discriminatory enforcement of consumer contracts: evidence from product returns

Meirav Furth-Matzkin[[1]](#footnote-1)\*

*Recent evidence suggests that sellers often selectively enforce standardized contract terms, authorizing employees to deviate from these terms on a case-by-case basis. This Article reports on a field experiment designed to test whether discretionary enforcement of consumer contracts leads to racial or gender discrimination. Nineteen testers of different races and genders were recruited and trained to return non-receipted clothing items to Chicago-based retail stores with formal receipt requirements for returns. The findings reveal that sellers are significantly more likely to treat white female customers more favorably than required by their return policies than they do African-American or male customers who use identical bargaining strategies. In particular, African-American male testers attempting to make a non-receipted return were 28 percent less likely to be offered a refund or store credit than were white female testers returning the same item and following an identical script. The results also show that store clerks are more likely to deviate from the formal policy to the detriment of African-American customers compared to white customers (for example, by refusing to accept a return even when the policy so allows). The results suggest that discretionary enforcement of consumer contracts facilitates on-the-ground discrimination.*

# Introduction

Racial and gender discrimination has been extensively documented in multiple contexts around the globe, including in housing, employment, adjudication, and policing (e.g., Riach & Rich, 2002; Pager, 2007; Bertnard & Duflo, 2017). Despite significant progress in the field, far too little is known about discrimination in business-to-consumer transactions, particularly regarding the role of race and gender in determining market outcomes. This paucity in the literature mirrors a legislative lacuna. While the various civil rights laws of the 1960s explicitly prohibit race and gender discrimination in employment and housing, they largely overlook discrimination against minorities in consumer markets.

The few studies to address discrimination against consumers by sellers have focused on only a handful of industries (e.g*.*, car sales, credit, and vacation rentals), using differential pricing or refusals to offer the good or service altogether as the main or sole measure of discrimination (e.g., Ayres & Siegleman, 1995; Goldberg, 1996; Zussman, 2013; Hanson et al., 2016; Edelman et al., 2017). Even fewer studies have documented more subtle and covert forms of differential treatment of certain consumer groups, such as longer waiting times and lower-quality service (e.g., Lee, 2000; Davidson, 2007; Ge et al., 2016), and most of this work has relied on qualitative interviews rather than on quantitative techniques. This Article contributes to discrimination scholarship by investigating a potentially overlooked form of marketplace discrimination: selective enforcement of sellers’ formal terms and conditions.

 As detailed in Section I below, both theory and anecdotal evidence suggest that businesses may write rigid, clear-cut terms into their standardized agreements while concurrently authorizing their employees to exercise discretion when applying these terms in their interactions with consumers on the ground. Key insights from economic theories of incomplete contracting suggest that sellers may use this strategy of combining rigid contract terms with flexibility in practice, rather than refer to all possible contingencies in the contract, because it enables them to protect themselves from opportunistic buyers who could exploit a more detailed and complete contract to their advantage (e.g., Gillette, 2005; Bebchuk & Posner, 2006; Johston, 2006).

Indeed, addressing potential consumer misconduct in this manner may be the main driver behind sellers’ preferences for selective enforcement of formal agreement terms and conditions over explicit specification of all possible conditions in agreements. That being said, empirical evidence from other domains, including policing, prosecution, adjudication, and employment, suggests that decision-makers often exercise their discretionary powers discriminatorily across gender and racial lines (e.g., Paternoster, 1984; Holmberg, 2000; Bushway & Morrison Piehl, 2001; Price, 2009). For example, studies have found that female, African-American, and other minority-group job applicants are significantly less likely to receive a callback from potential employers (e.g., Neumark et al., 1996; Pager, 2003), and that residential landlords are significantly more likely to refuse to rent out apartments to minority tenants (e.g., Ewens et al., 2014; Hanson and Hawley, 2011; Carpusor and Loges, 2006). This evidence raises concerns that sellers’ discretionary authority to depart from their formal agreements may be applied inconsistently, particularly to the disadvantage of certain consumer groups, such as African-American and female customers.

Building on this evidence, this Article uses an audit study to test whether selective enforcement of retailers’ return policies results in race or gender discrimination.

I focused on discriminatory enforcement of return policies for two main reasons. First, sellers’ implementation of their return policies is discretionary and lends itself to observation through an audit study. Second, the issue of product returns is of substantial economic importance to consumers and sellers. Consumers find the ability to withdraw from consumer transactions valuable, as they frequently regret their purchases after the fact.[[2]](#footnote-2) Sellers, on the other hand, face hundreds of billions in lost profits because of returned merchandise each year, with about twenty percent of all purchases made in retail stores ultimately returned.[[3]](#footnote-3) I hope that this study of product returns will encourage other researchers, and possibly regulators or civil rights advocates, to test for discriminatory enforcement of contracts in other settings, particularly in other business-to-consumers contexts.

Data collection took place in Chicago, Illinois, in 2019. Nineteen testers (five African-American male, four African-American female, five white male and five white female testers) were recruited and trained to return non-receipted clothing items to retail stores formally requiring receipts for refunds. The testers followed a uniform script and returned the items, unused and in their original packaging, to each of the audited stores. Overall, 203 audits of 59 stores were included in the study. It was found that retail sellers often deviate from the formal terms and conditions governing product returns, and that selective enforcement of return policies often leads to racial and gender disparities.

Sellers selectively enforced their contracts, deviating from the formal policy in roughly half (54%) of the audits. They deviated in favor of consumers in 34% of the audits, while deviating to consumers’ detriment in 20% of the audits.

Three types of racial disparities are particularly striking: (1) African-American consumers were 10 percentage points less likely to receive more lenient treatment than that required by the formal policy than were white consumers (28% v. 38%).[[4]](#footnote-4) (2) African-American consumers were 7 percentage points more likely to be treated worse than the policy required compared to white testers (24% v. 17%),[[5]](#footnote-5) and (3) African-American consumers were 23 percentage points less likely to see a manager upon request (59% v. 82%), and 15 percentage points less likely to obtain an improved outcome after speaking with a manager than were white consumers (19% v. 34%).[[6]](#footnote-6)

Sellers’ selective enforcement of contracts disproportionately benefited white consumers. African-American consumers were roughly 22 percentage points less likely to have their non-receipted return accepted (for either refund or store credit) compared to similarly-situated white consumers seeking to return an identical product and following an identical script.[[7]](#footnote-7)

Gender disparities were also observed, albeit pointing in somewhat conflicting directions. While male testers were 13 percentage points less likely to experience more lenient treatment than were female testers (27% v. 40%),[[8]](#footnote-8) and 9 percentage points more likely to be treated worse than the policy requires than female testers (25% v. 16%),[[9]](#footnote-9) women were significantly less likely to see a manager upon request than men (58% v. 84%).

Male and white customers were significantly more likely to benefit from complaining and negotiating the terms at the post-contract stage than were African-American and female customers. These findings suggest that while members of the majority benefit from complaining, minority consumers are penalized for doing so.[[10]](#footnote-10)

Since females were generally treated significantly better than men (except with respect to requests to speak with a manager), and white testers were treated significantly better than African-American testers, the largest differences were observed between white females and African-American males. For example, while white female testers had an 84% likelihood of having their non-receipted return accepted (for either refund or store credit), African-American male testers had only a 58% likelihood of obtaining such outcomes.[[11]](#footnote-11) White females were 21% more likely to experience a pro-consumer gap and 15% less likely to experience an anti-consumer gap compared to African-American males.[[12]](#footnote-12)

What can explain the observed disparate treatment of similarly-situated consumers returning the same product while using the same script? The study was not designed to answer this question, but the data may offer some clues.

One potential explanation is that since non-white consumers feel less confident or more uncertain about their rights,[[13]](#footnote-13) they are less likely to negotiate the terms of the agreement,[[14]](#footnote-14) and are thus less likely to benefit from more lenient treatment compared to white consumers. Yet, this experiment suggests that even when consumers’ negotiating behavior is uniform, African-American consumers are systematically treated worse by stores’ representatives.

The literature on discrimination offers two main theories for the causes of discrimination. Taste-based theories posit that a particular group is treated significantly worse because it is disfavored or hated.[[15]](#footnote-15) In contrast, statistical theories of discrimination posit that disparate treatment stems not from distaste for, or bias against, certain minority groups, but, rather, from a seller’s desire to maximize profits in a situation of imperfect information.[[16]](#footnote-16) When information about specific individuals is limited, decision-makers may draw statistical inferences based on an individual’s group affiliation (e.g., Posner, 1989).

This study’s findings do not support any single theory of discrimination. On one hand, I find that white clerks are significantly less likely to offer refunds or store credit to African-American testers than to white testers.[[17]](#footnote-17) This finding implies that in-group bias may play a role in white clerks’ tendency to discriminate against African-Americans. On the other hand, African-American clerks are also significantly less likely to offer refunds or store credit to African-American testers than to white testers, suggesting that statistical discrimination may also be at play.[[18]](#footnote-18) In fact, white testers fare significantly better among African-American clerks than among white clerks (70% likelihood of having return accepted when dealing with white clerks versus 82.5% likelihood when dealing with African-American clerks).[[19]](#footnote-19) At the same time, African-American testers do not obtain significantly worse results among white clerks than among African-American clerks.

When looking at the relationship between a store’s high-end status and racial discrimination, another interesting pattern appears. While African-American consumers are discriminated against in both mainstream and high-end stores, they are not treated significantly worse at mainstream and discount stores. This, again, suggests that statistical discrimination might be at play. While mainstream and discount stores in Chicago typically have a large African-American customer base, higher-end stores typically cater to higher-income, predominantly white, customers. It is therefore plausible that clerks in higher-end stores perceive African-American consumers as less valuable customers. Another plausible factor behind the observed discrimination is that store clerks at higher-end stores are more suspicious toward African-American customers, who do not usually frequent these stores.

Whatever the source of the observed disparate treatment may be, the results reveal that even though sellers typically adopt a standard, uniform contract or policy, which ostensibly applies identically to all consumers transacting with the seller, in practice, selective enforcement of the contract might disproportionately harm some consumers while benefiting others.

These findings have important policy implications. While it is unfeasible, and probably not desirable, to prohibit sellers from departing from the letter of their contracts, discriminatory enforcement of these contracts can be prohibited and sanctioned. This could be done either through direct legislation or by interpreting unfair or deceptive acts or practices statutes (UDAP laws) as prohibiting discriminatory enforcement of consumer contracts.

This Article proceeds as follows. Part I briefly explains the motivation for this study, while surveying the nascent literature on selective enforcement of consumer contracts and the limited evidence on discrimination in consumer markets. Building on this evidence, Part II describes the audit study designed to test for discriminatory enforcement of retailer return policies. Part III presents the results. Part IV discusses the normative implications of the findings.

# I. Background & Motivation

## Selective Enforcement of Consumer Contracts

It has traditionally been assumed that in consumer contracts, unlike in the context of relational, business-to-business agreements, sellers have little interest in deviating from the text of their standard form agreements.[[20]](#footnote-20) In a Harvard Law Review Article on “Contracts of Adhesion,” Todd Rakoff has noted that “[t]he characteristics of firms counsel the adoption of standard forms and rigidify allegiance to them.”[[21]](#footnote-21) As Rakoff explained, the use of standard form contracts keeps “wayward sales personnel” in check by ensuring adherence to the terms of the form contract.[[22]](#footnote-22) Similarly, Stewart Macaulay has observed that form contracts efficiently control salesmen by placing customers on notice of salespeople’s “limited authority.”[[23]](#footnote-23) As Macaulay explained, firms typically use standardized agreements or policies “to avoid being legally bound to expectations [their] salesmen […] created […] that are inconsistent with company policy.”[[24]](#footnote-24) Shmulik Becher and Esther Unger-Aviram have similarly 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.”[[25]](#footnote-25)

Some commentators have recently questioned the conventional wisdom that the written agreement generally reflects the true terms of the transaction.[[26]](#footnote-26) These scholars have speculated that in competitive markets, even terms that seem rigid and unconditional on paper may be relaxed by sellers in practice.[[27]](#footnote-27) Under this theory, sellers may adopt rigid or unconditional terms to protect themselves from opportunistic buyers, while at the same time allowing agents to selectively enforce these terms in view of reputational concerns.[[28]](#footnote-28) For example, a hotel may adopt a strict “no cancellation” policy, but allow for cancellations if customers show that they need to cancel their trip for health-related reasons.[[29]](#footnote-29) An airline may allow passengers who miss their flight due to unforeseen circumstances to get on the next available flight at no additional charge, even though the formal policy requires that passengers pay a fee.[[30]](#footnote-30) A retailer may adopt a stringent return policy of accepting only receipted returns of unopened merchandise, yet exhibit more accommodating or lenient behavior under certain circumstances.[[31]](#footnote-31) Credit card issuers may decide to waive late or annual fees in certain circumstances;[[32]](#footnote-32) or mortgage service providers may choose not to utilize their contractual right to foreclose on a defaulting borrower, depending, for example, on the borrower’s credit risks.[[33]](#footnote-33)

To date, most of the commentators who noted that sellers have incentives to adopt “selective enforcement” strategies (as I term them) have generally painted a rosy picture, suggesting that such strategies are efficient and beneficial for both sellers and consumers.[[34]](#footnote-34) The presence of an ostensibly rigid contract term enables sellers to use information that they can observe only ex post (after entering into the transaction) to screen out opportunistic buyers without having to bear the costs of verifying consumer misbehavior with arbitrators or courts.[[35]](#footnote-35) Selective enforcement of consumer contracts thus enables good faith buyers to enjoy better treatment than that for which they originally contracted, while helping sellers screen out the “bad apples” who would take advantage of a more lenient or flexible term in writing, thereby enabling sellers to keep prices low.[[36]](#footnote-36)

While many commentators have lamented the efficiency gains arising from selective enforcement strategies, only few have expressed concerns that discretionary enforcement of contracts could result in regressive distributional outcomes.[[37]](#footnote-37) This is surprising in view of the mounting evidence that decision-makers often exercise discretionary powers discriminatorily.[[38]](#footnote-38)

## Discrimination in Consumer Markets

Evidence from diverse domains, ranging from policing and adjudication to housing and employment, suggests that discretion is often exercised in ways that disproportionately harm minority populations.[[39]](#footnote-39) Indeed, even in the consumer setting, accumulating evidence indicates that sellers discriminate against minority consumers in some industries (e.g*.*, car sales, credit, and vacation rentals),[[40]](#footnote-40) by, for example, quoting significantly higher prices for the same products when facing African-American or female customers compared to similarly-situated white and male customers,[[41]](#footnote-41) refusing to enter into transactions with African-American customers altogether,[[42]](#footnote-42) or disproportionately foreclosing on defaulting borrowers from poor neighborhoods.[[43]](#footnote-43)

Taken together, this evidence raises concern that sellers’ discretionary authority to depart from their formal agreements may be applied inconsistently to the disadvantage of certain groups.[[44]](#footnote-44)

If standardized contracts are applied discriminatorily, selective enforcement could generate considerable socioeconomic harm. Selective enforcement of contracts can result in regressive redistribution if minority group members are treated less favorably. Furthermore, minority consumers, expecting inferior treatment, might be discouraged from entering into transactions even if those transactions are efficient and conducive to their well-being; or 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, when a problem arises.

Despite the importance of detecting inequitable treatment of minorities, research on discriminatory enforcement of consumer contracts is remarkably scarce.[[45]](#footnote-45) Building on the vast empirical evidence showing that decision-makers often exercise discretion discriminatorily,[[46]](#footnote-46) this Article tests whether selective enforcement of consumer retail contracts systematically and disproportionately discriminates against minority consumers.

# II. The Study

## Design

This Article reports on an audit study in which testers were recruited and trained to return non-receipted clothing items to retail stores in Chicago with formal receipt requirements for returns.[[47]](#footnote-47) Nineteen testers—five white females, five white males, four African-American females,[[48]](#footnote-48) and five African-American males—tried to return 203 items to 59 stores.[[49]](#footnote-49) The testers followed a uniform script and returned the items, unused and in their original packaging, but without a receipt, to each of the audited stores.

All of the audited stores had return policies explicitly requiring receipts to be produced in order to make returns.[[50]](#footnote-50) The study focuses on the enforcement of “receipt required” return policies because they are both very common and strongly disliked by many consumers, who perceive them as an unnecessary burden.[[51]](#footnote-51) At the same time, sellers routinely require receipts for returns mainly to protect themselves from shoplifters or from consumers trying to return items purchased at a different store.[[52]](#footnote-52) This generates mixed incentives for sellers with respect to enforcement. They want to weed out the “bad apples” seeking to take advantage of lenient policies, but they also have an incentive not to unnecessarily antagonize honest consumers. The area of consumer returns is also a fruitful ground for studying discrimination in the discretionary enforcement of contracts because store clerks and managers are typically granted a significant degree of discretion about accepting non-receipted returns.[[53]](#footnote-53)

In each store, testers attempted to return a clothing item[[54]](#footnote-54) that had been purchased in advance,[[55]](#footnote-55) in its original packaging and condition, with tags attached, but *without the receipt*. They were instructed to wait in line until a store clerk became available and ask the clerk what he or she could do for them. The testers were then instructed to say that they wanted to return the item, and to place the item, still in its original packaging, on the counter. If the store clerk asked them why they wanted to return the item, testers were instructed to say that they realized they did not need it after purchasing it.[[56]](#footnote-56) If asked for the receipt, testers would answer that they thought they had the receipt with them, but, after looking for it, they appeared to have lost it. They were then instructed to await the store clerk’s response. If the store clerk agreed to provide a refund, testers were instructed to accept the refund, thank the clerk, and leave the store.

If, however, testers were denied the return or were offered anything other than a refund (e.g., exchange or store credit), they were instructed to ask to speak to a manager. Whether the store clerk refused to call a manager, identified himself or herself as the manager, or called the manager, testers were to ask once again for a refund. If still denied a refund, testers would thank the clerk or manager and leave the store. After leaving each store, all testers filled out a detailed report, describing the outcomes of the attempted returns (see Appendix II).

## Uniformity and Controls

To provide confidence that any differences in outcomes did not result from unobserved differences between testers, measures were taken to reduce intertester variation and ensure uniformity in bargaining. For this purpose, testers were recruited according to uniform criteria:

1. *Education*: All testers were undergraduate students;
2. *Age*: All testers were between 18- and 25-years-old.

Testers were also instructed to wear similar attire and were trained to behave uniformly at the store:

1. *Attire:* All testers were instructed to wear casual clothing during the audits (jeans and t-shirt);
2. *Behavior at the store:* All testers were trained to behave uniformly at the audited stores. They memorized an identical script that they followed to the letter in their interactions with store clerks and managers. The testers were told that the study was about stores’ return policies and practices,[[57]](#footnote-57) and received a list of contingent responses to the questions they were likely to encounter. If asked, they gave uniform answers about the reason for making the return and about not having the receipt. Before auditing the stores, the testers attended training sessions at the University of Chicago, where they practiced their script and participated in numerous mock negotiations meant to help them behave uniformly during the audits.

Despite these efforts to enhance uniformity, some differences between testers undoubtedly remained. Yet it is highly unlikely that the observed differences in treatment along gender and racial lines can be explained by these residual differences or by minor divergences from the uniform bargaining script testers were trained to follow.

*Time of audits.* To avoid potential changes in stores’ return policies that typically occur during the holiday season, data collection took place between March and April 2019,[[58]](#footnote-58) well after the season had ended.[[59]](#footnote-59)

## Sample Descriptive Statistics

Information on basic retailer characteristics, including size (measured by number of employees), annual revenues (for the year of 2018), and age (defined as 2019 minus the year of establishment) was also collected.[[60]](#footnote-60) Finally, as a measure of store prestige, I collected data on the median prices of all clothing and accessory items listed on each store’s website.[[61]](#footnote-61)

[Table 1]

## Coding

To test whether sellers’ discretionary enforcement of their return policies systematically discriminates against African-American consumers, each store’s formal return policy was collected and coded as: (1) requiring receipts for all returns; (2) allowing for an exchange only if there was no receipt; or (3) allowing for store credit or exchange absent a receipt (while requiring receipts for refunds).

I then noted, based on testers’ reports, whether the store clerk or manager adhered to the formal policy, deviated from the formal policy in favor of consumers, or deviated from the formal policy to the detriment of consumers. Table 2 presents the full coding scheme.

Table 2. Coding Scheme: Selective Enforcement of Return Policies

|  |  |  |
| --- | --- | --- |
| **Formal Return Policy** | **What is a pro-consumer gap?** | **What is an anti-consumer gap?** |
| Receipt Required for ***all*** returns (36% of sampled stores) | Refund, store credit or exchange offered | \_ |
| Exchange is allowed without receipt. Receipt is required for store credit or refund (5% of sampled stores) | Refund or store credit offered | Exchange denied |
| Store Credit or exchange allowed without receipt. Receipt required for refunds only (59% of sampled stores) | Refund | Store credit or exchange is denied |

# III. Results

## Store Clerks’ Responses to Testers’ Return Requests

Store clerks’ and managers’ responses to the testers’ return requests were divided into four categories: return denied, exchange, store credit, or refund. Figure 2 shows the frequency of each response according to tester race and gender. As Figure 2 shows, white testers received a refund or store credit more frequently, and had their return denied less frequently, than did African-American testers.

Figure 2: Return Outcomes by Tester Race and Gender

Table 4 presents the main effect of testers’ race and gender on their return outcomes. The dependent variable is whether the return was accepted or denied.[[62]](#footnote-62) Columns 1 and 2 report testers’ first-stage outcomes (before asking to speak with a manager) as the dependent variable, while columns 3 and 4 report final-stage outcomes (after asking to speak with a manager) as the dependent variable.

Testers’ first-stage outcomes provide relatively well-controlled tests for discrimination. Because the store clerk’s initial reaction was made with relatively little intervention on the tester’s part, it is unlikely that the initial differences in treatment are driven by any unobserved differences between testers. On the other hand, testers’ final-stage outcomes may better reflect real-world racial and gender disparities, to the extent that in real life, consumers often continue to negotiate with sellers.

Table 3

|  |  |  |
| --- | --- | --- |
|  | First-Stage Likelihood of Having the Return Accepted | Final-Stage Likelihood of Having the Return Accepted |
| White females | 76% | 84% |
| White males | 67% | 80% |
| African-American females | 62% | 63% |
| African-American males | 55% | 58% |

 The results reveal large and significant racial disparities. During the first-stage, African-American testers were 14% less likely to have their return accepted than were white testers (58% v. 72%, *p* < 0.05). During the second stage, African-American testers were 22% less likely to have their return accepted than were white testers (60% v. 82%, *p <* 0.001). At the same time, the disparities between male and female testers (both across the entire sample and within each racial sub-group) in both the initial and final stages were insignificant (70% for female testers v. 61% for male testers in the initial stage; and 75% for female testers v. 70% for male testers across the sample in the final stage). The largest disparity was observed between white female testers and African-American male testers, with white female testers almost 20% more likely to have their return accepted than were African-American male testers at the initial stage (*p* < 0.05), and 26% more likely to have their return accepted than were African-American male testers at the final stage (*p* < 0.01).[[63]](#footnote-63)

Columns 2 and 4 of the regression table introduce additional control variables related to clerk and manager demographics (race and gender) and store characteristics (i.e., store’s age, status, and whether it is private or public). The racial effect remains large and significant across these specifications; indeed, it becomes even larger.[[64]](#footnote-64)

[*Table 4*]

## Discriminatory Enforcement of Return Policies

So far, this analysis has focused on differences in return outcomes across tester groups. I now turn to explore whether sellers’ selective enforcement of their return policies systematically discriminates against consumers based on race or gender.

Table 5 presents the main effect of testers’ race and gender on their likelihood of being treated differently than the store’s formal return policy requires. The dependent variable is an ordinal variable, which takes the value of “1” if a pro-consumer gap was observed, “0” if no gap was observed, and “-1” if an anti-consumer gap was observed. Columns 1 and 2 report testers’ first-stage outcomes (i.e., whether they experienced a pro-consumer gap, no gap, or an anti-consumer gap before asking to speak with a manager) as the dependent variable, while columns 3 and 4 report final-stage outcomes as the dependent variable.

Columns 2 and 4 of the regression table introduce additional control variables related to clerk and manager demographics (race and gender) and store characteristics (i.e., store’s age, high-end status and whether it is private or public).

 As the regression table shows, both race and gender had a strong and significant effect on the likelihood of being treated differently than required by the policy.[[65]](#footnote-65) At the initial stage, white testers were 9% more likely to experience a pro-consumer gap (29% v. 20%) and 9% less likely to experience an anti-consumer gap (21% v. 30%, *p* < 0.1)than were African-American testers. At the final stage, a similar pattern is observed, with white testers 10% more likely to experience a pro-consumer gap (38% v. 28%, *p* < 0.1) and 7% less likely to experience an anti-consumer gap (17% v. 24%, *p* < 0.1) than were African-American testers.

In terms of gender disparities, male testers were 12% less likely to experience a pro-consumer gap (18% v. 30%), and 12% more likely to experience an anti-consumer gap, than were female testers (31% v. 19%) at the initial stage. A similar pattern is observed at the final stage, with male testers 13% less likely to experience a pro-consumer gap (27% v. 40%), and 9% more likely to experience an anti-consumer gap (25% v. 16%) compared to female testers.





|  |  |  |
| --- | --- | --- |
|  | First Stage | Final Stage |
|  | Likelihood of Experiencing a Pro-Consumer Gap | Likelihood of experiencing an Anti-Consumer Gap | Likelihood of Experiencing a Pro-Consumer Gap | Likelihood of Experiencing an Anti-Consumer Gap |
| White females | 35% | 16% | 45% | 13% |
| White males | 21% | 27% | 30% | 22% |
| African-American females | 24% | 24% | 32% | 20% |
| African-American males | 15% | 36% | 24% | 28% |

As Figures 3 and 4 show, there are significant racial and gender differences in testers’ likelihood of experiencing both *more* and *less* favorable treatments than dictated by the formal policy.[[66]](#footnote-66) White testers are both more likely to receive more lenient treatment than required by the policy and less likely to experience a less favorable treatment than the policy allows compared to African-American testers. Similarly, female testers are both more likely to receive a more lenient treatment than required by the policy and less likely to experience a less favorable treatment than the policy allows compared to male testers.

The most striking difference is observed, once again, when comparing white females to African-American males. At the initial stage, white female testers are more than twice as likely to experience a pro-consumer gap compared to African-American male testers (35% v. 15%) and more than twice less likely to experience an anti-consumer gap compared to African-American male testers (16% v. 36%).[[67]](#footnote-67) Similar differences remain when looking at final-stage outcomes (45% v. 32% for pro-consumer gaps, 13% v. 28% for anti-consumer gaps).[[68]](#footnote-68) The racial effect remains large and significant with and without controls.[[69]](#footnote-69)

## Likelihood of Speaking with a Manager

As noted, testers were instructed to ask to speak with a manager if denied a refund at the initial stage. The findings reveal large and significant racial and gender differences in the likelihood of seeing a manager upon request. As Figure 6 shows, white female testers have a 69% likelihood of seeing a manager, white males, a 96% likelihood, African-American females, a 42% likelihood, and African-American males, a 72% likelihood.[[70]](#footnote-70) Both racial and gender gaps are strikingly large, with females 26% less likely to speak with a manager upon request (controlling for race) than males (*p* < 0.001), and African-Americans roughly 23% less likely to speak with a manager upon request (controlling for gender) than Whites (*p* < 0.001).

As Figure 7 shows, there are also significant racial differences in the likelihood of obtaining an improved outcome conditional on speaking with a manager, with white testers 15% more likely to obtain an improved outcome than African-American testers (*p* < 0.1). White females had a 32% likelihood of obtaining an improved outcome, white males, a 35% likelihood of obtaining an improved outcome, African-American females, only a 12% likelihood of obtaining an improved outcome, and African-American males, a 22% likelihood of obtaining an improved outcome Among African-American testers, female testers were roughly 45% less likely to obtain a better outcome upon speaking with a manager.



Figure 6. Regressions of Likelihood of Seeing a Manager by Tester Race and Gender

Predicted probabilities (with 90% confidence intervals) of seeing a manager upon request, based on logistic regression of outcomes of requests to see the manager by tester race and gender.



Figure 7. Predicted Likelihood of Improved Outcomes by Tester Race and Gender

Predicted probabilities (with 90% confidence intervals) of obtaining an improved return outcome after seeing a manager, based on logistic regression of post-manager return outcomes (for those who saw a manager) on tester race and gender.

]add regression table for likelihood to improve outcomes]

 In the next two sub-sections, I explore possible factors that may influence the racial and gender disparities documented in the study. In particular, I ask whether the demographics of the clerk or manager matter or whether the status of the store has an effect. Generally, I find that discrimination is remarkably robust across different factors.

## Effects by Store Clerk and Manager Demographics

Testers reported the perceived gender and race of the store clerk with whom they interacted. This enabled me to determine whether the findings changed based on the (perceived) demographics of the store clerks. If discrimination is driven by animus or in-group bias, then the clerk’s race and gender should matter. If animus or homophily were the primary factor driving differential treatment of consumers, then African-American customers would be expected to obtain better return outcomes when interacting with African-American clerks.

Table \_\_ reports on results of tests of whether the gender and race of the clerk affected the amount of discrimination.

White clerks are significantly less likely to offer refund or store credit to African-American testers than to white testers.[[71]](#footnote-71) This finding implies that in-group bias may play a role in white clerks’ tendency to discriminate against African-Americans. However, African-American clerks are also significantly less likely to offer refund or store credit to African-American testers than to white testers, suggesting that statistical discrimination may also be operative.[[72]](#footnote-72) In fact, white testers fare significantly better with African-American clerks than with white clerks (70% likelihood of having a return accepted when dealing with white clerks versus a 82.5% likelihood when dealing with African-American clerks).[[73]](#footnote-73) At the same time, African-American testers do not obtain significantly worse results among white clerks than among African-American clerks.

[add interaction tables]

## Interaction with Store Type

As a measure of store prestige, I collected data on the median prices of all clothing and accessory items listed on each store’s website.[[74]](#footnote-74) The average median price at the sampled stores was $60 (SD = $60), with the lowest median price being $5 and the highest being $350. Stores with median prices below the 25th percentile (i.e., stores with median prices lower than $25) were classified as “discount stores.” Stores with median prices between the 25th and 75th percentiles (i.e., between $25 and $70) were classified as “mainstream stores.” Finally, stores with median prices in the upper 25th percentile (i.e., with median prices higher than $70) were classified as “high-end stores.”[[75]](#footnote-75)

While African-American consumers are discriminated against in both mainstream and high-end stores, they are not treated significantly worse at the discount stores. This, again, suggests that statistical discrimination might be operative. While discount stores in Chicago typically have a large African-American customer base, higher-end stores typically cater to higher-income, predominantly white, customers. It is therefore plausible that clerks in higher-end stores perceive African-American consumers as less valuable customers. Another plausible factor behind the observed discrimination is that store clerks at higher-end stores are more suspicious toward African-American customers, as they are not usually buying at these stores.

# IV. Discussion

## Selective Enforcement generates Problematic Redistribution

The findings of the field experiment reveal that, at least in the context of retail product returns, discretionary enforcement of the formal return policies results in robust racial discrimination. White testers were significantly more likely to have their returns accepted than were African-American testers. White testers were also significantly less likely to be treated less favorably than the policy allows than were African-American testers.

The findings also reveal that both gender and race play a role in the bargaining process and outcomes. Male and white testers were significantly more likely than female and African-American testers to see a manager upon request. And African-American testers were significantly less likely to obtain an improved outcome after speaking with a manager compared to white testers.

These results contribute to a small but growing body of literature suggesting that selective enforcement of standardized consumer contracts may result in more adverse consequences than previously assumed.[[76]](#footnote-76) The study’s findings indicate that such discretionary performance of contractual terms might lead to discrimination and yield regressive distributional outcomes.[[77]](#footnote-77)

In the context of product returns, sellers’ discretionary authority to depart from their return policies in favor of some consumers was applied inconsistently, with African-American consumers experiencing a disadvantage compared to white consumers.

Perhaps more disturbingly, the findings reveal that sellers may also deviate from their contracts to the *detriment* of consumers, by refusing to perform the contract as written; and that selective deviations from the contract to the detriment of consumers (which could be considered breach of contract) also disproportionately harm African-American consumers relative to white customers.

## Why do sellers discriminate in contract enforcement?

 One important limitation of this experiment is that it does not enable identification of the mechanism causing worse outcomes for African-American customers. For example, it cannot be ruled out that sellers may draw inferences about the customer’s socioeconomic status, and consequently about the customer’s value to the store, based on the customer’s perceived race. Similarly, the experiment does not provide a sharp test of various theories of the sources of discrimination. The theoretical literature on discrimination typically distinguishes between statistical and taste-based (or animus-based) discrimination. While the study’s experimental design cannot rule out either mechanism, the findings suggest a more nuanced and multifaceted scenario than either of these classic explanations. Indeed, the results offer indications of in-group bias among white clerks. For example, white male and female clerks are significantly less likely to treat white customers less favorably than the formal policy allows than they are African-American customers. However, African-American clerks are also less likely to treat white customers less favorably than the formal policy allows than they are African-American customers. This finding suggests that the results may actually be driven by statistical inferences rather than by animus or in-group favoritism. At the very least, these findings seem to conflict with pure taste-based explanations for the observed discrimination.

There might be at least three possible statistical-based explanations for the observed racial discrimination. First, the observed racial discrimination might be driven by differences in store clerks’ inferences about the likelihood that each consumer group would abuse the store’s return policy, steal from the store, or return an item that was not purchased at the store. Second, the observed racial discrimination might stem from differences in store clerks’ inferences about consumers’ socioeconomic status and value to the store. If clerks typically believe that white customers are likely to be wealthier than African-American customers, and consequently more valuable customers, they might treat African-American customers less favorably than white customers. Third, the differential treatment might be driven by differences in the perceived likelihood—or unlikelihood—of minority consumers complaining and generating reputational harm if dissatisfied. If clerks assume that minority consumers are less likely to complain, and, in any event. less likely to generate reputational harm to the store, they might be less inclined to treat them more favorably than the policy requires.

All of these explanations may co-exist, and further research is needed to assess the importance and explanatory power of each.

## The Future of Discriminatory Enforcement

Although e-commerce retail sales currently account for only ~15% of all retail sales worldwide, online shopping is increasingly gaining popularity.[[78]](#footnote-78) According to a 2016 Pew Research Center survey of U.S. consumers, roughly 80% of Americans do at least some of their shopping online.[[79]](#footnote-79)

Should we expect to see more, less, or similar levels of discrimination with online platforms? Arguably, we may expect to see less discrimination, at least with respect to race and gender, as these characteristics would be less salient, and probably more difficult to observe on online marketplaces. Nonetheless, sophisticated and high-income customers will likely benefit from selective enforcement online more than lower-income, less sophisticated consumers, because they will negotiate more or ask for more lenient treatment. They may also be treated more favorably in view of the higher value to the firm of keeping them pleased. While it is plausible that the effects of race and gender will be less pronounced on the online settings, several studies cast doubt on this proposition. In particular, experiments have already shown that race and gender discrimination persist on the online marketplace. An experiment on Airbnb has found that guests with distinctively African-American names are significantly less likely to be accepted than similarly-situated guests with white-sounding names.[[80]](#footnote-80) Similarly, an experiment on eBay has found that women were disadvantaged as sellers compared to men when selling the exact same product in eBay auctions.[[81]](#footnote-81) These studies suggest that discrimination persists even on online markets.

A related question is whether technological advancements, such as the increased use of algorithmic intelligence and big data, will gradually replace salespeople’s discretion and lead to less biased outcomes. In the specific context of product returns, for example, sellers are already increasingly able to track serial returners automatically. Sellers may choose whether or not to behave more leniently towards customers seeking to make returns based on the customers’ purchasing histories.[[82]](#footnote-82) It is possible that market discrimination could be reduced if sellers have more transactional information on consumers, such as their purchasing history and past return behavior.[[83]](#footnote-83) In addition, sellers could be required to formally exclude protected characteristics, such as gender and race, from their algorithms. However, recent research in the domain of algorithmic credit pricing suggests that discrimination is likely to persist even when sellers scrutinize algorithmic input. The nearly endless range of correlations in big data, combined with the flexibility and complexity of machine learning, threaten to perpetuate existing disparities.[[84]](#footnote-84)

# V. Taking Away the License to Discriminate

This Article provides systematic empirical evidence showing that selective enforcement of consumer contracts might lead to racial and gender discrimination. In this experiment, sellers’ discretionary performance of their return policies disproportionately harmed African-American consumers compared to similarly-situated white consumers.

These findings suggest that sellers’ strategy of complementing seemingly harsh or incomplete contract terms with an internal policy granting their representatives discretion in exercising their contractual rights (or even in performing their duties) may often yield regressive distributional outcomes and generate welfare losses.

 One seemingly straight-forward solution is to prohibit sellers from deviating from the letter of the contract.[[85]](#footnote-85) Yet, this approach is probably very difficult, if not unfeasible, to implement and enforce, given sellers’ strong incentives to violate such prohibitions. Who is to know if a store clerk allows a customer to exchange an item or offers store credit even when the formal policy is “no returns”?

Furthermore, enforcement agencies will have very limited incentive to investigate and prosecute such violations, especially in the context of “small-stakes” violations or minor deviations from the contract terms. More importantly, perhaps, such a strict prohibition on deviations from the contract might generate unwarranted outcomes—for example, if sellers choose to enforce one-sided terms against all consumers as a result.

While enforcing the one-sided term against all consumers may be more equitable than enforcing it only against some consumers based on their race or other suspect grounds, it is still inferior to solutions that allow sellers flexibility in performing contracts while still prohibiting discrimination. That is, a better solution would allow sellers to distinguish between opportunistic and good-faith buyers and exercise discretion on a case-by-case basis, while still preventing them from discriminating against minority consumers.

Finally, many contracts are relational in nature and constantly evolving. It cannot be expected of sellers to predict all contingencies ex ante and write them into the contract. A solution that allows sellers discretion in the performance of their contracts but prohibits discriminatory enforcement of contracts is therefore preferable.

The problem is that, to date, many forms of marketplace discrimination, including discriminatory performance of consumer contracts, are not explicitly prohibited, either at the state or federal level. More generally, antidiscrimination law is a patchwork of federal and state statutes in which notable gaps remain.

Indeed, antidiscrimination laws exist, at both the state and federal levels. The Civil Rights Act of 1966 prohibits discrimination in hotels and other public accommodations based on race, color, religion, or national origin. Section 1981 specifically guarantees to “[a]ll persons within the jurisdiction of the United States” the same right as “white citizens” to “make and enforce contracts,” and prohibits discrimination on the basis of race, ethnicity, alienage, religion, and color.[[86]](#footnote-86) Section 1982 similarly prohibits intentional discrimination in real and personal property transactions.[[87]](#footnote-87) However, neither of these statutes recognizes disparate impacts or prohibits discriminatory enforcement of contracts.[[88]](#footnote-88)

Other federal regulations have traditionally focused on credit and housing discrimination. For example, the Equal Credit Opportunity Act (ECOA) prohibits discrimination in credit transactions against persons on the basis of race, color, religion, national origin, sex, age, and other suspect grounds. Similarly, the Fair Housing Act prohibits creditors involved in residential real-estate transactions from discriminating against any person on the basis of race, color, religion, sex, handicap, familiar status, or national origin. Yet, even in these limited contexts, it is unclear whether “disparate impact” discrimination is prohibited in addition to intentional, disparate treatment.

This Article proposes that legislatures explicitly prohibit and impose sanctions for discriminatory enforcement of consumer contracts. by. One possibility is to explicitly recognize that such discrimination is an “unfair” or “abusive” act or practice covered by federal and state laws prohibiting unfair, deceptive, and sometimes abusive acts or practices (“UDAP” laws). Both the Consumer Financial Protection Bureau (CFPB) and the Federal Trade Commission (FTC) have authority to take action to prevent service providers from engaging in unfair acts or practices when transacting with consumers.[[89]](#footnote-89)

An “unfair” (or “abusive”) act is defined as an act that is likely to cause substantial injury to consumers, is not reasonably avoidable, and is not outweighed by countervailing benefits to consumers or competition.[[90]](#footnote-90) Until explicit legislation is passed, courts could interpret current UDAP laws as prohibiting discriminatory performance of consumer contracts.[[91]](#footnote-91) Note that it is crucial not to require that regulators prove that the practice is intentional. Rather than requiring the showing of intent to discriminate, all regulators should be required to prove is disparate impact, i.e., that an ostensibly neutral policy or practice disproportionately harms members of protected classes. To demonstrate that certain selective enforcement policies lead to disparate impact, regulators may need to audit sellers, using an approach like the one used and reported on in this study—and much like the longstanding efforts to reduce discrimination in the housing, credit, and employment markets. In these markets, audits have served as critical tools in uncovering these more subtle, hidden forms of discrimination.

Of course, merely prohibiting discriminatory enforcement of consumer contracts is not enough. Legislative reforms should be complemented by strengthened enforcement efforts, both by state attorneys general and by federal agencies. I also propose that private enforcement be allowed, and in particular, that consumers who suffer racial or gender discrimination be allowed to file class actions against the discriminating seller or service provider. Here, as in the context of public enforcement, it is also imperative to allow for disparate impact claims, as intentional discrimination would often be difficult to prove.

The findings illustrate the importance of addressing a gaping hole in existing consumer protection efforts. To date, differential treatment of consumers based on their demographic characteristics is not explicitly prohibited or recognized as an unfair or deceptive act or practice. This study’s findings should bring this legislative omission to policymakers’ attention and encourage them to revisit the current state of the law. Until legislative or executive action is taken, courts may consider interpreting UDAP laws as prohibiting discrimination.

When assessing the seller’s violations and deciding on proper sanctions, regulators could consider various factors, including: (1) whether the discriminatory enforcement of the contract or policy is a repeat violation or a one-time occurrence; (2) the size of the consumer group affected; (3) the magnitude of the harm suffered by consumers; and (4) seller’s history, especially with respect to compliance with UDAP laws.

Sanctions may include requiring sellers to discontinue the discriminatory practice, to provide restitution or compensation to aggrieved consumers, and to pay a fine or a civil money penalty.

1. \* Assistant Professor, UCLA School of Law. S.J.D., LL.M. (Harvard University). 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.] 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). [Thanks to be added]. [↑](#footnote-ref-1)
2. (Becher & Zarsky, 2011; Kats, 2018) [↑](#footnote-ref-2)
3. (Reagan, 2016; Orendoff, 2019; Stojanovic, 2019) [↑](#footnote-ref-3)
4. This difference is significant at the 10% level. [↑](#footnote-ref-4)
5. This difference is significant at the 10% level. [↑](#footnote-ref-5)
6. This difference is significant at the 10% level. [↑](#footnote-ref-6)
7. 60% likelihood for African-American testers v. 82% likelihood for white testers. This difference is significant at the 0.1% level. [↑](#footnote-ref-7)
8. This difference is significant at the 5% level. [↑](#footnote-ref-8)
9. This difference is significant at the 5% level. [↑](#footnote-ref-9)
10. Cite articles from here: https://www.abajournal.com/news/article/showing\_anger\_in\_the\_courtroom\_can\_backfire\_for\_women\_lawyers\_study\_suggest [↑](#footnote-ref-10)
11. These differences are significant at the 0.1% level. [↑](#footnote-ref-11)
12. [↑](#footnote-ref-12)
13. [↑](#footnote-ref-13)
14. [↑](#footnote-ref-14)
15. [↑](#footnote-ref-15)
16. (e.g., Phelps, 1972; Arrow, 1973; Aigner & Cain, 1977). [↑](#footnote-ref-16)
17. The difference is significant at the 10% level, based on a logistic regression of return outcomes on tester race, clerk race, and the interaction between tester and clerk race. [↑](#footnote-ref-17)
18. The difference is significant at the 10% level, based on a logistic regression of return outcomes on tester race, clerk race, and the interaction between tester and clerk race. [↑](#footnote-ref-18)
19. The difference is significant at the 10% level, based on a logistic regression of return outcomes on tester race, clerk race, and the interaction between tester and clerk race. [↑](#footnote-ref-19)
20. *See, e.g.,* Jason Scott Johnston, *The Return of Bargain: An Economic Theory of How Standard-Form Contracts Enable Cooperative Negotiation between Businesses and Consumers*, 104 Mich. L. Rev. 857, 857 (2006) (“Among attorneys, judges, and legal academics, there is virtual consensus that the widespread use by business firms of standard-form contracts in their dealing with consumers has completely eliminated bargaining in consumer contracts.”). [↑](#footnote-ref-20)
21. Todd D. Rakoff, *Contracts of Adhesion: An Essay in Reconstruction*, 96 Harv. L. Rev. 1173, 1222–23 (1983). [↑](#footnote-ref-21)
22. *Id. See also* Todd D. Rakoff, *The Law and Sociology of Boilerplate*, 104 Mich. L. Rev. 1235, 1235 (2006). [↑](#footnote-ref-22)
23. Stewart Macaulay, *Private Legislation and the Duty to Read - Business Run by IBM Machine, the Law of Contracts and Credit Cards,* 19 Vand. L. Rev. 1051, 1059 (1966). [↑](#footnote-ref-23)
24. *Id.* [↑](#footnote-ref-24)
25. Shmuel I. Becher & Esther Unger-Aviram, *The Law of Standard Form Contracts: Misguided Intuitions and Suggestions for Reconstruction*, 8 DePaul Bus. & Com. L.J. 199, 201 (2010). [↑](#footnote-ref-25)
26. *See, e.g.,* Lucian A. Bebchuk & Richard A. Posner, *One-Sided Contracts in Competitive Consumer Markets*, 104 Mich. L. Rev. 827, 828 (2006) (suggesting that sellers’ concerns for reputations may lead them to behave more leniently than their contracts demand); Clayton P. Gillette, *Rolling Contracts as an Agency Problem*, 2004 Wis. L. Rev. 679, 704–12 (2004); Clayton P. Gillette, *Pre-Approved Contracts for Internet Commerce*, 42 Houston L. Rev. 975, 977 (2005); Johnston, *supra* note 1. [↑](#footnote-ref-26)
27. *See, e.g.,* Johnston, *supra* note 1, at 858 (“firms use clear and unconditional standard-form contract terms not because they will insist upon these terms, but because they have given their managerial employees the discretion to grant exceptions from the standard-form terms on a case-by-case basis.”). [↑](#footnote-ref-27)
28. Bebchuk & Posner, *supra* note 7, at 827–28 (“[a] seller concerned about its reputation can be expected to treat consumers better than is required by the letter of the contract. But the seller’s right to stand on the contract as written will protect it against opportunistic buyers.”). [↑](#footnote-ref-28)
29. Bebchuk & Posner, *supra* note 7, at \_\_\_. [↑](#footnote-ref-29)
30. *See, e.g.*,Claire Nowak, *This Little-Known Airplane Rule Can Help When You Miss Your Flight*, Reader’s Digest, available at https://www.rd.com/advice/travel/missed-flight-flat-tire-rule/; Amanda Harding, *This Surprising Airline Rule Can Help If You Miss Your Flight*, Shobiz Cheatsheet (May 5, 2018), available at https://www.cheatsheet.com/culture/this-surprising-airline-rule-can-help-if-you-miss-your-flight.html/. [↑](#footnote-ref-30)
31. [↑](#footnote-ref-31)
32. *See, e.g., See* Michelle Crouch, *Poll: You Can Get Better Credit Card Terms Just by Asking* (March 27, 2017), available at https://www.creditcards.com/credit-card-news/late-fee-waiver-poll.php. [↑](#footnote-ref-32)
33. [↑](#footnote-ref-33)
34. *See, e.g.,* Johnston, *supra* note 1, at 858 (“a firm will often provide benefits to consumers […] beyond those that its standard form obligates it to provide. […] Were firms legally required to extend such benefits […]—then both firms and their customers would be worse off”); Bebchuk & Posner, at 828 (“A one-sided contract may thus be preferred ex ante by informed parties as a cheaper mechanism for inducing efficient outcomes, should contingencies arise during the performance of the contract, than a more “balanced” contract that, because of imperfect enforcement, could create costs as a consequence of consumers’ enforcing protective provisions in the contract.”). [↑](#footnote-ref-34)
35. *See, e.g.,* Bebchuk & Posner, *supra* note 1. [↑](#footnote-ref-35)
36. *See, e.g.,* Gillette, *Rolling Contracts*, *supra* note 7, at 705 (noting, for example, that “if sellers systematically provide redress where goods are clearly defective, but systematically contest less credible disputes about product quality, then the insertion of a clause into an RC that disfavors buyers may be less problematic, because the clause is applied disproportionately against bad claimants.”).  [↑](#footnote-ref-36)
37. *See, e.g., See* Shmuel I. Becher & Tal Z. Zarsky, *Minding the Gap*, 51 Conn. L. Rev. 69, 91 (2019) (suggesting that “uninformed and weak groups of consumers” will be disadvantaged, as “sophisticated and informed” groups will plausibly be treated more forgivingly or generously); Eyal Zamir, *Contract Law and Theory: Three Views of the Cathedral*, 81 U. Chi. L. Rev. 2077, 2100 (2014) (suggesting 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”). [↑](#footnote-ref-37)
38. e.g., Applegate, 1982; Paternoster, 1984; Holmberg, 2000; Bushway & Morrison Piehl, 2001; Price, 2009 [↑](#footnote-ref-38)
39. [↑](#footnote-ref-39)
40. There is also evidence that buyers exercise their discretion discriminatorily, for example by giving African-American taxi drivers significantly lower tips, or by offering lower prices to female sellers in online markets. *See* Ian Ayres, Fredrick E. Vars & Nasser Zakariya, *To Insure Prejudice: Racial Disparities in Taxicab Tipping,* 114 Yale L. J. 1613 (2005) (documenting racial discrimination in taxicab tipping); Tamar Kricheli-Katz & Tali Regev, *How Many Cents on the Dollar” Women and Men in Product Marekts*, 2 Sci. Adv. 1 (2016) (documenting gender discrimination in eBay auctions). [↑](#footnote-ref-40)
41. *See, e.g.,* Ian Ayres and Peter Siegelman, *Race and Gender Discrimination in Bargaining for a New Car*, 85 Am. Econ. Rev. 304, 305–06 (1995).In this well-known study, pairs of testers were trained to bargain uniformly and then sent to negotiate for the purchase of a new car at 153 Chicago dealerships. Notwithstanding the identical approach to bargaining, Ayres and Siegelman have found that white males were quoted significantly lower prices than white female and African-American (male and female) buyers. *See also* Ian Ayres, *Fair Driving: Gender and Race Discrimination in Retail Car Negotiations*, 104 Harv. L. Rev. 817, 822–27 (1991). [↑](#footnote-ref-41)
42. (e.g., Ayres & Siegleman, 1995; Goldberg, 1996; Zussman, 2013; Hanson et al., 2016; Edelman et al., 2017). There is also qualitative evidence suggesting that minority consumers feel they are discriminated against by sellers. *See, e.g.,* Edith F. Davidson, *Shopping while Black: Perceptions of Discrimination in Retail Settings*, PhD diss., University of Tennessee (2007), available at <https://trace.tennessee.edu/utk_graddiss/147/> (reporting, based on a series of interviews, that African-Americans feel that they are constantly subject to racial profiling in retail stores); Aronte M. Bennet et al., *Shopping while Nonwhite: Racial Discrimination among Minority Consumers*, 49 J. of Consumer Affairs 328 (2015) (reporting, based on a survey, that non-white customers are significantly more likely to feel discriminated against in stores than white customers). [↑](#footnote-ref-42)
43. Manisha Padi, *Contractual Inequality*, \_\_\_ Michi. L. Rev. (2021). [↑](#footnote-ref-43)
44. [↑](#footnote-ref-44)
45. For a notable exception, *see* Manisha Padi, *Contractual Inequality*, \_\_\_ Michi. L. Rev. (2021) (finding, based on a national sample of mortgage agreements, that mortgage servicers disproportionately exercise their contractual right to foreclose on borrowers from poor neighborhoods). [↑](#footnote-ref-45)
46. *See* Section I, *infra*. [↑](#footnote-ref-46)
47. The methods used in this study have been approved by the University of Chicago Institutional Review Board prior to data collection. The study (IRB Study Number: IRB 18-1529), including its hypotheses, analyses, and design, was pre-registered in AsPredicted (see Post-Contract Discrimination in the Retail Market, #16928, created in 11/23/2018). [↑](#footnote-ref-47)
48. I initially hired 5 African-American female testers. However, one tester dropped during training, after expressing safety concerns and reluctance to follow the script while facing white male clerks and managers in the audited stores. [↑](#footnote-ref-48)
49. The sample initially consisted of all 192 retail stores with a Chicago location appearing in the ReferenceUSA and Hoover’s Company Directories’ databases. Before launching the study, research assistants coded each store’s return policy, as appearing at the retailer’s website, in-store sign (if present), or receipts. Stores whose return policy did not explicitly require a receipt for exchange, store credit, or cash refund (n = 31) were excluded from the sample. Dollar stores, stores that did not offer items for $30 or less, or stores that did not have a downtown Chicago location (n = 101) were also excluded from the sample to make the study more manageable. One store was excluded from the sample due to deviations from the script. The final sample includes 59 retail stores located in downtown Chicago (for a list of the audited stores, see Appendix 1). Each store was supposed to be audited four times by four testers—an African-American male, an African-American female, a White a male, and a White female. However, due to discarded tests and scheduling difficulties, the final sample included 203 audits instead of 236. The 203 audits include 59 audits by white female testers, 51 by white male testers, 51 by African-American male testers, and 42 by African-American female testers. [↑](#footnote-ref-49)
50. In the initial, larger sample of 192 retail stores, receipt was explicitly required for returns in 84% (161 out of 192 policies) of the return policies. [↑](#footnote-ref-50)
51. For evidence that receipt requirements are perceived as a hassle by consumers, see, e.g., Janakiraman, Syrdal, & Freling, *supra* note 38. [↑](#footnote-ref-51)
52. Another opportunistic behavior that retail sellers often try to protect themselves from is “buy-to-rent” strategies, whereby consumers purchase a product in order to use it for a specific purpose or event, and then return it for a full refund afterwards, thereby obtaining free rent. In order to protect themselves from such opportunistic behavior, retailers often formally require consumers to return items “unused” and “with tags attached.” The receipt requirement is less effective in protecting sellers from the “buy-to-rent” problem, since consumers can buy the item, use it, and then return it with a receipt. In order to minimize concern that sellers’ reluctance to accept the returns would be driven by their suspicions that the items were stolen, testers were instructed to return the items unused, in their original packaging, and with tags attached. [↑](#footnote-ref-52)
53. In informal interviews I conducted with store clerks working in Chicago, several interviewees mentioned being granted a considerable degree of discretion in deciding whether or not to accept a return. [↑](#footnote-ref-53)
54. To minimize differences acrossstores, purchasers were instructed to buy a clothing accessory (i.e., a hat, gloves, socks, scarves, purses, or bags). If no accessories were available, they were instructed to buy a shirt, pants, or another clothing item. They were specifically instructed to refrain from buying 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. Products’ prices were kept constant at between $20 to $30. [↑](#footnote-ref-54)
55. Research assistants (purchasers) were sent to purchase the items in advance. They paid in cash so that sellers would not obtain any personal information from the purchase. The items were then returned by different members of the research team—the testers. This design was chosen for several reasons. First, it allowed for the purchase of identical items from each store, so that items would not vary within stores. Second, it mitigated the concern that in some stores, store clerks would identify the person making the return whereas in others, testers would encounter different store clerks. Still, this design raises the concern that, in some stores, store clerks were more suspicious of the testers making the returns because they did not identify them. This may mean that, in real-life, sellers are even more likely to deviate from their formal return policy, when they recognize the customer seeking to make the return. Of course, increased suspicion cannot explain the differential treatment based on race and gender. [↑](#footnote-ref-55)
56. It is possible that store clerks’ willingness to depart from the formal policy would vary depending on 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 or 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 more credible and reasonable. [↑](#footnote-ref-56)
57. The testers were not informed that the study tested for discrimination, although some testers reported, in their post-audit reports, that they felt discriminated against on the basis of race. [↑](#footnote-ref-57)
58. For consistency, testers were instructed to audit the stores on weekdays in the afternoons. [↑](#footnote-ref-58)
59. As data collection ended well before the COVID-19 pandemic broke, changes in return policies as a result of the pandemic also could not affect the results. [↑](#footnote-ref-59)
60. This data was also obtained from Bloomberg and Hoover’s Company Directories. [↑](#footnote-ref-60)
61. Python was used to scrape the stores’ websites. Coders and programmers were instructed to derive the median prices of the items based on clothing items only, in order to keep the analysis tractable across stores with different offerings. Median prices were chosen instead of mean prices, as mean prices—unlike median prices—are affected by outliers (i.e., extremely expensive or very cheap products). Some stores blocked access to their websites, and these websites (n = 17) were manually coded. One store’s website was impossible to manually code, so its median price was coded as missing. One store did not have items for sale online, so its median price was also coded as missing. [↑](#footnote-ref-61)
62. If testers were offered a refund or store credit, the return outcome was coded as “return accepted.” Otherwise, it was coded as “return rejected.” [Please see robustness checks with different dependent variables in Appendix \_\_] [↑](#footnote-ref-62)
63. Within the female sub-group, white female testers were 14% more likely to have their return accepted than African-American female testers at the first stage, and 19% more likely at the final stage. The differences between white and African-American females are significant at the 5% and 1% levels when store and clerk controls are included in the regression. [↑](#footnote-ref-63)
64. When adding store and clerk controls, white female testers are likely to have their return accepted for refund or store credit 79% of the time, white males—75% of the time, and African-American male and female testers—only 58% of the time. The differences between white females and African-American males and females in likelihood to be offered a refund or store credit are significant at the 5% level under the specifications which include the controls. At the final stage, the differences in return outcomes between white and African-American males is significant, and so is the difference between white and African-American females, and between white females and African-American males. The difference between white males and African-American females is not significant. [↑](#footnote-ref-64)
65. The racial difference is significant at the 10% level, and the gender difference is significant at the 10% level (for final outcomes). [↑](#footnote-ref-65)
66. Again, the racial effect remains strong even when controlling for tester and store characteristics, with white female testers likely to experience an anti-consumer gap only 12% of the time, white males—20% of the time, African-American females—33% of the time, and African-American males—as frequently as 32% of the time. [↑](#footnote-ref-66)
67. The difference becomes significant at the 1% level when controls are added to the regression. [↑](#footnote-ref-67)
68. White female testers are also 50% more likely to experience a pro-consumer gap than African-American female testers When controlling for store and clerk characteristics, the differences between white and African-American female testers become significant at the 1% level. [↑](#footnote-ref-68)
69. At the initial stage, when adding store and clerk controls, white female testers are likely to experience a more lenient treatment 41% of the time, white males—28% of the time, and African-American male and female testers—only 17% of the time. At the final stage, when adding the controls, white female testers are 49% likely to experience a pro-consumer gap, white male testers- 37% likely, African-American male testers—20% likely, and African-American female testers—22% likely. At the final stage, white females are 13% likely to experience an anti-consumer gap, white males—22% likely, African-American females—21% likely, and African-American males—28% likely. The differences between white male and female testers, and between white females and African-American male testers are significant at the 10% and 1% levels respectively. [↑](#footnote-ref-69)
70. The differences between white male and female testers are significant at the 1% level, as well as the differences between African-American females and white male and female testers. The differences between African-American male testers and white female testers are small and insignificant. The racial and gender differences remain strong and significant even when controlling for tester and store characteristics. [↑](#footnote-ref-70)
71. The difference is significant at the 10% level, based on a logistic regression of return outcomes on tester race, clerk race, and the interaction between tester and clerk race. [↑](#footnote-ref-71)
72. The difference is significant at the 10% level, based on a logistic regression of return outcomes on tester race, clerk race, and the interaction between tester and clerk race. [↑](#footnote-ref-72)
73. The difference is significant at the 10% level, based on a logistic regression of return outcomes on tester race, clerk race, and the interaction between tester and clerk race. [↑](#footnote-ref-73)
74. Python was used to scrape the stores’ websites. Coders and programmers were instructed to derive the median prices of the items based on clothing items only, in order to keep the analysis tractable across stores with different offerings. Median prices were chosen instead of mean prices, as mean prices—unlike median prices—are affected by outliers (i.e., extremely expensive or very cheap products). Some stores blocked access to their websites, and these websites (n = 17) were manually coded. One store’s website was impossible to manually code, so its median price was coded as missing. One store did not have items for sale online, so its median price was also coded as missing. [↑](#footnote-ref-74)
75. This analysis was repeated with classification based on my RA’s coding of each store as “discount,” “mainstream,” and “high-end” based on their impressions from being at the store’s physical location. The results were similar and robust. [↑](#footnote-ref-75)
76. Lucian A. Bebchuk & Richard A. Posner, *One-Sided Contracts in Competitive Consumer Markets*, 104 Mich. L. Rev. 827, 828 (2006) (suggesting that “reputational considerations” may “induce the seller to treat the buyer fairly even when such treatment is not contractually required”); Clayton P. Gillette, *Rolling Contracts as an Agency Problem*, 2004 Wis. L. Rev. 679, 704–12 (2004) (suggesting that sellers may use a “contract clause that assigns an entitlement to the seller [in order to protect themselves from consumer misbehavior], but that the seller may under-enforce [the clause] when it is dealing with a good claimant”); Clayton P. Gillette, *Pre-Approved Contracts for Internet Commerce*, 42 Houston L. Rev. 975, 977 (2005) (making a similar observation); Jason Scott Johnston, *The Return of Bargain: An Economic Theory of How Standard-Form Contracts Enable Cooperative Negotiation between Businesses and Consumers*, 104 Mich. L. Rev. 857, 858 (2006). For a similar observation in the context of franchise agreements, see Benjamin Klein, *Transaction Cost Determinants of “Unfair” Contractual Arrangements*, 70 Am. Econ. Rev. 356, 358–60 (1980) (explaining that franchisors may include harsh termination clauses in their agreements to prevent cheating, while reputational mechanisms will constrain them from relying on these terms opportunistically). For a more recent paper suggesting that firms’ behavior is often more generous than their contractual language, see Shmuel I. Becher & Tal Z. Zarsky, *Minding the Gap*, 51 Conn. L. Rev. 69 (2019). [↑](#footnote-ref-76)
77. See, e.g., Manisha Padi, Contractual Inequality, \_\_ Michi. L. Rev \_\_ (2022); Becher & Zarsky, *supra* note \_\_, at 91 (suggesting that “uninformed and weak groups of consumers” will be disadvantaged, as “sophisticated and informed” groups will plausibly be treated more forgivingly or generously); Eyal Zamir, *Contract Law and Theory: Three Views of the Cathedral*, 81 U. Chi. L. Rev. 2077, 2100 (2014) (suggesting 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”). [↑](#footnote-ref-77)
78. *See, e.g.*, Stojanovic, *supra* note 27. [↑](#footnote-ref-78)
79. [↑](#footnote-ref-79)
80. Benjamin Edelman, Michael Luca, and Dan Svirsky, *Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment*, 9 American Economic Journal 1 (2017). [↑](#footnote-ref-80)
81. Tamar Kricheli-Katz & Tali Regev, *How Many Cents on the Dollar? Women and Men in Product Markets*, 2 Science Advances (2016). [↑](#footnote-ref-81)
82. *See, e.g.*, Orendoff, *supra* note 28. [↑](#footnote-ref-82)
83. *See, e.g.*, Lior Jacob Strahilevitz, *Reputation Nation: Law in an Era of Ubiquitous Personal Information*, Nw. U. L. Rev. 102 (2008); Lior Jacob Strahilevitz, *Less Regulation, More Reputation, in* The Reputation Society 71 (2012) (suggesting that “an important potential upside of new reputation tracking technologies is their potential to displace statistical discrimination on the basis of race, gender, age, appearance, and other easily observable characteristics”). [↑](#footnote-ref-83)
84. *See, e.g.,* Talia Gillis, *The Input Fallacy*, \_\_\_ Minnesota Law Review \_\_ (2022); Talia Gillis & Jann L. Spiess, *Big Data and Discrimination*, 86 U. Chi. L. Rev. 459 (2018). [↑](#footnote-ref-84)
85. A similar suggestion has been proposed by Ian Ayres, Fredrick E. Vars, and Nasser Zakariya to eliminate racial disparities in taxicab tipping. After document such discrimination, the researchers propose to prohibit all tipping or to adopt a *service compris* regulation. [↑](#footnote-ref-85)
86. 42 U.S.C. § 1981. McDonald v. Santa Fe Trail Transp. Co., 427 U.S. 273, 287 (1976) (noting that § 1981 was enacted to protect persons of “every race and color”); [↑](#footnote-ref-86)
87. [↑](#footnote-ref-87)
88. Saint Francis Coll. v. Al-Khazraji, 481 U.S. 604, 613 (1987) (holding that § 1981 was “intended to protect from discrimination identifiable classes of persons who are subjected to intentional discrimination solely because of their ancestry or ethnic characteristics”). [↑](#footnote-ref-88)
89. 12 U.S.C. § 5531 (CFPB UDAAP authority); 15 U.S.C. § 45(n) (FTC UDAP authority). [↑](#footnote-ref-89)
90. Id. [↑](#footnote-ref-90)
91. For an even broader suggestion to interpret current UDAP laws as prohibiting all types of discrimination in consumer markets, see Stephen Hayes and Kali Schellenberg, *Discrimination is “Unfair”: Interpreting UDA(A)P to Prohibit Discrimination*, Student Borrower Protection Center Paper (April 2021). [↑](#footnote-ref-91)