some stores, store clerks were more suspicious of the testers making the returns because they could not identify them. While this increased suspicion, to the extent that it existed, cannot explain the observed differential treatment based on race (or gender), it may mean that sellers are more likely to discriminate against minority consumers when they cannot recognize the customers seeking to make the return.

¹³It is possible that store clerks’ responses to testers’ return requests were sensitive to the reason offered by testers. For example, if testers had said that they had bought the wrong size or received the item as a gift and did not like it, store clerks may have responded differently. The generic excuse used in the study was meant to allow testers to request a refund rather than merely exchange the item or obtain store credit. Unlike returning a gift or exchanging an item for a different size, explicitly saying that they did not need the product made asking for a cash refund rather than an exchange or store credit arguably more credible and reasonable. In any event, the type of excuse chosen cannot, on its own, explain the observed racial disparities in treatment.

At each store, the testers were instructed to wait in line until a store clerk became available, inform the clerk that they wanted to return the item, and await the store clerk’s response. If the store clerk immediately agreed to provide a refund, testers were instructed to accept the refund, thank the clerk, and leave the store.

If, however, testers were told they could not return the item or were offered anything other than a refund (for example, an exchange or store credit), they were instructed to ask to speak to a manager. Whether the store clerk refused to call someone or whether the store clerk called a “manager,” testers were instructed to once again ask for a refund. Regardless of the clerk’s or manager’s response, testers would then offer thanks and leave the store. Upon leaving each store, testers filled out a detailed report describing the outcomes of each attempted return, including the (perceived) race and gender of the store representatives (to be used as controls in the regression analysis).

In most audit studies, auditors are typically informed of the goal of the study. However, as noted by Heckman (1998), Riach and Rich (2002), and Turner, Fix, and Struyk (1991), this might encourage auditors (consciously or unconsciously) to interact with a decision-maker in a manner that is meant to generate more evidence of discrimination. To alleviate these concerns, testers in this study were blind to the true purpose of the study. Rather, they were told that the study’s purpose was to explore differences in retailers’ return policies. In addition, testers were asked at the end of audits what they thought the study was about, and none of them indicated they believed the study tested for discrimination.

3 Results

This section reports the results of the audit study. The regression tables present models with various specifications, using ordinary least squares regressions.¹⁴ Subsection 3.1 reports the main results, using immediate return outcomes as the dependent variable. Subsection 3.2 uses additional dependent variables—the testers’ eventual outcomes—as a robustness check.¹⁵ Subsection 3.3 looks into the potential role of the tester’s gender, and at whether racial disparities persist within each gender sub-group.¹⁶

¹⁴Ordinary least squares regressions are reported in the main text for simplicity of understanding, but logistic regressions were also conducted, and the results are roughly the same. Jackknife and bootstrap resampling operations were also conducted to ensure robustness. Resampling is a way to reuse data to generate new, hypothetical samples (called resamples) that are representative of the underlying population. While both bootstrap and jackknife could be used as resampling methods, jackknife is a more conservative approach and is considered more suitable for small original data samples (see, for example, Efron Gong 1983). The jackknife is used to estimate the bias of an estimator by dropping one observation each time and recalculating the estimator based on the remaining dataset. I therefore report both jackknife and bootstrap standard errors in brackets. The results generally remain statistically significant when the standard errors are calculated using the jackknife and bootstrap operations

¹⁵Additional robustness checks, with regressions using other dependent variables, are reported in the appendix.

¹⁶Additional regression tables of the return outcomes on the four race-gender tester sub-groups are reported in the appendix.

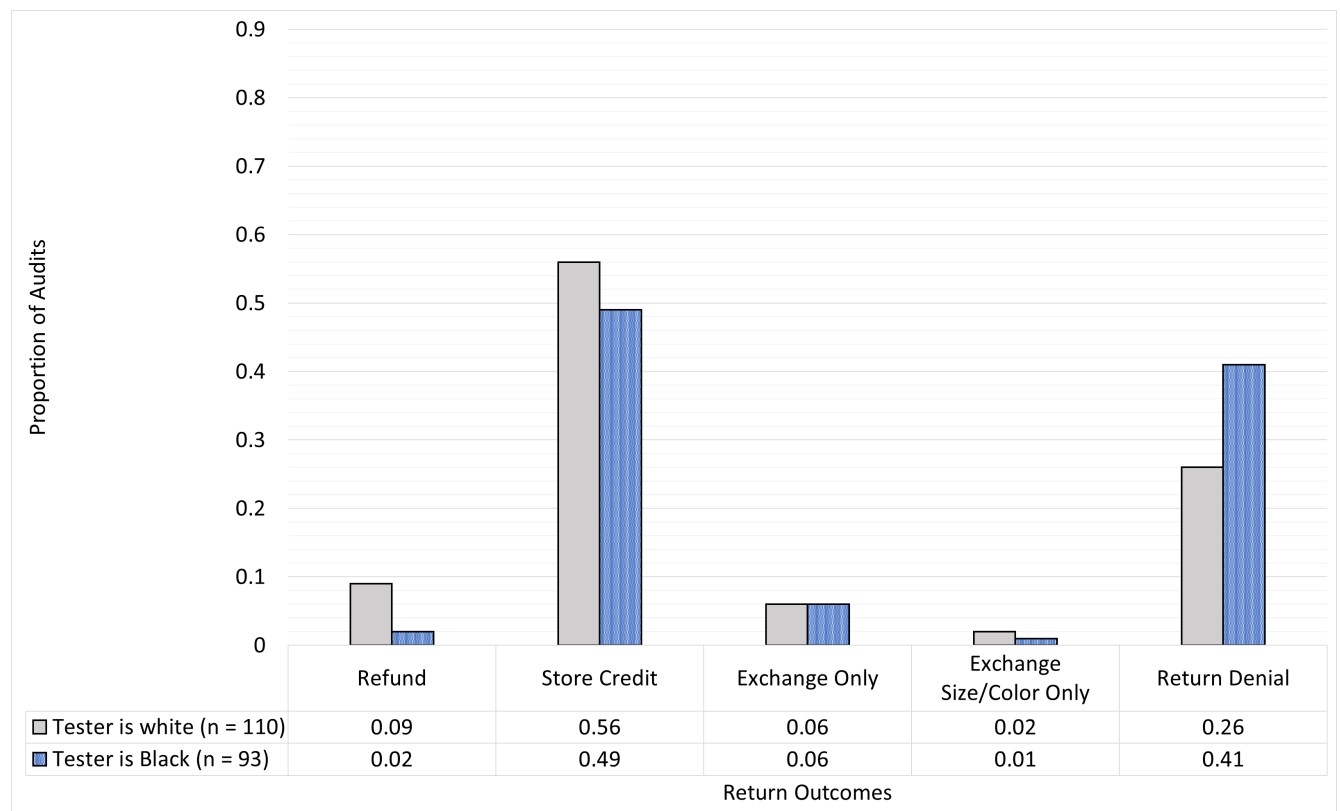
3.1 The Effect of Race on Immediate Return Outcomes

This subsection reports the main results, using the tester’s immediate return outcome (defined as the store clerk’s first response to the tester’s return request), as the main outcome measure. Because the store clerks’ immediate reactions were made following relatively little interaction with the testers, this measure provides a relatively well-controlled test for discriminatory behavior. In other words, it is unlikely that the racial disparities in the immediate treatment were driven by any uncontrolled differences between testers.

3.1.1 Raw Outcomes

The testers’ *immediate* return outcomes were classified as follows: (1) a *refund*—if the tester was offered a refund, in addition to store credit or exchange; (2) a *store credit or exchange*—if the tester was not offered a refund, but was offered an exchange or store credit; (3) an *exchange only*—if the tester was only allowed to exchange the item (but was not allowed a refund or store credit); (4) an *exchange size/color only*—if the tester was only allowed to exchange the item for a different size or color; or (5) a *return denial*—if the return request was denied completely, and the tester was not even able to exchange the item or obtain a store credit. Figure 1 shows the proportion of each response by race.

Figure 1. Immediate Return Outcome by Tester Race



Note. Sum may not add up to 100 percent due to rounding.

As Figure 1 illustrates, white testers were 7 percentage points more likely to be immediately offered a refund than were similarly situated Black testers. White testers were also 7 percentage points more likely to be immediately offered store credit compared to Black testers.

Overall, white testers were 15 percentage points more likely than Black testers to be immediately offered a “concession,” a term that includes a refund, a store credit, or an exchange, rather than to have their return request completely rejected.

3.1.2 Regression Analysis

In the regression analysis that follows, I consider two definitions of the dependent variable: (1) *Immediate refund*. A dummy variable which equals “1” if the tester was offered a refund immediately after making the initial return request, “0” otherwise; and (2) *Immediate concession*. A dummy variable which equals “1” if the tester was offered any concession, whether in the form of a refund, store credit, or exchange, after making the initial return request, and “0” if the return request was denied completely.

Table 2 presents the results of six regression models. Immediate refund offers are used in all six models as the dependent variable. Standard errors are clustered at the store and tester levels to account for correlations in the error terms (Bertrand, Duflo, and Mullainathan 2004).

Model 1 regresses the dependent variable on the race of the tester. Model 2 adds the race of the clerk (as perceived and reported by the tester) as a control. Model 3 adds an interaction term between the race of the clerk and the race of the tester. Model 4 adds store fixed effects. Model 5 adds time of day and day of week fixed effects, as well as month of audit fixed effects. Model 6 adds various controls related to the store characteristics, including an indicator for whether the store was a discount, mainstream or high-end store,¹⁷ store age (defined as years since incorporation), an indicator for whether the retail company was publicly or privately owned, and an indicator for whether the store formally allowed for non-refund concessions without receipts, subject to store clerks’ discretion.

¹⁷Stores were classified as either high-end, mainstream, or discount, based on the median prices at the stores. The top 25th percentile was classified as high-end, the lowest 25th percentile was coded as discount, and the remaining stores were coded as mainstream.

Table 2: The Effects of Tester Race on Immediate Return Outcomes (Refund as Dependent Variable)

| | (1) | (2) | (3) | (4) | (5) | (6) |
|----------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|-------------------------|
| White | 0.0694** (0.0314) | 0.0720** (0.0347) | 0.0565 (0.0647) | 0.0817** (0.0340) | 0.0790** (0.0345) | 0.0848** (0.0304) |
| | [0.0315] | [0.0350] | [0.0657] | [0.0433] | [0.0442] | [0.0377] |
| | {0.0345} | {0.0371} | {0.0647} | {0.0379} | {0.0359} | {0.0355} |
| Clerk is white | | 0.000424 (0.0464) | -0.0435 (0.0432) | 0.0626 (0.0584) | 0.0568 (0.0603) | 0.0665 (0.0433) |
| Clerk is other non-white | | -0.0442 (0.0409) | -0.00502 (0.0577) | -0.0432 (0.0648) | -0.0542 (0.0701) | -0.00664 (0.0429) |
| White tester | | | 0.0950 | | | |
| # clerk is white | | | (0.295) | | | |
| White tester | | | -0.0680 | | | |
| # clerk is other non-white | | | (0.0799) | | | |
| Store is mainstream | | | | | | |
| Store is high-end | | | | | | |
| Age of Store | | | | | | 0.0752 (0.0983) |
| Store is privately-owned | | | | | | 0.0580 (0.121) |
| | | | | | | -0.000696 (0.000953) |
| | | | | | | -0.115 (0.0542) |
| Lenient Policy | | | | | | 0.0954 (0.0803) |
| Constant | 0.0215 (0.156) | 0.0337 (0.363) | 0.0435 (0.315) | 0.00451 (0.928) | 0.0645 (0.549) | -0.0643 (0.697) |
| Observations | 203 | 203 | 203 | 203 | 203 | 200 |
| R^2 | 0.021 | 0.029 | 0.049 | 0.069 | 0.078 | 0.095 |

Note. Results are from OLS regressions of the tester's immediate return outcomes, with immediate refund (a dummy variable which equals "1" if the tester was offered a refund, "0" otherwise) as the dependent variable. Black is the reference category for the race of the tester and the clerk. Store fixed effects are included in models 4 and 5. Time of day, day of week, and month of audit fixed effects are included in models 5 and 6. Model 6 also includes store controls, including the store type (discount, mainstream, or high-end, with discount as the reference category) and age (number of years since incorporation), whether the store is public or private (with public as the reference category), and policy type (lenient or not). Standard errors are clustered at the store and tester levels and are presented in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Jackknife standard errors (203 replications) are in brackets. Bootstrap standard errors (50 replications) are in curly brackets.

As table 2 shows, across specifications, white testers were significantly more likely to be offered an immediate refund than their Black counterparts. The pro-white gap remained significant and similarly large in magnitude, at approximately 7 percentage points, across specifications, even after store characteristics and clerk demographics were controlled for in the regression analysis.

The interaction term between the race of the tester and the store clerk was not significant. I therefore cannot rule out the possibility that there was equal racial disparity in treatment across store clerks' racial groups.¹⁸

Next, I explored whether white testers were more likely than their Black counterparts to immediately receive any concession, a term that includes a refund, a store credit, or an exchange, as opposed to an absolute denial of their return request. Table 3 presents the results of the same six models, this time using a dependent variable which takes the value of "1" if any immediate concession was offered, and "0" if the return request was denied completely.

¹⁸The sample size unfortunately does not allow me to draw any meaningful conclusions about the relationship between the race of the tester and the store clerk. This important question is therefore left for future research.

Table 3: The Effects of Tester Race on Immediate Return Outcomes (Concession as Dependent Variable)

| | (1) | (2) | (3) | (4) | (5) | (6) |
|----------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| White | 0.145** (0.030) | 0.156** (0.023) | 0.347*** (0.005) | 0.148** (0.038) | 0.170** (0.020) | 0.182** (0.010) |
| Clerk is white | [0.0665] {0.0611} | [0.0686] {0.0697} | [0.1248] {0.1115} | [0.0833] {0.0737} | [0.0848] {0.0927} | [0.0869] {0.0842} |
| Clerk is other non-white | | 0.0351 (0.656) | 0.181 (0.160) | 0.0682 (0.550) | 0.0671 (0.537) | 0.0824 (0.411) |
| White tester | | -0.0878 (0.298) | 0.0987 (0.495) | -0.205* (0.077) | -0.165 (0.147) | -0.0772 (0.437) |
| # clerk is white | | | -0.248 (0.124) | | | |
| White tester | | | -0.302* (0.088) | | | -0.0638 (0.779) |
| # clerk is other non-white | | | | | | 0.0394 (0.888) |
| Store is mainstream | | | | | | -0.00320 (0.148) |
| Store is high-end | | | | | | -0.0892 (0.477) |
| Age of Store | | | | | | 0.0714 (0.701) |
| Store is privately-owned | | | | | | |
| Lenient Policy | | | | | | |
| Constant | 0.591 (0.000) | 0.599 (0.000) | 0.478 (0.000) | 0.627 (0.000) | 0.849 (0.000) | 0.921 (0.017) |
| Observations | 203 | 203 | 203 | 203 | 203 | 200 |
| R^2 | 0.024 | 0.036 | 0.053 | 0.078 | 0.117 | 0.109 |

Note. Results are from OLS regressions of the tester's immediate return outcomes, with immediate concession (a dummy variable which equals "1" if the tester was offered refund, store credit, or exchange, "0" otherwise) as the dependent variable. Black is the reference category for the race of the tester and the clerk. Store fixed effects are included in models 4 and 5. Time of day, day of week, and month of audit fixed effects are included in models 5 and 6. Model 6 also includes store controls, including the store type (discount, mainstream, or high-end, with discount as the reference category) and age (number of years since incorporation), whether the store is public or private (with public as the reference category), and policy type (lenient or not). Standard errors are clustered at the store and tester levels and are presented in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Jackknife standard errors (203 replications) are in brackets. Bootstrap standard errors (50 replications) are in curly brackets.

As table 3 shows, white testers were significantly more likely to be offered an immediate concession (as opposed to complete rejection of their return requests) than their Black counterparts. The pro-white gap remained significant, at between 14 and 18 percentage points, across specifications. Put differently, Black testers were between 14 and 18 percentage points more likely than their white counterparts to have their immediate return request denied completely.

Notably, as column 3 in table 3 shows, the race of the clerk significantly interacted with the race of the tester. Specifically, white testers were significantly more likely to be offered a concession when interacting with Black clerks than with other non-white (for example, Hispanic or Asian) clerks.

3.2 Robustness: Eventual Return Outcome as the Dependent Variable

As a robustness check, I also tested whether racial disparities persisted even if the eventual results of those who were initially denied a refund (and then asked to speak to management) were included in the analysis. Admittedly, the immediate return outcomes represent a “cleaner” test of discrimination, as they required considerably less interaction between the testers and the clerks. However, including the post-complaint results allows us to test whether complaining or negotiating moderates the observed racial disparities in treatment. After all, one could hypothesize that, even if we observed racial disparities in treatment after the initial return request, these disparities should decrease (or even disappear) once the tester lodges a complaint and again requests a refund. To test this hypothesis, I also looked at the testers’ final outcomes.

Table 4 presents the results of this robustness check. In models 1 to 3, the dependent variable equals “1” if refund was offered, either immediately or after the tester continued to negotiate. It equals “0” if testers’ refund requests were eventually denied. Model 1 regresses the dependent variable on tester’s race, while model 2 includes clerk controls and store fixed effects, and model 3 adds time of audit fixed effects. In models 4 to 6, the dependent variable equals “1” if any concession was offered, either immediately or after the tester continued to negotiate. It equals “0” if the tester’s return request was eventually denied.

Table 4: The Effect of Tester Race on Final Return Outcomes

| | DV: Refund | | | DV: Concession | | |
|--------------------------|----------------------|----------------------|---------------------|----------------------|----------------------|----------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| White | 0.0735* | 0.0808** (0.0378) | 0.0687* (0.0386) | 0.231*** (0.0609) | 0.230*** (0.0702) | 0.243*** (0.0631) |
| | [0.0397] | [0.0506] | [0.0519] | [0.0642] | [0.0702] | [0.0737] |
| | {0.0372} | {0.0410} | {0.0492} | {0.5556} | {0.0642} | {0.0581} |
| Clerk is white | | 0.0248 (0.0662) | 0.00310 (0.0674) | | 0.323 (0.0939) | 0.0166 (0.0928) |
| Clerk is other non-white | | -0.0385 (0.0730) | -0.0775 (0.0789) | | -0.164* (0.0827) | -0.152* (0.0876) |
| Constant | 0.0538** (0.0235) | 0.0524 (0.0536) | 0.323** (0.142) | 0.624*** (0.0505) | 0.663*** (0.0727) | 1.048*** (0.215) |
| Observations | 203 | 203 | 203 | 203 | 203 | 203 |
| R^2 | 0.016 | 0.031 | 0.090 | 0.070 | 0.126 | 0.151 |

Note. Results are from OLS regressions of the tester's final return outcomes. Models 1, 2, and 3 use eventual refund (a dummy variable which equals "1" if the tester was eventually offered a refund, "0" otherwise) as the dependent variable. Models 4, 5 and 6 use eventual concession (a dummy variable which equals "1" if the tester was eventually offered refund, store credit, or exchange, "0" otherwise) as the dependent variable. Black is the reference category for the race of the tester and the clerk. Models 2, 3, 5, and 6 include store fixed effects. Models 3 and 6 introduce time of day, day of week, and month of audit fixed effects. Standard errors are clustered at the store and tester levels and are presented in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Jackknife standard errors (203 replications) are in brackets. Bootstrap standard errors (50 replications) are in curly brackets.

As table 4 shows, the results are robust to using the tester’s final return outcome as the dependent variable. Indeed, white testers were between 7 and 8 percentage points more likely to receive a refund than their Black counterparts and white testers were between 23 and 24 percentage points more likely to receive a concession (refund or non-refund)—either immediately or eventually—than were Black testers, and these racial differences were statistically significant across all specifications.

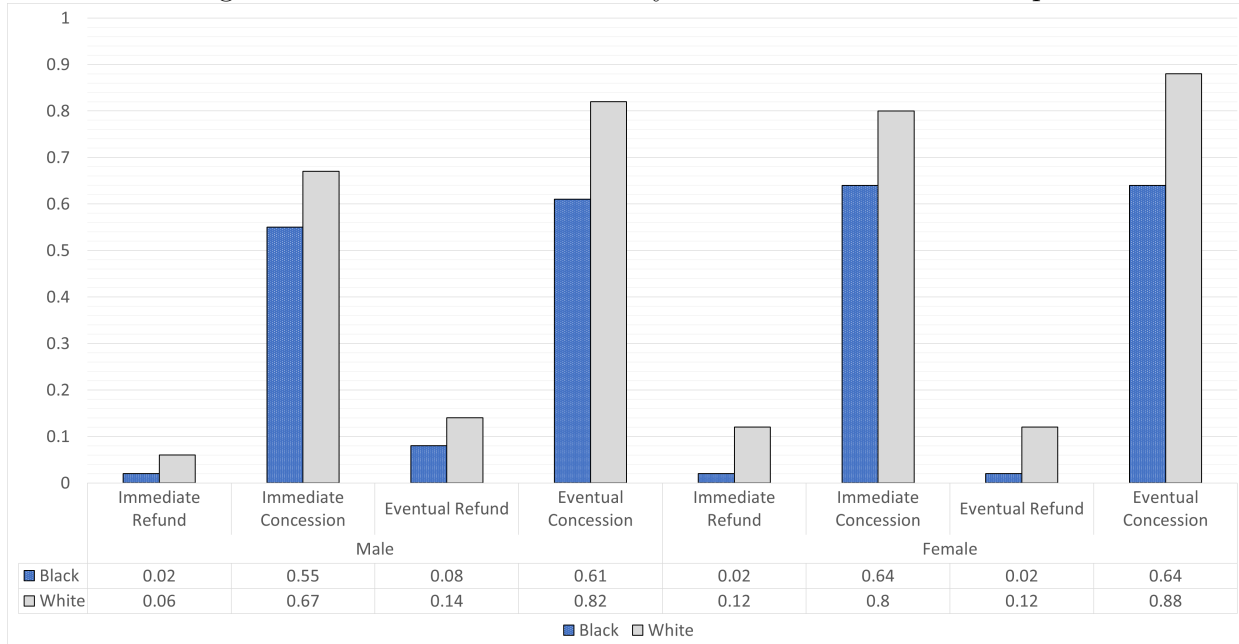
As shown in tables A1 and A2 (in the appendix), across specifications, white testers were between 22 and 23 percentage points more likely to speak with management upon request than were Black testers who made the same request (these differences are significant at the 1 percent level). Not only were white testers significantly more likely to have their request to speak with a manager met, but white testers were also between 15 and 20 percentage points more likely than Black testers to obtain an improved outcome—defined as any additional concession beyond what was offered when the tester made the initial return request. These differences, both in likelihood of speaking with management and in likelihood of obtaining an improved outcome conditional on seeing a manager, can help explain the racial disparities in eventual refund and concession offers reported in table 4.

One cautionary note: although the testers were blind to the goal and hypotheses of the study, a few of the Black testers indicated in their post-audit reports that they believed that their race had negatively affected the way they were treated at the store. While these testers may have witnessed discriminatory behavior on the part of the salesclerks, it is also possible that they went into the audits expecting to encounter mistreatment or discrimination and that their postures or beliefs contributed to the ultimate result. The short time frame during which the results were measured weighs against the possibility that it was the testers’ behavior that was driving the observed disparities. Nevertheless, readers should note this issue and researchers may want to consider it in the design of future studies.

3.3 Racial Disparities within Gender Sub-Groups

In addition to the robustness checks described above, I also tested whether the racial disparities persisted within each gender group to examine whether the results were driven or moderated by gender. Figure 2 shows the racial disparities in return outcomes, both immediate and eventual, within each gender group.

Figure 2. Raw Return Outcome by Race within Gender Groups



Note. Sum may not add up to 100 percent due to rounding.

As figure 2 reveals, within each gender sub-group, both male and female, Black testers did worse than their white counterparts, using all four dependent variables (both immediate and eventual refund and concessions).

Table 5 presents these results. All models regress the dependent variable on the tester's race and on store and clerk control variables and audit fixed effects for each of the gender sub-groups. Model 1 uses an immediate refund as the dependent variable, model 2 uses an immediate concession as the dependent variable, model 3 uses an eventual refund as the dependent variable, and model 4 uses an eventual concession as the dependent variable. In each model, the left column represents the female sub-group and the right column—the male sub-group.

Table 5: The Effects of Tester Race on Return Outcomes by Gender Sub-Groups

| Gender Group | (1) | | (2) | | (3) | | (4) | |
|---|-------------------------|------------------------|-----------------------|-------------------------|-------------------------|--------------------------|-----------------------|------------------------|
| | All Female | All Male | All Female | All Male | All Female | All Male | All Female | All Male |
| White | 0.103* (0.0603) | 0.00233 (0.0370) | 0.243** (0.0971) | 0.147 (0.105) | 0.0865 (0.0591) | 0.0287 (0.0628) | 0.329*** (0.0951) | 0.261*** (0.0978) |
| Clerk is white | 0.0173 (0.0741) | -0.0857* (0.0458) | 0.0148 (0.119) | 0.0355 (0.130) | 0.00419 (0.0726) | -0.0839 (0.0777) | -0.0172 (0.117) | 0.0299 (0.121) |
| Clerk is other non-white | -0.0466 (0.0741) | -0.0930** (0.0455) | -0.0467 (0.119) | -0.0758 (0.129) | 0.0126 (0.0727) | -0.110 (0.0772) | -0.101 (0.117) | -0.0189 (0.120) |
| Clerk is male | 0.0361 (0.0591) | -0.0154 (0.0380) | 0.180* (0.0952) | 0.0531 (0.108) | -0.00894 (0.0580) | 0.0692 (0.0646) | 0.0645 (0.0933) | -0.0974 (0.101) |
| Clerk is other (gender) | 0.145 (0.115) | | 0.331* (0.186) | | 0.0880 (0.113) | | 0.280 (0.182) | |
| Store is Mainstream | 0.0583 (0.0741) | 0.0423 (0.0471) | -0.0124 (0.119) | 0.0560 (0.134) | 0.0478 (0.0727) | -0.00641 (0.0800) | -0.00540 (0.117) | -0.156 (0.125) |
| Store is High-end | 0.119 (0.0829) | 0.0482 (0.0557) | 0.158 (0.133) | 0.265* (0.158) | 0.189** (0.0812) | -0.0350 (0.0946) | 0.0140 (0.131) | -0.0781 (0.147) |
| Age of Store | 0.00174** (0.000676) | 0.000320 (0.000461) | -0.00118 (0.00109) | -0.00271** (0.00131) | 0.00157** (0.000663) | -0.0000461 (0.000783) | -0.00101 (0.00107) | -0.000275 (0.00122) |
| Store is privately-owned | -0.0318 (0.0628) | -0.0138 (0.0395) | -0.0417 (0.101) | -0.165 (0.112) | -0.0366 (0.0616) | 0.0523 (0.0670) | -0.0647 (0.0991) | -0.103 (0.104) |
| Formal Policy allows non-refund concessions | -0.0140 (0.0590) | -0.0133 (0.0397) | 0.349** (0.0950) | 0.0545 (0.113) | 0.0475 (0.0578) | -0.117* (0.0674) | 0.147 (0.0930) | -0.0280 (0.105) |
| Constant | 0.0529 (0.307) | 0.100 (0.124) | 0.377 (0.494) | 1.097*** (0.352) | 0.0254 (0.301) | 0.388* (0.211) | 0.599 (0.484) | 1.229*** (0.328) |
| Observations | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| R ² | 0.226 | 0.122 | 0.250 | 0.143 | 0.255 | 0.182 | 0.174 | 0.133 |

Standard errors are clustered at the store and tester levels and are presented in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All models include time of day, day of week, and month of audit fixed effects.

As table 5 shows, among female testers, white females were significantly more likely to receive an immediate refund or concession, as well as an eventual concession, compared to Black females. Among male testers, white males were more likely to receive an immediate or eventual concession, as well as an eventual refund. However, except for the context of eventual concessions, in which the racial disparities were significant at the 1 percent level, these differences were not significant.

Tables A3 and A4 (in the Appendix) show regression models of the dependent variables on the four race-gender tester sub-groups, with white female testers as the reference category. These tables show that Black testers—both female and male—were significantly less likely to receive an immediate refund, an immediate concession, or an eventual concession than white female testers and that the greatest disparities in treatment (in all outcome measures other than eventual refund) were between white female testers and Black male testers. This suggests that the intersection of race and gender plays a role in determining return outcomes.

Indeed, the findings of this study could be read as showing that Black consumers—both male and female—are treated significantly less favorably than white testers, particularly white females. Additionally, the findings reveal that the intersection of race and gender in this context places Black male consumers in an especially disadvantaged position, while white female customers enjoy an especially large “white female” premium.¹⁹

4 Discussion

4.1 Summary and Implications

This audit study reveals that, at least in the context of retail product returns, stores selectively enforce seemingly uniform contracts and policies, deviating from the terms of their form contracts on a case-by-case basis. These findings are consistent with the small but growing body of research that suggests that sellers often complement their allegedly standardized agreements with internal policies granting their representatives discretion in applying these agreements on the ground (Bebchuk and Posner, 2005; Gillete, 2004; Johnston, 2005).

Most important, however, the study reveals that such discretionary performance of contractual terms can be exercised in a discriminatory manner. Specifically, the findings revealed that—in the setting of retail product returns—there was significant racial discrimination in the enforcement of return policies to the disadvantage of Black consumers. The study showed that, in stores with formal policies requiring receipts as a precondition to returns, white consumers (testers) without receipts were significantly more likely to be offered a refund or any concession—whether in the form of a refund, store credit, or exchange—than similarly situated Black consumers (testers). These racial disparities were robust across specifications and restrictions and persisted even when the analysis included the eventual outcome of those who were initially denied a refund and requested to speak with management. In sum, Black testers—male and female—obtained worse outcomes—compared to white male and female

¹⁹These findings are consistent with prior research suggesting that Blacks, and particularly Black men, are often penalized for displaying assertiveness in social and professional settings. As previous work reveals that in various professional domains, Blacks, and particularly Black men, often suffer a “penalty” for taking actions that appear assertive or angry, as opposed to comparable whites.

testers.

This study suggests that racial discrimination in the enforcement of consumer contracts—retailer return policies being just one example—may be an overlooked form of systemic marketplace discrimination. And yet, as research on marketplace discrimination suggests, discriminatory enforcement of consumer contracts may be particularly harmful to minority consumers. In the specific context of retail race discrimination, for example, it has been shown that Black consumers’ encounters with unfair treatment—their “Shopping while Black” experiences²⁰—have cumulative debilitating effects on their self-esteem and life satisfaction, and even on their mental health, (Columbres, 2002; Feagin, 1991; Gabbidon and Higgins, 2020; Paradies et al., 2015; Schmitt et al., 2014).²¹

Discriminatory enforcement of consumer contracts could also result in regressive redistribution if poor consumers are treated less favorably than wealthier consumers (either directly or indirectly—as a result of the relationship between wealth and other characteristics) (see Padi 2022). Furthermore, consumers who belong to discriminated-against groups might be discouraged from entering into transactions, even if those transactions are conducive to their well-being; alternatively, they might enter into transactions optimistically believing that sellers will behave more leniently than their contracts dictate, only to face strict (and discriminatory) observance of the terms *ex post*.

Discriminatory enforcement of consumer contracts may also be counterproductive for sellers in terms of sales and profits. In the context of retail clothing, for example, Black consumers spend almost as much as white consumers on goods and services (U.S. Department of Labor, 2020). According to a nationally representative consumption data, Black consumers devote larger shares of their expenditure bundles to visible goods such as clothing than do white consumers with comparable incomes (Charles, Hurst and Roussanov, 2009, p. 425-67). Treating Black customers as second-class citizens may ultimately translate into lower sales and lost profits (Chui et al., 2021).²²

4.2 Future Research

To test for discrimination in the enforcement of consumer contracts, the study focused on discrimination against *Black* consumers, using one test case—*retail product returns*—and was conducted in one city—*Chicago*.

While the results reveal striking racial disparities in the enforcement of retail return policies, future research should examine the generalizability of these findings. In particular,

²⁰Discrimination while shopping (or consumer racial profiling in stores) is most commonly associated with Black consumers, perhaps due to media coverage and highly publicized litigation cases involving Black consumers. Yet other racial and ethnic minorities (including Hispanics, Asians, and Native Americans) report similar experiences of discrimination, also known as “shopping while brown or non-white.” See, for example, Dunlap 2021.

²¹For a general assessment of the social costs produced by marketplace race discrimination, see Siegelman 1998, p. 12.

²²As sociologist Cassi Pittman Claytor notes, “[e]very year Black Americans spend an estimated 1.2 trillion dollars. This combined spending power exceeds several nations’ gross domestic product . . . [yet] for decades and decades, Black consumers have been regularly overlooked by companies that do not see them as a priority demographic.” (Pittman, 2020, p. 7) Discriminating against Black consumers could easily prove to be a multibillion-dollar mistake.

future studies should explore whether discriminatory enforcement of consumer contracts exists against other minority groups, in other types of contracts, and in other markets.

For example, studies could investigate whether and to what extent other racial and ethnic minorities, including Hispanics, Asians, and Native Americans, are similarly subject to discriminatory enforcement of consumer contracts. Researchers could also test whether contracts are enforced differently within the Black population based on different skin tones. Although colorism and racism are often intertwined, people belonging to the same race might suffer from different degrees (or forms) of discrimination based on their skin tones.²³

Finally, because this study only included bricks-and-mortar sales, future research is needed to determine whether online contracts reflect similar racial disparities in enforcement. This is important because, although e-commerce retail sales currently account for only 15 percent of all retail sales worldwide (U.S Department of Commerce, 2022), online shopping is increasingly gaining popularity among Black and white consumers alike (Wang 2022).²⁴

4.3 Potential Explanations

The findings reveal significant racial disparities in salespeople’s treatment of consumers’ return requests. The exact cause of the observed discrimination remains open to debate. We do not know, for example, what assumptions or beliefs on the part of the salesclerks led to the observed behavior. Existing research, however, points to at least three possible explanations. First, there is evidence that store clerks tend to be more suspicious of Black customers generally, particularly Black men, and that they unfairly target them with surveillance and calls to the police (Gabbidon and Higgins, 2011, p. 208; Rappaport, 2018, p. 2290; Schreer, Smith and Thomas, 2009, p. 1432-1444). This general suspicion of Black customers has been documented in the lab. In a 2000 study, survey participants—all undergraduate marketing students in Minnesota—were asked to imagine that they were the “managers of tomorrow’s retail establishments.” When asked about the “typical shoplifter,” most participants described a young Black male, even though law enforcement statistics in the area showed that the typical shoplifter was a white female (Asquith and Bristow, 2000, p. 272).

Second, the observed racial discrimination might stem from differences in store clerks’ inferences about a particular consumer’s socioeconomic status, buying power, and value to the store. As researchers have shown, “Blacks have been viewed as second-class citizens in consumer markets” both historically and in contemporary times (Pittman, 2020, p. 4). In fact, research suggests that people often associate race with class, (Fiske et al., 2002; Penner

²³Indeed, <https://www.overleaf.com/project/634ee9205438e80d5503fb23>irical research has shown that, at least in the employment context, darker skin tones have a negative impact on employment possibilities and outcomes (see, for example, Kricheli Katz et al., 2020).

²⁴While it is plausible that the effects of race (and gender) will be less pronounced in online settings, several studies have already shown that discrimination persists in the online marketplace. Indeed, some researchers have even suggested that minority consumers may find it helpful to conceal their racial identities by, for example, using home addresses of family relatives who live in more racially diverse neighborhoods, or by using white-sounding names. In one oft-cited case, an online job applicant’s decision to change his name from “Jose” to “Joe” significantly increased the positive responses to his job application. Future research on selective enforcement of contracts in online settings could shed light on these questions (Matthews, 2014).

and Saperstein, 2013) and that class is closely intertwined with what it means to be “Black” or “white” (Brewster, 2012; Brewster, Brauer and Lynn, 2015; Freeman and Ambady, 2011; Ridgeway and Kricheli-Katz, 2013). Thus, if clerks typically believe that white customers are likely to be wealthier than Black customers and, therefore, more valuable customers with greater buying power, they might well treat Black customers less favorably than white customers.

If clerks hold those beliefs, the fact that the study was conducted in downtown Chicago, a predominantly white area, may have exacerbated the situation. In other words, it is plausible that Black testers were treated with more suspicion in that area because they do not frequently shop there or because they were perceived to be one-time customers and thus less valuable than repeat buyers. Indeed, previous interview-based research suggests that Black consumers face a greater degree of discrimination when shopping in predominantly white areas (Lee, 1998). In any event, the expectations and beliefs associated with the socio-economic status of the testers could explain the observed racial discrimination.

Notably, the study described in this Article attempted to control for class and socio-economic backgrounds by instructing all testers to wear similar casual clothing. It must be acknowledged, however, that the same clothing (in this case, jeans and a t-shirt) may convey different meanings when worn by members of different racial or ethnic groups. To the extent that people conflate race with class, such inferences may have played a role in the store clerks’ and managers’ decisions. Moreover, it must be noted that, in real life, outside of the context of a controlled field study, these inferences may, in fact, be bolstered or attenuated by other status signals, such as a consumer’s brand-name or expensive attire. And this might mean that selective enforcement of contracts might also have regressive distributional outcomes, with poor consumers less likely than affluent consumers to enjoy favorable concessions or deviations from the language of the agreement.

Third, the observed racial disparities might be driven by differences in the perceived likelihood that Black consumers will complain or generate reputational harm if dissatisfied. If clerks assume that minority consumers are less likely to complain or generate reputational harm to the store than their white counterparts, then salesclerks might be more willing to ignore Black customers’ requests.

In fact, there is evidence to suggest that store clerks believe that white customers harbor higher expectations of service than Black customers of the same socioeconomic background and that those white customers are more likely to display assertiveness when these expectations are not met (for example, Williams, 2006, p. 111). This underlying belief in the difference in expectations—specifically the white customer’s sense of entitlement—could prompt clerks to treat white customers more favorably than their Black counterparts.

In this study, all testers were instructed to ask to speak with a manager if their refund request was initially denied. In the real world, however, consumers may be less likely to complain if they believe that their chances of obtaining redress are low (Granbois, Summers and Frazier, 1977, p. 18; Richins, 1982; Richins, 1983, p. 76). This, in turn, can create a vicious cycle: if Black consumers do not feel entitled to complain or do not anticipate a successful outcome after complaining, they may be less likely to register a complaint, thereby becoming even less likely to receive redress. At the same time, white consumers, who may not suffer from the same sense of impotence, will continue to complain, disproportionately benefitting from preferential treatment from the seller.

This insight also carries the possibility that differences in expectations or bargaining styles associated with race or gender could have influenced the results of this study. The fact that, during their post-audit interviews, a few Black testers mentioned that they believed that they had been discriminated against (see Part 3.2.2) raises the possibility that those testers’ perceptions influenced the results. In other words, if those Black testers expected to be treated with suspicion at the store—an artifact of their “shopping while Black” experience—they might have behaved differently when interacting and negotiating with store clerks and managers, either by signaling acquiescence or by infusing even the initial conversation with defensive posture or subtle hostility. And this, too, can lead to a vicious cycle of discrimination.

The potential magnitude of the so-called “expectations problem” should not be ignored. Research indicates that most Black consumers feel they are often discriminated against in stores (Williams and Snuggs, 1997; Williamson, 2000). According to every Gallup poll (Jones and Lloyd, 2021) since 1997, Black consumers have been most likely to report unfair treatment while shopping. Indeed, in the most recent Gallup poll (July 2021), as many as 35 percent of the Black respondents reported having recently experienced unfair treatment while shopping in a store—more than in any other situation, including interactions with the police, at the workplace, or in healthcare or entertainment facilities.

Thus, while it is plausible that the racial disparities seen in the study were driven by the salesclerks’ belief systems, it is possible that the results were also driven by the Black testers’ expectations, which in turn affected the salesclerks’ behavior. Of course, these potential explanations are not mutually exclusive and may coexist. Moreover, regardless of the potential explanations underlying salespeople’s observed behavior, the findings of this study have important implications since they reveal significant and robust racial disparities in the enforcement of consumer contracts.

5 Conclusion

This Article provides empirical evidence that significant racial disparities exist in the enforcement of certain consumer contracts and that such discrimination is pervasive and robust. Indeed, as this experiment shows, sellers’ discretionary performance of return policies disproportionately harms Black consumers.

In this study, Black testers, posing as consumers, were significantly less likely than their white counterparts to obtain refunds or concessions beyond the dictates of sellers’ formal return policies. These racial disparities not only persisted across gender groups but widened after the testers whose refund requests were initially denied requested to speak with management.

These findings illustrate the importance of addressing a gaping hole in existing anti-discrimination laws, which do not seem to currently prohibit this form of marketplace discrimination (Columbres, 2002, p. 215; Kennedy, 2001, p. 307; Singer, 1995; Williams et al., 2015, p. 86). Future research should focus on exploring whether and to what extent discriminatory enforcement of consumer contracts exists in other markets and contracts, which types of consumers face this form of discrimination, and the underlying causes of such discrimination.

If this study’s findings are corroborated by future research and if discriminatory enforcement of consumer contracts is shown to be widespread, policymakers will need to consider adopting appropriate legislation to reduce the social and economic harms that this form of ongoing discrimination generates.

6 Appendices

6.1 Study Materials

6.1.1 Instructions to Buyers

Retail Return Policies Study: Instructions

Thank you for agreeing to participate in the study. This is a study about retail stores’ return practices. You will be asked to go and buy items from stores located in downtown Chicago, while recording various details about the stores and the items purchased. You will be provided with a list of stores and their addresses. You will be provided with money so that you could pay in cash for the items.

In each store, please buy four identical items that cost between 20 and 30 dollars. If there are no items within that price range in the store, please indicate so in the survey form and leave the store. Please try to buy a clothing accessory if possible (i.e., a hat, gloves, socks, a scarf, a purse, or a bag). Otherwise, please buy a shirt, pants or another clothing item. Please do not buy underwear, swimwear, jewelry, electronic devices, “last act”/clearance/sale items, or any item that you suspect will not be eligible for returns (or will have stricter return requirements). Please try to buy gender-neutral items.

When making the purchase, please pay in cash and ask for separate bags and receipts for each item. Please keep the receipts—you will be required to bring the original receipts to the researcher.

If the store clerk offers you to join a credit card/membership/sign up for something or asks for any personal details or contact information, please politely refuse and do not provide any such information. [. . .] While at the store, please see if there is a return policy sign anywhere. Please answer the related questions on the survey form [. . .].

Form (Buyers):

Name of store:

Date:

Time of visit:

Item purchased:

Item price:

If the store does not offer items within the requested price range, please check here:

Please do not buy any item and leave the store.

6.1.2 Instructions to Testers

Thank you for agreeing to participate in the study. This is a study about retail stores' return practices. You will be asked to go and make non-receipted returns to stores located in Chicago, while following a memorized script, and to report the outcomes of your attempted returns as explained in detail below.

Preparation (before going to stores): Please memorize the script and follow it to the letter. We will practice the script during our simulation sessions.

You will be provided with a list of stores and their addresses. All stores are located in downtown Chicago. You will also be provided with items that you will need to return to the stores. Each item will be in its original package with tags attached. You will not have the receipt.

Please come to the stores dressed in casual clothes: jeans and t-shirt.

At the store: When you enter the store, please approach the first available store clerk at the service/check-out counter. If no clerk is available, please wait in line until a store clerk becomes available. If more than one store clerk is available, please choose the store clerk standing to the right. Please report the perceived race and gender of the store clerk in the form. Please follow the script from that point onward.

Immediately after leaving each store, please fill out the below form:

Form (Auditors):

Auditor's Name:

Store audit number (how many stores have you audited before this one):

Name of store:

Date of visit:

Time of visit:

Item Type:

Item Price:

[. . .]

For the store clerk you spoke with:

1. What do you think their race was? Please check one:

1. Black
2. White
3. Asian American
4. Hispanic
5. Native American
6. Mixed
7. Other

2. What do you think their gender was? Please check one:

1. Female

2. Male
3. Other

At first, what did the store clerk offer you? Please check one:

1. Refund
2. Store Credit
3. Exchange
4. Exchange for a different size/color only
5. Nothing offered
6. Other (please explain)

When you asked to see a manager, what did the store clerk say/do?

1. Said that he/she was the manager
2. Said that the manager is not available
3. Refused to call manager, saying he/she already knew what the manager was going to say
4. Called the manager
5. Went and spoke to manager
6. Other (please explain)

Did you see a manager?

1. Yes
2. No

If so, What do you think their race was? Please check one:

1. Black
2. White
3. Asian American
4. Hispanic
5. Native American
6. Mixed
7. Other (please explain)

What do you think their gender was? Please check one:

1. Female
2. Male
3. Other (please explain)

[for those who saw manager] When you talked to the manager, what did they offer you?
Please check one:

1. Refund
2. Store Credit
3. Exchange

4. Exchange for a different size/color only
5. Nothing offered

Do you have any other comments or issues that arose from your visit? If so, please explain:

What do you think was the purpose of the study?

6.2 Additional Robustness Checks

6.2.1 Likelihood of Speaking with Management as the Dependent Variable

Table A1 reports the results of six models using ordinary least squares regressions. The dependent variable is whether the tester saw a manager upon request (a dummy variable which equals “1” if the tester saw a manager after requesting to speak with one, “0” otherwise). Model 1 regresses the dependent variable on the race of the tester. Model 2 adds the race of the clerk (as perceived and reported by the tester) as a control. Model 3 adds an interaction term between the race of the clerk and the race of the tester. Model 4 adds store fixed effects. Model 5 adds time of day and day of week fixed effects and month of audit fixed effects. Model 6 adds various controls related to the store characteristics, including an indicator for whether the store is a discount, mainstream or high-end store,²⁵ store age (defined as years since incorporation), an indicator for whether the retail company is publicly or privately owned, and an indicator for whether the store formally allows for non-refund concessions, subject to store clerks’ discretion.

²⁵The sample included stores of various price ranges. To control for this variation in prices, store websites were scraped using Python and the median prices of the clothing items offered at each website were computed. Some stores blocked access to their websites or did not offer items for sale online. These websites ($n = 17$) were manually coded. Median prices were chosen instead of mean prices, as mean prices—unlike median prices—are affected by outliers (that is, extremely expensive or very cheap products). Median prices ranged between \$5 and \$350, with a mean of \$60 and a median of \$45. Stores were then classified as either high-end, mainstream, or discount, based on the median priced at the stores. The top 25th percentile was classified as high-end, the lowest 25th percentile was coded as discount, and the remaining stores were coded as mainstream.

Table A1: The Effects of Tester Race on Likelihood of Speaking with Management Upon Request

| | (1) | (2) | (3) | (4) | (5) | (6) |
|--|----------------------|----------------------|---------------------|----------------------|---------------------|----------------------|
| White | 0.231** (0.0649) | 0.230** (0.0648) | 0.268 (0.132) | 0.229** (0.0740) | 0.220** (0.0736) | 0.221** (0.0698) |
| Clerk is white | | 0.0993 (0.0830) | -0.114 (0.132) | 0.131 (0.0937) | 0.0786 (0.0907) | 0.0861 (0.101) |
| Clerk is other non-white | | 0.234*** (0.0769) | 0.295** (0.140) | 0.240** (0.106) | 0.162 (0.117) | 0.210** (0.101) |
| White tester # Clerk is white | | | -0.0144 (0.169) | | | |
| White tester # Clerk is other non-white | | | -0.101 (0.166) | | | |
| Store is mainstream | | | | | | -0.218 (0.233) |
| Store is high-end | | | | | | -0.0590 (0.291) |
| Age of Store | | | | | | 0.00232 (0.00214) |
| Store is privately-owned | | | | | | 0.0302 (0.123) |
| Formal Policy allows non-refund concessions | | | | | | -0.217 (0.181) |
| Constant | 0.589*** (0.0521) | 0.478*** (0.0761) | 0.455*** (0.108) | 0.461*** (0.0742) | 1.205*** (0.207) | 1.232*** (0.375) |
| Observations | 190 | 190 | 190 | 190 | 190 | 188 |
| R^2 | 0.065 | 0.107 | 0.109 | 0.109 | 0.192 | 0.223 |

Standard errors are in parentheses.

$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ Store fixed effects are included in models 4 and 5. Time of day, day of week, and month of audit fixed effects are included in models 5 and 6. Standard errors are clustered at the store and tester levels.

As table A1 shows, white testers were 22-23 percentage points more likely to see a manager upon request than comparable Black testers across specifications. This may have contributed to the observed racial disparities in eventual return outcomes.

6.2.2 Likelihood of Obtaining an Improved Outcome (Conditional on Speaking with Management) as the Dependent Variable

As table A1 reveals, white testers were significantly more likely to speak with a manager after asking to speak with one. But were they also more likely to obtain better treatment compared to their immediate return outcomes than Black testers?

Table A2 presents the results of six models using ordinary least squares regressions in an attempt to address that question. The dependent variable is whether the tester obtained an improved outcome (compared to what the tester was offered after making the initial request) after speaking with a manager. Model 1 regresses the dependent variable on the race of the tester. Model 2 adds the race of the clerk (as perceived and reported by the tester) as a control. Model 3 adds an interaction term between the race of the clerk and the race of the tester. Model 4 adds store fixed effects. Model 5 adds time of day and day of week fixed effects, and month of audit fixed effects. Model 6 adds various controls related to the store characteristics, including an indicator for whether the store is a discount, mainstream or high-end store, store age (defined as years since incorporation), an indicator for whether the retail company is publicly or privately owned, and an indicator for whether the store formally allows for non-refund concessions, subject to store clerks' discretion.

Table A2: The Effects of Tester Race on Likelihood of Obtaining an Improved Outcome upon Speaking with Management

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|----------------------|--------------------|--------------------|------------------------|---------------------|----------------------|
| White | 0.149* (0.0753) | 0.154* (0.0785) | 0.0692 (0.157) | 0.192** (0.0825) | 0.157* (0.0827) | 0.194** (0.0923) |
| Clerk is white | | 0.0510 (0.0993) | -0.0400 (0.150) | -0.0000480 (0.0922) | -0.0252 (0.0936) | 0.0316 (0.126) |
| Clerk is other non-white | | 0.0753 (0.0985) | 0.0222 (0.164) | 0.0440 (0.124) | -0.0226 (0.134) | -0.0190 (0.132) |
| White tester # clerk is white | | | 0.146 (0.201) | | | |
| White tester # clerk is other non-white | | | 0.0722 (0.205) | | | |
| Store Fixed Effects | No | No | No | Yes | Yes | No |
| Time of day, day of week Fixed Effects | No | No | No | No | Yes | Yes |
| Month fixed effects | No | No | No | No | Yes | Yes |
| Store is Mainstream | | | | | | -0.195 (0.263) |
| Store is High-end | | | | | | -0.205 (0.348) |
| Age of Store | | | | | | 0.00399 (0.00251) |
| Store is privately-owned | | | | | | -0.201 (0.151) |
| Formal Policy allows non-refund concessions | | | | | | 0.147 (0.239) |
| Constant | 0.189*** (0.0541) | 0.139 (0.0909) | 0.200 (0.129) | 0.146 (0.0952) | 0.418** (0.156) | 0.256 (0.433) |
| Observations | 136 | 136 | 136 | 136 | 136 | 136 |
| \bar{R}^2 | 0.026 | 0.031 | 0.034 | 0.057 | 0.108 | 0.183 |

Standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Store fixed effects are included in models 4 and 5. Time of day, day of week, and month of audit fixed effects are included in models 5 and 6. Standard errors are clustered at the store and tester levels.

6.2.3 Regressions of Return Outcomes on all Race-Gender Tester Sub-groups

Regression tables A3 and A4 present the results of ordinary least squares regressions of testers' return outcomes on the four race-gender tester sub-groups.

Table A3 uses refund as the dependent variable. Models 1 to 3 use immediate refund as the dependent variable. Models 4 to 6 use eventual refund as the dependent variable. Models 1 and 4 regress the dependent variable on the tester's race-gender sub-group, with white female testers as the reference category. Models 2 and 5 add store clerk controls and store and audit fixed effects. Models 3 and 6 include store control variables and parent company fixed effects.

Table A4 uses concession as the dependent variable. Models 1 to 3 use immediate concession as the dependent variable. Models 4 to 6 use eventual concession as the dependent variable. Models 1 and 4 regress the dependent variable on the tester's race-gender sub-group, with white female testers as the reference category. Models 2 and 5 add store clerk controls and store and audit fixed effects. Models 3 and 6 include store control variables and parent company fixed effects.

Table A3: Regressions of Return Outcomes on the four race-gender tester sub-groups: Refund as Dependent Variable

| | (1) | (2) | (3) | (4) | (5) | (6) |
|--------------------------|-----------------------|----------------------|-------------------------|----------------------|-----------------------|-----------------------|
| White Male | -0.0598 (0.0540) | -0.0735 (0.0465) | -0.0755* (0.0449) | 0.0186 (0.0646) | -0.00805 (0.0599) | -0.0175 (0.0573) |
| Black Female | -0.0948* (0.0487) | -0.0877* (0.0449) | -0.112** (0.0508) | -0.0948* (0.0487) | -0.0724 (0.0436) | -0.0886* (0.0505) |
| Black Male | -0.0990** (0.0468) | -0.129** (0.0492) | -0.129** (0.0589) | -0.0402 (0.0570) | -0.0803 (0.0676) | -0.0765 (0.0732) |
| Clerk is white | | 0.0316 (0.0546) | 0.0462 (0.0461) | | 0.00107 (0.0497) | 0.00466 (0.0583) |
| Clerk is other non-white | | -0.0769 (0.0678) | -0.0320 (0.0549) | | -0.0289 (0.0712) | -0.0224 (0.0684) |
| Clerk is male | | 0.0593 (0.0411) | 0.0465 (0.0388) | | 0.0665 (0.0559) | 0.0500 (0.0554) |
| Clerk is other (gender) | | 0.0452 (0.0955) | 0.106 (0.109) | | 0.124 (0.111) | 0.113 (0.116) |
| Mainstream | | | 0.0653 (0.0932) | | | 0.0414 (0.0920) |
| High-end | | | 0.00556 (0.142) | | | 0.0457 (0.151) |
| Age | | | -0.000675 (0.000864) | | | 0.000125 (0.00135) |
| Store is privately-owned | | | -0.0956 (0.0575) | | | -0.0706 (0.0986) |
| Lenient Policy | | | 0.101 (0.0897) | | | -0.0191 (0.147) |
| Constant | 0.119*** (0.0425) | 0.268** (0.109) | 0.109 (0.155) | 0.119*** (0.0425) | 0.328 *** (0.0944) | 0.279 (0.188) |
| Observations | 203 | 203 | 203 | 203 | 203 | 203 |
| R ² | 0.030 | 0.118 | 0.135 | 0.020 | 0.085 | 0.075 |

Standard errors are in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Store fixed effects are included in models 4 and 5. Time of day, day of week, and month of audit fixed effects are included in models 5 and 6. Standard errors are clustered at the store and tester levels.

Table A4: The Effects of Tester Race and Gender on Return Outcomes: Concession as Dependent Variable

| | (1) | (2) | (3) | (4) | (5) | (6) |
|--------------------------|-----------------------|-----------------------|------------------------|-----------------------|-----------------------|------------------------|
| White Male | -0.130 (0.0851) | -0.127 (0.0836) | -0.125 (0.0817) | -0.0578 (0.0687) | -0.0275 (0.0680) | -0.0235 (0.0686) |
| Black Female | -0.154* (0.0915) | -0.191* (0.101) | -0.178* (0.104) | -0.238*** (0.0859) | -0.253*** (0.0871) | -0.246*** (0.0888) |
| Black Male | -0.248*** (0.0881) | -0.289*** (0.0918) | -0.291*** (0.0984) | -0.274*** (0.0811) | -0.307*** (0.0812) | -0.321*** (0.0861) |
| Clerk is white | | 0.0550 (0.0882) | 0.0439 (0.0921) | | 0.0235 (0.0849) | 0.0454 (0.0823) |
| Clerk is other non-white | | -0.128 (0.0873) | -0.127 (0.0961) | | -0.135* (0.0788) | -0.135* (0.0757) |
| Clerk is male | | 0.0805 (0.0804) | 0.0995 (0.0837) | | -0.0533 (0.0742) | -0.0344 (0.0706) |
| Clerk is other (gender) | | 0.229 (0.199) | 0.188 (0.189) | | 0.220 (0.180) | 0.190 (0.163) |
| Mainstream | | | -0.0931 (0.0932) | | | -0.299*** (0.0920) |
| High-end | | | 0.0346 (0.173) | | | -0.316* (0.160) |
| Age | | | 0.00322** (0.00142) | | | -0.00513* (0.00273) |
| Store is privately-owned | | | -0.0469 (0.161) | | | -0.0787 (0.157) |
| Lenient Policy | | | 0.0855 (0.214) | | | 0.347* (0.194) |
| Constant | 0.797*** (0.0529) | 1.040*** (0.243) | 1.305*** (0.329) | 0.881*** (0.0425) | 1.234 *** (0.231) | 1.699*** (0.284) |
| Observations | 203 | 203 | 200 | 203 | 203 | 200 |
| R ² | 0.038 | 0.126 | 0.142 | 0.073 | 0.138 | 0.217 |

Standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Models 1 to 3 use immediate concession as the dependent variable. Models 4 to 6 use eventual concession as the dependent variable. Models 1 and 4 regress the dependent variable on the tester's race-gender sub-group, with white female testers as the reference category. Models 2 and 5 add store clerk controls and store and audit fixed effects. Models 3 and 6 include store control variables and parent company fixed effects.

Tables A3 and A4 show that Black testers—both female and male—were significantly less likely to receive an immediate refund, an immediate concession, or an eventual concession than white female testers and that the greatest disparities in treatment (in all outcome measures other than eventual refund) were between white female testers and Black male testers. This suggests that the intersection of race and gender plays a role in determining return outcomes.

References

- Arbel, Yonathan A. 2019. Reputation Failure: The Limits of Market Discipline in Consumer Markets. *Wake Forest Law Review* 54: 1239–1303.
- Asquith, Jo Ann L, and Dennis N Bristow. 2000. To Catch a Thief: A Pedagogical Study of Retail Shoplifting. *Journal of Education for Business* 75: 271–276.
- Avery, Robert B, Raphael W Bostic, Paul S Calem, and Glenn B Canner. 1996. Credit Risk, Credit Scoring, and the Performance of Home Mortgages. *Federal Reserve Bulletin* 82: 621–648.
- Ayres, Ian. 1991. Fair Driving: Gender and Race Discrimination in Retail Car Negotiations. *Harvard Law Review* 817–872.
- Ayres, Ian, and Peter Siegelman. 1995. Race and Gender Discrimination in Bargaining For a New Car. *The American Economic Review* 304–321.
- Ayres, Ian, Fredrick E Vars, and Nasser Zakariya. 2004. To Insure Prejudice: Racial Disparities in Taxicab Tipping. *Yale Law Journal* 114: 1613–74.
- Bakos, Yannis, Florencia Marotta-Wurgler, and David R Trossen. 2014. Does Anyone Read the Fine Print? Consumer Attention to Standard-Form Contracts. *The Journal of Legal Studies* 43: 1–35.
- Bebchuk, Lucian A, and Richard A Posner. 2005. One-Sided Contracts In Competitive Consumer Markets. *Michigan Law Review* 104: 827–835.
- Becher, Shmuel I, and Esther Unger-Aviram. 2009. The Law of Standard Form Contracts: Misguided Intuitions and Suggestions For Reconstruction. *DePaul Business and Commercial Law Journal* 8: 199–228.
- Becher, Shmuel I, and Tal Z Zarsky. 2019. Minding the Gap. *Connecticut Law Review* 51: 69.
- Becher, Shmuel I, and Tal Zarsky. 2011. Open Doors, Trap Doors, and the Law. *Law and Contemporary Problems* 63–90.
- Bertrand, Marianne, and Esther Duflo. 2017. Field Experiments on Discrimination. *Handbook of Economic Field Experiments* 1: 309–393.
- Bielen, Samantha, Wim Marneffe, and Naci Mocan. 2021. Racial Bias and In-Group Bias in Virtual Reality Courtrooms. *The Journal of Law and Economics* 64: 269–300.
- Bond, Casey. 2019. Credit Card Late Fees: What You Need to Know. <https://money.usnews.com/credit-cards/articles/credit-card-late-fees-what-you-need-to-know6002/>.
- Brewster, Zachary W. 2012. Racialized Customer Service in Restaurants: A Quantitative Assessment of the Statistical Discrimination Explanatory Framework. *Sociological inquiry* 82: 3–28.

- Brewster, Zachary W, Jonathan R Brauer, and Michael Lynn. 2015. Economic Motivations and Moral Controls Regulating Discrimination Against Black and Hispanic Diners. *The Sociological Quarterly* 56: 506–538.
- Bushway, Shawn D, and Anne Morrison Piehl. 2001. Judging Judicial Discretion: Legal Factors and Racial Discrimination in Sentencing. *Law and Society Review* 733–764.
- Butler, Henry N, Christopher R Drahozal, and Joanna Mehlhop Shepherd. 1998. *Economic Analysis for Lawyers*. Carolina Academic Press Durham.
- Carpusor, Adrian G, and William E Loges. 2006. Rental Discrimination and Ethnicity in Names 1. *Journal of Applied Social Psychology* 36: 934–952.
- Charles, Kerwin Kofi, Erik Hurst, and Nikolai Roussanov. 2009. Conspicuous Consumption and Race. *The Quarterly Journal of Economics* 124: 425–467.
- Chui, Michael, Brian Gregg, Sajaland Kohli, and Shelley Stewart. 2021. A \$300 Billion Opportunity: Serving the Emerging Black American Consumer. <https://www.mckinsey.com/featured-insights/diversity-and-inclusion/a-300-billion-dollar-opportunity-serving-the-emerging-black-american-consumer>.
- Columbres, Claudine. 2002. Targeting Retail Discrimination with Parens Patriae. *Columbia Journal of Law and Social Problems* 36: 209–236.
- Covey, Russell D. 2016. Rules, Standards, Sentencing, and the Nature of Law. *California Law Review* 104: 447–498.
- Dabney, Dean A, Laura Dugan, Volkan Topalli, and Richard C Hollinger. 2006. The Impact of Implicit Stereotyping on Offender Profiling: Unexpected Results From an Observational Study of Shoplifting. *Criminal Justice and Behavior* 33: 646–674.
- Dunlap, Michelle. 2021. *Retail Racism: Shopping While Black and Brown in America*. Rowman & Littlefield.
- Edelman, Benjamin, Michael Luca, and Dan Svirsky. 2017. Racial Discrimination in The Sharing economy: Evidence from a field experiment. *American Economic Journal: Applied Economics* 9: 1–22.
- Elliott, Christopher. 2018. Travel Nightmares: What to Do If You Miss Your Flight. <https://www.usatoday.com/story/travel/advice/2018/03/25/missed-flight-passenger-rights/450366002/>.
- Ewens, Michael, Bryan Tomlin, and Liang Choon Wang. 2014. Statistical Discrimination or Prejudice? A Large Sample Field Experiment. *Review of Economics and Statistics* 96: 119–134.
- Feagin, Joe R. 1991. The Continuing Significance of Race: Antiblack Discrimination in Public Places. *American Sociological Review* 56: 101–116.

- Fischman, Joshua B, and Max M Schanzenbach. 2012. Racial Disparities Under the Federal Sentencing Guidelines: The Role of Judicial Discretion and Mandatory Minimums. *Journal of Empirical Legal Studies* 9: 729–764.
- Fiske, Susan T., Amy J. C., Cuddy, Peter Glick, and Jun Xu. 2002. A Model of (often mixed) Stereotype Content: Competence and Warmth Respectively Follow From Perceived Status and Competition. *Journal of Personality and Social Psychology* 82: 878–902.
- Freeman, Jonathan B, and Nalini Ambady. 2011. A Dynamic Interactive Theory of Person Construal. *Psychological Review* 118: 247–279.
- Gabbidon, Shaun L, and George E Higgins. 2011. Public Opinion on the Use of Consumer Racial Profiling to Identify Shoplifters: An Exploratory Study. *Criminal Justice Review* 36: 201–212.
- Gabbidon, Shaun L, and George E Higgins. 2020. *Shopping While Black: Consumer Racial Profiling in America*. Routledge.
- Gillete, Clayton P. 2004. Rolling Contracts as an Agency Problem. in *Consumer Protection in the Age of the 'Information Economy*. edited by Jane K. Winn. New York University Press.
- Granbois, Donald, John O Summers, and Gary L Frazier. 1977. Correlates of Consumer Expectation and Complaining Behavior. *Consumer Satisfaction, Dissatisfaction and Complaining Behavior* 8: 18–25.
- Hanson, Andrew, and Zackary Hawley. 2011. Do Landlords Discriminate in the Rental Housing Market? Evidence from an Internet Field Experiment in US Cities. *Journal of Urban Economics* 70: 99–114.
- Heckman, James J. 1998. Detecting Discrimination. *Journal of Economic Perspectives* 12: 101–116.
- Heckman, James J., and Peter Siegelman. 1993. “The Urban Institute Audit Studies: Their Methods and Findings.”
- Hoffman, David A. 2016. From Promise to Form: How Contracting Online Changes Consumers. *New York University Law Review* 91: 1595–1645.
- Johnston, Jason Scott. 2005. The Return of Bargain: An Economic Theory of How Standard-Form Contracts Enable Cooperative Negotiation Between Businesses and Consumers. *Michigan Law Review* 104: 857–897.
- Jones, Jeffrey M., and Camille Lloyd. 2021. Black Americans’ Reports of Mistreatment Steady or Higher. <https://news.gallup.com/poll/352580/black-americans-reports-mistreatment-steady-higher.aspx>.
- Kaplow, Louis. 1992. Rules Versus Standards: An Economic Analysis. *Duke Law Journal* 42: 557–629.

- Katz, Avery. 1990. The Strategic Structure of Offer and Acceptance: Game Theory and the Law of Contract Formation. *Michigan Law Review* 89: 215–293.
- Kennedy, Deseriee A. 2001. Consumer Discrimination: The Limitations of Federal Civil Rights Protection. *Missouri Law Review* 66: 275–339.
- Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Cass R Sunstein. 2018. Discrimination in the Age of Algorithms. *Journal of Legal Analysis* 10: 113–174.
- Kricheli-Katz, Tamar, and Tali Regev. 2016. How Many Cents on The Dollar? Women and Men in Product Markets. *Science Advances* 2: e1500599.
- Kricheli Katz, Tamar, Tali Regev, Shay Lavie, Haggai Porat, and Ronen Avraham. 2020. Those who Tan and Those who Don't: A Natural Experiment on Colorism. *PloS One* 15: e0235438.
- Lee, In Ho. 1998. Market Crashes and Informational Avalanches. *Review of Economic Studies* 65: 741–759.
- Macaulay, Stewart. 1965. Private Legislation and the Duty to Read—Business Run by IBM Machine, the Law of Contracts and Credit Cards. *Vanderbilt Law Review* 19: 1051–1121.
- Matthews, Cate. 2014. He Dropped One Letter in his Name while Applying for Jobs, and the Responses Rolled in. *The Huffington Post* 2.
- Neumark, David, Roy J Bank, and Kyle D Van Nort. 1996. Sex Discrimination in Restaurant Hiring: An Audit Study. *The Quarterly Journal of Economics* 111: 915–941.
- Padi, Manisha. 2021. Contractual Inequality. *Michigan Law Review, Forthcoming*.
- Paradies, Yin, Jehonathan Ben, Nida Denson, Amanuel Elias, Naomi Priest, Alex Pieterse, Arpana Gupta, Margaret Kelaher, and Gilbert Gee. 2015. Racism as a Determinant of Health: a Systematic Review and Meta-Analysis. *PloS One* 10: e0138511.
- Paternoster, Raymond. 1984. Prosecutorial Discretion in Requesting the Death Penalty: A Case of Victim-Based Racial Discrimination. *Law and Society Review* 18: 437–478.
- Penner, Andrew M, and Aliya Saperstein. 2013. Engendering Racial Perceptions: An Intersectional Analysis of how Social Status Shapes Race. *Gender and Society* 27: 319–344.
- Pittman, Cassi. 2020. “Shopping While Black”: Black Consumers’ Management of Racial Stigma and Racial Profiling in Retail Settings. *Journal of Consumer Culture* 20: 3–22.
- Price, Melynda J. 2009. Performing Discretion or Performing Discrimination: Race, Ritual, and Peremptory Challenges in Capital Jury Selection. *Michigan Law Review* 15: 57–107.
- Rakoff, Todd D. 1983. Contracts of Adhesion: An Essay in Reconstruction. *Harvard Law Review* 1174–11285.
- Rappaport, John. 2018. Criminal Justice, Inc. *Columbia Law Review* 118: 2251–2322.

- Riach, Peter A, and Judith Rich. 2002. Field Experiments of Discrimination in the Market Place. *The Economic Journal* 112: F480–F518.
- Richins, Marsha L. 1982. An Investigation of Consumers' Attitudes Toward Complaining. *ACR North American Advances*.
- Richins, Marsha L. 1983. Negative Word-of-Mouth by Dissatisfied Consumers: A Pilot Study. *Journal of Marketing* 47: 68–78.
- Ridgeway, Cecilia L, and Tamar Kricheli-Katz. 2013. Intersecting Cultural Beliefs in Social Relations: Gender, Race, and Class Binds and Freedoms. *Gender and Society* 27: 294–318.
- Schmitt, Michael T, Nyla R Branscombe, Tom Postmes, and Amber Garcia. 2014. The Consequences of Perceived Discrimination for Psychological Well-Being: a Meta-Analytic Review. *Psychological Bulletin* 140: 921–948.
- Schreer, George E, Sandra Smith, and Kirsten Thomas. 2009. “Shopping while black”: Examining racial discrimination in a retail setting 1. *Journal of Applied Social Psychology* 39: 1432–1444.
- Siegelman, Peter. 1998. Racial Discrimination in “Everyday” Eommercial Transactions: What Do We Know, What Do We Need to Know, and How Can We Find Out. *A National Report Card on Discrimination in America: The Role of Testing* 69–98.
- Singer, Joseph William. 1995. No Right to Exclude: Public Accommodations and Private Property. *Northwestern Law Review* 90: 1283–1498.
- Sunstein, Cass R. 1995. Problems with Rules. *California Law Review* 83: 953–1027.
- Turner, Margery Austin, Michael Fix, and Raymond J Struyk. 1991. *Opportunities Denied, Opportunities Diminished: Racial Discrimination in Hiring*. The Urban Insitute.
- U.S Census Bureau, QuickFacts. 2021. Population Estimates, July 1 2021 (V2021) – Chicago city, Illinois. <https://www.census.gov/quickfacts/chicagocityillinois>.
- U.S Department of Commerce, Census Bureau. 2022. Quarterly Retail E-Commerce Sales 1Q 2022. <https://www.statista.com/statistics/187443/quarterly-e-commerce-sales-in-the-us/>.
- U.S. Department of Labor, Bureau of Labor Statistics. 2020. Consumer Expenditures in 2020.Report No. 1096. www.bls.gov/opub/reports/consumer-expenditures/2020/home.htm.
- Wilkinson-Ryan, Tess. 2013. A Psychological Account of Consent of Fine Print. *Iowa Law Review* 99: 1745–1784.
- Williams, Christine L. 2006. *Inside Toyland: Working, Shopping, and Social Inequality*. Univ of California Press.

- Williams, Jerome D, and Thelma Snuggs. 1997. "Survey of Attitudes Toward Customer Ethnocentrism and Shopping in Retail Store: The Role of Race." 161–62, Postdam, NY: Society for Consumer Psychology.
- Williams, Jerome, Geraldine Henderson, Sophia Evett, and Anne-Marie Hakstian. 2015. Racial Discrimination in Retail Settings: A Liberation Psychology Perspective. *Race and Retail: Consumption Across the Color Line* 263–276.
- Williamson, Oliver E. 2000. The New Institutional Economics: Taking Stocks, Looking Ahead. *Journal of Economic Literature* 38, No. 3: 595–613.
- Zamir, Eyal. 2014. "Contract Law and Theory: Three Views of the Cathedral."
- Zamir, Eyal, and Doron Teichman. 2018. *Behavioral Law and Economics*. Oxford University Press.