Can customer ratings be discrimination amplifiers?  
Evidence from a gig economy platform

Fei Teng  
Yale School of Management, fei.teng@yale.edu

Tristan Botelho  
Yale School of Management, tristan.botelho@yale.edu

K. Sudhir  
Yale School of Management, k.sudhir@yale.edu*

This paper investigates the conjecture that rating systems can lead to discriminatory spillovers and become “discrimination amplifiers.” Because rating systems memorialize differences in individual ratings (impacted by group-based statistical or taste-based discrimination) as differences in worker quality, displaying average ratings on a platform can lead to discrimination spillovers for customers who do not discriminate and amplify discrimination for those who do, leading to greater inequity. After demonstrating the idea using a stylized analytical model, we investigate the question of discrimination amplification empirically using data from an online labor market platform that connects service workers with customer jobs. Using a model of customer’s job cancellations and rating choice and allowing for unobserved heterogeneity in discriminatory behaviors, we identify three segments: one shows no difference in behavior towards minority and White workers; the second cancels minority workers at higher rates, hurting minority earnings; a third cancels minorities more and rates them lower than Whites. We find that customer discrimination increases sub-5 star ratings (5 is the highest rating) for the average minority worker by 23% and decreases earnings by 4%. Displaying ratings amplifies discrimination and increases the minority rating gap in 5 star ratings by 22% and the minority earnings gap by 115% relative to not displaying ratings.

Key words: statistical discrimination, taste discrimination, prejudice, minorities, earnings gap, customer rating systems, online platforms, gig economy

* The authors are grateful for insightful discussions and comments from Khai Chiong, Soheil Ghili, Vineet Kumar, Jiwoong Shin, Kosuke Uetake and participants in the Yale Quantitative Marketing Brown Bag Seminar.
1. Introduction

The proliferation of digital commerce platforms (e.g., Amazon, eBay, Taobao) and mediated marketplaces (e.g., Airbnb, Uber, TaskRabbit) has institutionalized customer feedback mechanisms—such as five-star or thumbs up/thumbs down evaluation system—as an integral part of the transactional landscape. These platforms transcend a wide range of sectors, with individuals relying on peer feedback to make choices in industries ranging from travel (e.g., Tripadvisor) and dining (e.g., Yelp) to healthcare (e.g., Healthgrades), education (e.g., Coursera), and legal services (e.g., Avvo). Moreover, these consumer-sourced rating systems have evolved into primary tools for evaluating and semi-automating the management of disaggregated workforces in gig economy platforms, such as TaskRabbit, Uber, and Upwork (Rosenblat et al. 2016). On the surface, this appears to be a “win-win” for customers, service providers, and platforms alike; the prevailing assumption (or hope) is that rating systems bolster customer confidence, thereby stimulating a greater volume of exchange, while enhancing revenue for workers and platforms.

The conventional wisdom is also that online rating systems can mitigate discrimination against marginalized groups (Cui et al. 2020). The argument is that these systems, by offering quantifiable performance metrics based on customers’ post-transaction experiences, can minimize statistical discrimination—where customers make choices based on beliefs about differences in group performance based on individual characteristics (e.g., race, gender).¹ However, a growing body of research challenges this optimistic view, documenting racial and gender biases in customer ratings across online platforms (e.g., Hannák et al. 2017, Botelho and Gertsberg 2022). This dissonance has led to concerns that customer rating systems could actually serve as “vehicles for discrimination,”

¹ Statistical discrimination as a mechanism for discrimination was proposed by (Phelps 1972, Arrow 1973), in contrast to taste-based discrimination proposed by Becker (1957), where discrimination arises due to animus or prejudice. More recently, there has been interest in whether statistical discrimination arises from accurate or inaccurate beliefs of differences between groups (Bordalo et al. 2016). However, Lang and Spitzer (2020) argue that inaccurate group belief based discrimination is better understood as taste based discrimination—when the evidence against the inaccurate beliefs are clear.
exacerbating earnings and employment discrimination against workers from underrepresented or protected groups, such as racial and ethnic minorities or women (Rosenblat et al. 2016, Hanrahan et al. 2018).

In this paper, we propose that customer rating systems can serve as more than vehicles for discrimination; they could act as “discrimination amplifiers.” We posit that when ratings capture customer preferences influenced by either taste-based or statistical discrimination—even if only from a segment of customers—the public display of these aggregate ratings to subsequent consumers can amplify discrimination over time. This effect can even create a discrimination spillover among customers who otherwise do not discriminate. Because customers view ratings as metrics of quality, displaying workers’ average ratings “memorializes” any differences in ratings that may have arisen from discriminatory preferences or beliefs as differences in “quality.” As such, it impacts the beliefs about the “quality” of the rated individuals for future customers. Thus, this mechanism not only intensifies discriminatory practices among those who already engage in them but also creates spillovers affecting future customers who may not have initially discriminated. A preference for the highest-rated workers leads to a widening gap in ratings and, by extension, earnings between minority and White workers, even if both groups may have been similar on quality at the outset. In essence, the very act of displaying customer ratings can perpetuate and amplify systemic disparities between minority and White workers in both ratings and earnings, even if only a fraction of customers discriminate.

To clarify and build intuition for our hypothesis, we develop a simple and stylized analytical model to formalize the intuition of how customer rating systems can act as discrimination amplifiers. Our model demonstrates that displaying aggregate customer ratings leads to two outcomes (i)
it allows the discriminatory preferences of one customer segment to spill over to the other customer segment who would not otherwise discriminate, and (ii) it reinforces and amplifies the discrimination practiced by the segment already engaged in discriminatory behavior. Importantly, this inequity is not reflective of the actual performance quality of minority workers. As such, it serves as an illustration of how systemic disadvantages can manifest due to the structure of the evaluation processes, even when the platform itself may have had no intent to disadvantage minorities.

Next, we explore the issue empirically in the context of an online labor market platform that connects customers with home service workers. Unlike prior research that documents the presence of discrimination in ratings, our study aims to quantify how discriminatory ratings translate into long-term earnings inequities over time, which presents several challenges. First, we need to account for worker quality independent of the effect of displayed ratings. We exploit a feature of the platform’s rating display rule to obtain an estimate of worker quality independent of display ratings. Specifically, the platform does not display ratings for the first five rated jobs—this helps to identify the effect of displayed rating, independent of worker quality. Second, customer discrimination is not homogeneous; there is heterogeneity in both the prevalence and degree of discriminatory behavior among customers. Therefore, our model accounts for this unobserved variability. Third, discrimination can occur at different stages. Some customers may choose to cancel appointments when paired with minority providers, while others may engage with the provider, but differentially refrain from leaving reviews/ratings for minority providers. Some may leave a rating but rate minority providers lower than White providers. Each form of discriminatory behavior has its own implications for the dynamic evolution of ratings.

To address these challenges, we develop a flexible model that captures discrimination at different stages of the consumer journey, while also accounting for unobserved heterogeneity. Our empirical approach benefits from the two-sided panel data structure, which includes both customers and workers. This allows for identification of unobserved heterogeneity. In our context, customer jobs are assigned to both minority and White workers, with some customers using the platform repeatedly.
Concurrently, workers perform a large number of jobs over time, serving a variety of customer segments and generating a range of displayed metrics, such as ratings and number of past jobs completed.

We use the estimates of the model and counterfactual analyses to address three key research questions. First, is there evidence of customers discriminating against minority workers, and if so, does this discrimination manifest heterogeneously, such as through increased job cancellations or biased ratings? Second, to what extent does customer discrimination affect the ratings and earnings of minority workers compared to their White counterparts? Third, does the display of customer ratings serve as “discrimination amplifiers,” spilling over to those who do not discriminate? If yes, how much does this amplification contribute to the ratings and earnings gap between minority and white workers?

Our findings yield several key insights. Our model estimates reveal that there are three distinct customer segments: a neutral segment that does not engage in discrimination against minority workers, and two segments that do—but in different ways. The first discriminatory segment primarily cancels jobs assigned to minority workers, while the second segment discriminates at both the job cancellation and rating stages. Counterfactual analysis shows that customer discrimination contributes 23% of the ratings gap and 4% of the earnings gap between minority and White workers. Moreover, consistent with our conjecture, the public display of customer ratings induces discrimination spillovers among the non-discriminating customer segment and amplifies discrimination among the two customer segments that already engage in discriminatory behavior. Overall, the display of average ratings increases discrimination; it exacerbates the ratings gap by 22% and widens the earnings gap by 115% between minority and White workers.

Though our paper focuses on the empirical setting of online ratings, the fundamental principle we expose—that the memorialization of discrimination-tainted metrics as markers of “merit” or “quality” can spillover to amplify discrimination—has implications far beyond evaluation processes on digital platforms. Our logic is relevant in other societal and economic areas including organizational hiring/promotions, education, and criminal justice. For example, Owens and McLanahan
Can customer ratings be discrimination amplifiers?

(2020) show that teacher biases can lead to greater minority student suspensions and expulsions. When these punitive actions become part of a student’s academic record—or the implicit impressions about a student—they can have long-term negative impacts, both on the student’s subsequent disciplinary actions and academic performance. Similarly, initial disparities in law enforcement regarding drug use between minorities and Whites can create a biased criminal records, influencing future legal outcomes for the same illegal behaviors. In the context of the Indian caste system and college admission, Subramanian (2019) argues that relabeling structural disadvantages as differences in observable “merit” not only legitimizes existing inequalities but also portrays efforts to redress these structural inequities as antithetical to a merit-based society. Our analytical approach offers a framework to formally explore and substantiate such phenomena, especially in settings where the evolution of “merit” and outcomes can be fully observed.

From a modeling perspective, our paper contributes to the burgeoning discourse on fairness and algorithmic discrimination within machine learning. A unique contribution is our advancement of the important role of unobserved heterogeneity among discriminatory agents. Although extant research often focuses on “average bias” or bias conditional on observables, our research highlights how unobserved heterogeneity in bias dynamically influences the “observable metrics” that are subsequently presented to customers. This detail is crucial for machine learning applications—and the study of the conditions under which bias occurs in evaluations more broadly—as it underscores the need to account for such heterogeneity in efforts to mitigate bias. To the extent that “observable metrics” used by a system carry the imprint of past discrimination, they will not only embed but also amplify existing structural inequities and disparities over time.

From a social standpoint, our focus on unobserved heterogeneity challenges the notion of “average discrimination.” Framing discrimination as an average phenomenon often leads to public resistance to research findings. It is common for individuals to consider themselves merit-oriented and non-discriminatory, which is likely accurate. By allowing for unobserved heterogeneity in discrimination, our model presents a more accurate representation of market dynamics that allows for interaction
Can customer ratings be discrimination amplifiers?

between discrimination and “merit.” Furthermore, our modeling approach offers a valuable lens when customer choices may be influenced not only by discriminatory attitudes but also by evolving control variables, including the lingering effects of lagged discrimination.

The remainder of this paper is structured as follows: Section 2 situates our work within the broader literature on racial discrimination and the ratings and earnings gap. Section 3 presents a stylized analytical model that formalizes our core hypothesis and associated intuition, elaborating how biased rating systems can adversely affect minority groups through bias spillovers and amplification. Section 4 describes the empirical setting and the data. Section 5 outlines the empirical model we use to analyze customer cancellations and ratings. Section 6 presents our results—descriptive analysis, model estimates, and findings. Section 7 delves into our counterfactual results. Finally, Section 8 offers concluding remarks.

2. Related Literature

We situate this paper within the broader literature on racial discrimination and the ratings/earnings gap between minorities and Whites. We categorize the existing literature on consumer-related racial discrimination into three primary streams, as outlined in Table 1: (i) discrimination against consumers by firms or workers; (ii) discrimination against suppliers by consumers; and (iii) discrimination against workers by consumers or firms. It is important to acknowledge that the literature in these areas is vast; thus, the references we provide are merely examples for the categories.

From a marketing and organizations perspective, the literature on customer discrimination holds particular relevance. This body of work has notably concentrated on discrimination in the housing (e.g., Yinger 1995) and credit (e.g., Blanchflower et al. 2003) markets. Compared to these areas, research on consumer product market discrimination is relatively small (e.g., Ayres and Siegelman 1995, Edelman et al. 2017). The literature also extends to supplier discrimination. On Ebay, prices for products from Blacks (Ayres et al. 2015) and females (Kricheli-Katz and Regev 2016) are lower than their White and male counterparts. Similarly, Black and female hosts on Airbnb generate lower revenues than White and male hosts for similar properties (Marchenko 2019). Zhang et al. (2021) address the potential for a remedy for such discrimination using an algorithmic pricing tool.
The literature on racial discrimination of workers and its impacts on the labor market was kickstarted by Becker (1957). Becker organizes discrimination in labor markets to arise from the racial preferences of employers, co-workers, and customers. Becker notes that while competitive pressures can potentially eliminate discrimination arising from the first two preferences (employers or co-workers), discrimination arising from customer preferences is likely more insidious and persistent—as businesses will be rewarded by customers for discrimination against racial minorities.

Table 1: Illustrative Sample of Literature on Racial Discrimination in Markets

<table>
<thead>
<tr>
<th>Example Papers</th>
<th>Measured Outcome of Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Discrimination of consumers by market type</strong></td>
<td></td>
</tr>
<tr>
<td>Housing</td>
<td>Yinger (1995)</td>
</tr>
<tr>
<td></td>
<td>Homes shown for buying and rental</td>
</tr>
<tr>
<td>Credit</td>
<td>Blanchflower et al. (2003)</td>
</tr>
<tr>
<td></td>
<td>Credit approval for small business</td>
</tr>
<tr>
<td>Consumer Products</td>
<td>Ayres and Siegelman (1995)</td>
</tr>
<tr>
<td></td>
<td>Bargained Car Prices</td>
</tr>
<tr>
<td>Airbnb</td>
<td>Edelman et al. (2017)</td>
</tr>
<tr>
<td></td>
<td>Guest Acceptance</td>
</tr>
<tr>
<td><strong>Discrimination of suppliers by setting</strong></td>
<td></td>
</tr>
<tr>
<td>Ebay</td>
<td>Kricheli-Katz and Regev (2016)</td>
</tr>
<tr>
<td></td>
<td>Prices for Products by Gender</td>
</tr>
<tr>
<td></td>
<td>Ayres et al. (2015)</td>
</tr>
<tr>
<td></td>
<td>Prices for Products by Race</td>
</tr>
<tr>
<td>Airbnb</td>
<td>Marchenko (2019)</td>
</tr>
<tr>
<td></td>
<td>Prices for Hosts by Gender and Race</td>
</tr>
<tr>
<td><strong>Discrimination of workers by source of bias</strong></td>
<td></td>
</tr>
<tr>
<td>Employer</td>
<td>Bertrand and Mullainathan (2004)</td>
</tr>
<tr>
<td></td>
<td>Hiring of Workers</td>
</tr>
<tr>
<td>Co-worker</td>
<td>Bodvarsson and Partridge (2001)</td>
</tr>
<tr>
<td></td>
<td>NBA team composition</td>
</tr>
<tr>
<td>Consumer</td>
<td>Lynn and Sturman (2011)</td>
</tr>
<tr>
<td>(Our focus)</td>
<td>Ratings of Servers</td>
</tr>
<tr>
<td></td>
<td>Brewster and Lynn (2014)</td>
</tr>
<tr>
<td></td>
<td>Tips for Servers</td>
</tr>
<tr>
<td></td>
<td>Bar and Zussman (2017)</td>
</tr>
<tr>
<td></td>
<td>Hiring in Labor-intensive Services</td>
</tr>
<tr>
<td></td>
<td>Kuppuswamy and Younkin (2020)</td>
</tr>
<tr>
<td></td>
<td>Hiring in Hollywood movies</td>
</tr>
<tr>
<td></td>
<td>Botelho and DeCelles (2023)</td>
</tr>
<tr>
<td></td>
<td>Job Cancellations on a Labor Platform</td>
</tr>
</tbody>
</table>

Our focus is on the discrimination of workers (service providers) by consumers—an issue that holds particular relevance to marketers across various service industries. While the literature in this area is relatively limited, existing studies corroborate the influence of consumer preferences on labor market discrimination, as initially put forth by Becker. For example, Lynn and Sturman (2011) show that customers tend to rate servers of the same race more favorably on promptness and attentiveness than those of a different race. Brewster and Lynn (2014) document a “Black tip
Can customer ratings be discrimination amplifiers?

penalty,” wherein White and Black restaurant goers tip Black servers less than White ones, irrespective of the quality of service. Similarly, Bar and Zussman (2017) find customer discrimination against Arab workers in labor-intensive services within the Israeli market, affecting both hiring practices and service pricing; firms with Arab workers charge lower prices for similar services. Here, Jewish customers prefer firms employing Jewish, rather than Arab workers. In the entertainment sector, Kuppuswamy and Younkin (2020) find that consumer preferences drive the lack of minority hiring in Hollywood movies. In online labor markets, Botelho and DeCelles (2023) observe higher cancellation rates when jobs are assigned to minority workers than when assigned to White workers. Overall, these results directly support Becker’s predictions that customer preferences drive labor market discrimination.

Discrimination in the workplace has far-reaching economic implications, leading to a significant lifetime earnings gap and stunted wealth accumulation for minorities. Although racial differences in social and economic backgrounds contribute to wealth disparities (Altonji and Doraszelski 2005), they cannot fully account for the observed gaps. Even after controlling for a variety of demographic, marital, and socioeconomic factors, more than 70% of the racial gap in wealth remains unexplained (Oliver and Shapiro 2013, Shapiro et al. 2004). Comprehensive reviews of the literature further support this, indicating that differential socioeconomic resources are insufficient to explain the wealth and earnings gap. Moreover, even when controlling for relevant human capital characteristics, Blacks are less likely to be hired, have to search longer for jobs, have less work experience, and tenure, and earn lower wages compared to Whites (Pager and Shepherd 2008, Tomaskovic-Devey et al. 2005).

Unlike the traditional regression-based approaches that aim to quantify the minority earnings gap across the economy—or even a large industry sector—our paper uses a case study methodology from industrial organization. We build a process model of consumer choices that allows for discrimination within a specific empirical context. Through counterfactual analysis, we quantify the impact of platform-specific or policy decisions—such as the display of online ratings—on the earnings gap. The approach is particularly useful for firms; it enables them to make design decisions that could mitigate discriminatory outcomes within their unique setting.
3. A Stylized Analytical Model

To fix ideas and intuition for how rating systems can be discrimination amplifiers that magnify the ratings and earnings gap, we develop a stylized analytical model of an online labor market where workers encounter a mixing of customers and study how displaying ratings impact ratings and earnings.

3.1. Model Setup

Assume that there are two workers in the market. One worker belongs to the advantaged group (denoted as $A$) and the other worker belongs to the disadvantaged group (denoted as $D$). We assume that one group has an advantage in the market because either (i) some consumers may prefer to do business with them (which captures the notion of taste-based discrimination), or (ii) may hold a potentially inaccurate belief that the group is more skilled or provides better quality of service (which captures the notion of statistical discrimination based on inaccurate beliefs).\(^3\) Let $w$ denote worker type. We have $w \in \{A, D\}$. We allow for the possibility that true quality can differ between the two types of workers. We normalize the quality of the advantaged worker to be $q^A = 1$ and the quality of the disadvantaged worker to be $q^D = 1 + q$. This allows for the quality of the disadvantaged worker to be equal to the advantaged worker ($q = 1$), greater than the advantaged worker ($q > 0$) or lower than the advantaged worker ($q < 0$).

We assume a unit mass of customers who are of two types $c = \{N, P\}$. Type $N$ denotes customers who are neutral, in that they do not discriminate in their behavior towards the two groups; type $P$ denotes partial customers, in that they are partial and discriminate in favor of the advantaged group.\(^4\) Let $\alpha \in (0, 1)$ be the share of partial customers and $1 - \alpha$ be the share of neutral customers. The partial customer may discriminate either at the cancellation stage by canceling jobs or at the

\(^3\) The assumption that there are no objective differences between the two groups rules out statistical discrimination based on accurate beliefs about differences between the groups, but our key insights around spillovers and bias magnification due to display of ratings are robust even if we allow for such differences.

\(^4\) We use the words “neutral” and “partial” and avoid the use of the words “biased,” and “unbiased” so that the discriminatory behavior can encompass both taste-based and statistical discrimination.
rating stage by giving a lower rating. We start with a model that specifically focuses on the rating discrimination, assuming each customer offers one job in each period. Subsequently, we extend the analysis to incorporate discrimination in job cancellations.

3.2. Case A: No Job Cancellations

Suppose each customer offers one job in each period. We denote the rating for worker \( w \in \{A, D\} \) by customer \( c \in \{N, P\} \) at time \( t \in \{1, 2, 3, \ldots\} \) as \( r_{tc}^{wC} \). Let \( r_t^w \) be the composite rating of worker \( w \) across all jobs from the neutral and partial customers for the current period \( t \), which we operationalize as the geometric mean of the ratings of jobs from both types of customers:

\[
r_t^w \equiv \left( r_{tN}^w \right)^{1-\alpha} \left( r_{tP}^w \right)^{\alpha}
\]

Similarly, we denote the cumulative composite rating of worker \( w \) across all jobs up to period \( t \) as \( R_t^w \), operationalized as the geometric mean of all ratings (including jobs from period \( t \)):

\[
R_t^w \equiv \left( \prod_{t'=1}^{t} r_t^w \right)^{\frac{1}{t}} = \left( R_{t-1}^w \right)^{1-\alpha} \left( r_t^w \right)^{\frac{1}{t}}
\]

As a normalization, we assume that \( R_0^w = 0 \forall w \in \{A, D\} \).

Past research has shown that early reviews impact future reviews by consumers (Park et al. 2021). We capture this idea in the model by allowing for correlation between the displayed rating for the worker and their rating for new jobs. We do so by assuming that the rating \( r_{tc}^{wC} \) is the product of the worker quality and the displayed ratings. Additionally, partial customer discounts the rating of the disadvantaged customer relative to the advantaged customer by a factor of \( 1 - \delta \), where \( \delta \) is the degree of rating discrimination.

For the first period, since there can be no displayed rating, we normalize \( R_0^w = 1 \), both types of customers rate the workers by work quality. So for the advantaged worker, \( r_1^{Ac} = 1 \forall c \in \{N, P\} \) and \( R_1^A = 1 \). For the disadvantaged worker, \( r_1^{DN} = 1 + q \), \( r_1^{DP} = (1 - \delta)(1 + q) \), and \( R_1^D = (1 - \delta)^{\alpha}(1 + q) \).

\(^5\) Our propositions hold numerically even with the arithmetic mean of past ratings and the intuitions are the same. The geometric mean helps analytical tractability to obtain closed-form results.

\(^6\) We also present model-free evidence in our setting for this key assumption.
We restrict the parameter space for the analysis such that parameters \( \{\delta, \alpha, q\} \) satisfies \( 0 < (1 - \delta)^\alpha (1 + q) < 1 \). This parameter range allows for the presence of discrimination, and allows for \( q \) to be positive, zero or negative, i.e., disadvantaged groups can have higher, equal or lower intrinsic quality. Starting from the second period, how the ratings evolve depends on whether past ratings are displayed or not displayed to the customer. We discuss each of the cases below.

i. Ratings not displayed: Here the ratings of the advantaged and disadvantaged workers are straightforward. For the advantaged worker, neutral and partial customers give a rating of 1 in all periods; i.e, \( r^A_t = r^{AP}_t = 1, \forall t \). So the geometric mean is \( r^A_t = 1, \forall t \) and the cumulative geometric mean up to period \( t \) is \( R^A_t = 1, \forall t \).

For the disadvantaged worker, neutral customers give a rating of \( 1 + q \) and partial customers give a rating of \( (1 - \delta)(1 + q) \) for all periods, i.e., \( r^D_t = r^{DN}_t = 1 \) and \( r^{DP}_t = (1 - \delta)(1 + q) \), \( \forall t \). The geometric mean rating is \( r^D_t = (1 - \delta)^\alpha (1 + q) \) in each period \( t \). The cumulative geometric mean ratings up to period \( t \) is \( R^D_t = (1 - \delta)^\alpha (1 + q) \).

ii. Ratings displayed: Starting from the second period, when customers are shown the geometric mean of past ratings of the worker \( R^w_{t-1} \), the ratings of the current period are summarized in the following table:

<table>
<thead>
<tr>
<th>( c )</th>
<th>( R^A_{t-1} )</th>
<th>( (1 + q)R^D_{t-1} )</th>
<th>( (1 - \delta)(1 + q)R^D_{t-1} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( c = N )</td>
<td>( r^{DN}_t )</td>
<td>( R^D_t )</td>
<td>( R^D_t )</td>
</tr>
<tr>
<td>( c = P )</td>
<td>( r^{AP}_t )</td>
<td>( R^A_t (1 - \delta) )</td>
<td>( R^D_t )</td>
</tr>
</tbody>
</table>

Again, it is straightforward to see that the overall rating of the advantaged worker remains 1 across time: \( r^A_t = 1 \) and \( R^A_t = 1 \ \forall t \).

For the disadvantaged worker, as we discuss before, in the first period, \( R^D_1 = (1 - \delta)^\alpha (1 + q) \). For period \( t \geq 2 \), following Table 2, the geometric mean of the rating across two types of customers for the current period is

\[
r^D_t = (r^{DP}_t)^\alpha (r^{DN}_t)^{1-\alpha} = (1 - \delta)^\alpha (1 + q)R^D_{t-1}
\]
The cumulative geometric mean ratings at the end of period \( t \) is given by

\[
R^D_t = (R^D_{t-1})^{\frac{\delta_t}{\alpha_t}} = (R^D_{t-1})^{\frac{1}{\alpha_t}}((1 - \delta)^\alpha(1 + q)R^D_{t-1})^{\frac{1}{\alpha_t}} = (1 - \delta)^{\frac{\delta_t}{\alpha_t}}(1 + q)^{\frac{1}{\alpha_t}}R^D_{t-1}
\]

Given \( R^D_1 = (1 - \delta)^\alpha(1 + q) \), the recursive equation above can be reduced to

\[
R^D_t = \left((1 - \delta)^\alpha(1 + q)\right)^{\frac{1}{\alpha_t}}
\]

Since the displayed ratings cause future customers to interpret the impact of the partial customer’s discriminatory rating as a metric of quality, it creates a spillover of the discrimination onto neutral customers and makes the partial customer even more discriminatory. We highlight these points in the following proposition and the graph below. The analysis leads to the following two propositions on discriminatory spillovers and discriminatory amplification.

**Proposition 1 (Discriminatory Spillover to Neutral Customers).** When ratings are displayed, the neutral customer’s rating for the disadvantaged worker monotonically decreases in rating partiality \( \delta \) and the size of the partial customer segment \( \alpha \); i.e., \( \partial r^{DN}_t / \partial \delta < 0 \), \( \partial r^{DN}_t / \partial \alpha < 0 \).

In contrast, when ratings are not displayed, the neutral customer’s rating for the disadvantaged worker is not impacted by the rating partiality and the size of the partial customer segment; i.e., \( \partial r^{DN}_t / \partial \delta = 0 \), \( \partial r^{DN}_t / \partial \alpha = 0 \).

**Proposition 2 (Discrimination Amplification).** When ratings are displayed, the rating of disadvantaged worker decreases over time for both neutral and partial customers i.e., \( \partial r^{DP}_t / \partial t < 0 \) and \( \partial r^{DP}_t / \partial t < 0 \). The rating gap between the advantaged and disadvantaged workers also expands over time, i.e. \( \partial (R^A_t - R^D_t) / \partial t > 0 \).

In contrast, when the ratings are not displayed, the rating from both neutral and partial customers to the disadvantaged worker and the rating gap remains constant over time i.e., \( \partial r^{DN}_t / \partial t = 0 \), \( \partial r^{DP}_t / \partial t = 0 \) and \( \partial (R^A_t - R^D_t) / \partial t = 0 \).

Figure 1 graphically shows the results of Proposition 1 and Proposition 2, where we illustrate how displaying ratings leads to discrimination spillover to neutral customers and amplifies the discrimination for both the neutral and partial segments. For the figure, we set \( q = 0 \) to ensure
that the observed differences between advantaged and disadvantaged workers are not influenced by quality differences between the workers.

The first plot shows the rating from the neutral customers, where the rating remains to be 1 when the cumulative ratings are not displayed. However, when ratings are displayed, the ratings for the disadvantaged workers are negatively affected by partial customers. This captures our idea of discrimination spillover. The second plot shows the ratings from the partial customers. From the first two plots, we observe that when ratings are displayed, both the neutral and partial customer’s ratings become less favorable to the disadvantaged worker. Moreover, the partial customer always remains less favorable by a factor of $1 - \delta$. Finally, the third plot shows the overall ratings of advantaged worker and the disadvantaged worker when ratings are displayed. While the rating curve for the advantaged worker remains flat, the downward sloping trend for the disadvantaged worker shows that displaying rating further amplifies the impact of partiality, expanding the rating gap between the workers.

![Figure 1](image_url)  
*Figure 1  Illustration of Proposition 1 and Proposition 2: Ratings for the Disadvantaged Worker.*

Note: $q = 0$, $\delta = 0.2$ and $\alpha = 0.2$.

### 3.3. Case B: Customers Cancel Some Jobs Based on Ratings

In the setting where there are job cancellations, We assume that the customer’s decision to accept/cancel a job upon being connected with the worker depends on the worker’s displayed rating, which is a signal of worker quality. In each period $t$, the neutral customers accept a job with probability $R^w_{t-1}$ for workers $w \in \{A, D\}$. Thus if the cumulative rating $R^w_{t-1} = 1$, the job is
Can customer ratings be discrimination amplifiers?

always accepted and not canceled. The partial customer, however, accepts a job for the advantaged worker at probability $R_{t-1}^A$ but accepts a job for the disadvantaged worker at probability $\gamma R_{t-1}^D$, discounting the worker’s cumulative rating. We have $0 < \gamma < 1$ and $1 - \gamma$ indicates the level of discrimination in cancellation. Again, to confine the parameter space so that the problem is within an interesting domain, we assume $0 < (1 - \delta) \frac{\alpha \gamma}{\alpha + \alpha \gamma} (1 + q) < 1$.

We compare the case when ratings are not displayed versus when ratings are displayed. To keep track of the workers’ earnings, we denote the number of jobs accepted by customer type $c$ and completed by worker type $w$ in period $t$ as $n_{wc}^t$. Note that $n_{wc}^t$ only accounts for the net job count after cancellation. The total number of jobs completed by worker $w \in \{A, D\}$ in period $t$ is $n_w^t \equiv n_w^{wN} + n_w^{wP}$. The total number of jobs completed at the end of period $t$ is given by $N_w^t \equiv \sum_{t'=1}^{t} n_{w}^{t'}$.

i. Ratings not displayed:

When ratings are not displayed, similar to the previous case, it is straightforward to see that $R_{t}^A = 1 \forall t$. Hence, all jobs are accepted for the advantaged worker. Therefore, $n_{t}^A = 1 \forall t$ and $N_{t}^A = t \forall t$.

As for the disadvantaged worker, the neutral customers always accept the job, while the partial customers accept jobs with probability $\gamma$, reflecting their discrimination in cancellation process. For the disadvantaged worker, the total job count for period $t$ is given by

$$n_{t}^D = 1 - \alpha + \alpha \gamma$$

(2)

The cumulative job count up to period $t$ is therefore

$$N_{t}^D = t(1 - \alpha + \alpha \gamma)$$

Similar to the case without cancellation, ratings are given by $r_{t}^{DN} = 1 + q$, $r_{t}^{DP} = (1 - \delta)(1 + q)$, $r_{t}^{D} = (1 - \delta)^\alpha (1 + q)$ and $R_{t}^{D} = (1 - \delta)^\alpha (1 + q)$.

ii. Ratings displayed: Table 3 shows the expected job count (after accounting for cancellation probabilities) and the rating as functions of cumulative rating:
Table 3  Expected job count and rating at period $t$

<table>
<thead>
<tr>
<th>Worker $w$</th>
<th>Customer $c$</th>
<th>Job Count ($n_{wc}^t$)</th>
<th>Rating ($r_{wc}^t$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>N</td>
<td>$(1 - \alpha)R_{t-1}^A$</td>
<td>$R_{t-1}^A$</td>
</tr>
<tr>
<td>A</td>
<td>P</td>
<td>$\alpha R_{t-1}^A$</td>
<td>$R_{t-1}^A$</td>
</tr>
<tr>
<td>D</td>
<td>N</td>
<td>$(1 - \alpha)R_{t-1}^D$</td>
<td>$(1 + q)R_{t-1}^D$</td>
</tr>
<tr>
<td>D</td>
<td>P</td>
<td>$\alpha \gamma R_{t-1}^D$</td>
<td>$(1 - \delta)(1 + q)R_{t-1}^D$</td>
</tr>
</tbody>
</table>

It is trivial to see that, similar to the case when ratings are not displayed, the advantaged worker always receives a rating of 1 from both types of customers and thus receives one job in each period. That is, $n_t^A = 1$, $r_t^A = 1$, $R_t^A = 1$, and $N_t^A = t$, $\forall t$. For the disadvantaged worker, based on Table 3, the total job count for period $t$ is given by

$$n_t^D = (1 - \alpha + \alpha \gamma)R_{t-1}^D$$  \hspace{1cm} (3)

The average rating for period $t$ is

$$r_t^D = \left( (1 + q)R_{t-1}^D \right)^{\frac{1 - \alpha}{1 - \alpha + \alpha \gamma}} \left( (1 - \delta)(1 + q)R_{t-1}^D \right)^{\frac{\alpha \gamma}{1 - \alpha + \alpha \gamma}} = (1 - \delta)^{\frac{\alpha \gamma}{1 - \alpha + \alpha \gamma}}(1 + q)R_{t-1}^D$$  \hspace{1cm} (4)

The cumulative job count can be written as

$$N_t^D = (1 - \alpha + \alpha \gamma) \sum_{t'=0}^{t-1} R_{t'}^D$$

The cumulative rating can be shown to be

$$R_t^D = (R_{t-1}^D)^{\frac{N_{t-1}^D}{N_{t-1}^D + n_t^D}} \left( r_t^D \right)^{\frac{n_t^D}{N_{t-1}^D + n_t^D}} = \left( (1 - \delta)^{\frac{\alpha \gamma}{1 - \alpha + \alpha \gamma}}(1 + q) \right)^{\sum_{t'=0}^{t-1} R_{t'}^D} R_{t-1}^D$$  \hspace{1cm} (5)

We can show that Proposition 1 and Proposition 2 continue to hold for the case where customers cancel jobs based on ratings. That is, the discrimination in ratings extends to the neutral customers and is amplified over time.

Additionally, from Eq. (2), we see that there is an earnings gap for the disadvantaged relative to the advantaged. Comparing equations (2) and (3), we see that cumulative ratings further magnify

7 We initialize $R_0^D = 1$ so that in the first period, $n_0^{DN} = 1 - \alpha$ and $n_0^{DP} = \alpha \gamma$, which is consistent with the case when ratings are not displayed.
the earnings gap as \( R_i^D < 1 \). Hence, when cancellations are induced by ratings, there will be an earnings gap for the disadvantaged relative to the advantaged. Further, the earnings gap will be magnified by displaying ratings. This is summarized in Proposition 3.

**Proposition 3** (Earnings Gap). *In the presence of cancellations due to ratings, there will be an earnings gap for the disadvantaged worker. The earnings gap will be larger when ratings are displayed, compared to when ratings are not displayed.*

Fig. 2 shows the corresponding earnings gap when cancellation is based on ratings. The first plot shows that, compared to the advantaged worker, the accumulated earnings gap from the partial customers continues to widen across time, and the gap is even wider when ratings are displayed. Moreover, as the second plot of Fig. 2 shows, when ratings are displayed, the spillover effect is also carried over to the job count, where there is earning gap between the two workers even from neutral customers. As a result, overall the earnings gap is larger when ratings are displayed relative to when ratings are not displayed.

**Figure 2 Illustration of Proposition 3: Earnings Gap**

![Graphs showing earnings gap](image)

*Note:* \( q = 0, \alpha = 0.2, \delta = 0.2 \) and \( \gamma = 0.8 \). In the second plot, the line for the advantaged worker and the line for the disadvantaged worker without displayed ratings coincide.

Finally, we note that even if the disadvantaged workers is of higher quality, i.e., \( q > 0 \), there can be positive ratings and earnings gaps when \((1 - \delta)^{\frac{-\alpha\gamma}{\alpha + \gamma}} < 1\). This can be seen from our restriction on the parameter space that \((1 - \delta)^{\frac{-\alpha\gamma}{\alpha + \gamma}} < 1\) and is summarized in the corollary below:

**Corollary 1.** If \((1 - \delta)^{\frac{-\alpha\gamma}{\alpha + \gamma}} < 1\), then there exists \( q > 0 \), where positive ratings and earnings gaps for the disadvantaged worker arise.
3.4. Ratings Gap when disadvantaged workers have favorable initial conditions

In Cases A and B discussed above, we assume that in each period, each worker encounters a fixed mixing of neutral and partial customers. However, in real-world situations, the proportion of partial customers encountered by disadvantaged workers can vary over time. We illustrate a setting where in the initial period (i.e. the first period), the disadvantaged worker is fortunate not to encounter partial customers. Then if \( q > 0 \), i.e., disadvantaged workers are of higher quality, the disadvantaged worker can have higher initial rating. But over time, as they encounter partial customers, their rating can go below the advantaged workers.

To illustrate this crossover succinctly in the model, we assume that there is no job cancellation as in Section 3.2. Suppose in the first period, \( \alpha = 0 \). That is, all customers encountered by the disadvantaged worker are neutral. Therefore, the rating for the first period is \( R^D_1 = r^D_1 = 1 + q \).

Starting from the second period, the advantaged worker encounters \( \alpha \) partial customers and \( 1 - \alpha \) neutral customers. The cumulative rating for the disadvantaged worker in this case is:

\[
R^D_t = \frac{1}{(1 - \delta)^\alpha} ((1 - \delta)^\alpha (1 + q)) \sum_{i=1}^t \frac{1}{\delta^i}
\]

If \( q > 0 \) and \( (1 - \delta)^\alpha (1 + q) < 1 \), then in the initial periods, the advantaged worker has a higher rating. However, as the disadvantaged worker becomes exposed to partial customers over time, they eventually receive a lower rating compared to the advantaged worker. Fig. 3 presents the pattern.

**Figure 3**  Illustration of the Impact of Initial Conditions on Ratings Gap

![Diagram showing the impact of initial conditions on ratings gap.](image)

*Note:* \( q = 0.15, \delta = 0.4 \) and \( \alpha = 0.4 \).
4. Empirical Setting and Descriptive Analysis

We first describe the empirical setting of the labor market platform. We then provide model-free evidence in support of the subsequent theory and empirical models.

4.1. The Online Labor Market Platform

Our empirical setting is a North American online labor market platform. We use a pseudonym, ServicesConnect (“SC” for short) to refer to the platform as the platform wishes to remain anonymous. The platform connects customers and workers for a range of home-service jobs and manages the entire transaction process.

Workers on SC are mostly small-business owners who specializes in certain types of jobs. Before taking any jobs on SC, workers must go through SC’s screening process, which includes skill verification and a criminal background check. Moreover, customers are informed that workers must undergo the screening process to take any jobs on SC. Therefore, this verification process should alleviate concerns regarding significant discrepancies in the quality of workers, both for customers and the researchers.

Services provided by the workers on SC typically require only a few hours of work. Examples of the services include appliance, electrical work, maintenance services, and plumbing. Once a customer submits a service request by selecting from a predetermined list of service categories, a minimum cost and an hourly rate are generated. The costs are uniform within the service category, do not differ by worker, and are non-negotiable for both the customer and the worker. The customer is then directed to choose from available time slots for the job to be completed.

After a job is submitted, SC employs an algorithm to allocate workers to customers. The algorithm selects a small set of workers based on the number of jobs the worker has completed and the worker’s average rating. The small set of workers is then privileged to accept the job within a 15-minute window on a first-come, first-served basis. If the job is not accepted within the time frame, it becomes open to all workers eligible for the service category, again on a first-come, first-served basis. Importantly, the algorithm is blind to worker demographics and customer characteristics,
Can customer ratings be discrimination amplifiers?

as they are not inputs. Since the customer has no control over this paring process, the pairing between the workers and the customers can be considered as good as random with regard to the relationship between perceived worker race and customer characteristics.

Once a worker claims a job, the customer receives a message from SC, showing the worker’s name, photo, average rating to date, and number of jobs completed to date. The average rating to date is only available to customers after the worker has received five ratings. A customer can cancel the job at any time before the worker indicates that they are on their way to complete the job. After a job is completed, customers are sent another email asking them to rate their experience with the worker from 1 to 5 stars.

We obtain data on all jobs from one metro area in North America during the four-year period from 2016-19 for the purposes of this analysis. We deleted the 11 female workers in the data as the platform workers are primarily male. Additionally, given that the platform does not collect workers’ demographic information, we use the worker’s photo to code a customer’s perception of race. Hence, we excluded jobs where the worker’s photo was not available. The data consists of 86157 jobs requested by 34110 unique customers and assigned to 633 workers.

SC categorizes service jobs into several categories, including maintenance, plumbing, appliance, electrical etc. Among the service categories, most can be classified by whether the jobs require worker credentials. For example, electrical work and HVAC requires credentials, while snow removal does not. However, appliance and electrical are the two exceptions as some jobs require credentials, while others do not. We split appliance and electrical each into two categories, based on whether credentials are needed. The most frequently requested service category is maintenance, followed by plumbing and appliance. The cancellation rates and the average ratings vary across service categories. Urgent service categories such as locksmith and appliance have the lowest cancellation

---

8 SC does not collect worker or customer demographics.

9 Fig. A.1 in Appendix 8 shows the two messages sent to the customers.

10 Although we observe service category in the data, due to NDA, we cannot provide the summary statistics such as job proportion, cancellation rates, and average ratings by service category.
Can customer ratings be discrimination amplifiers?

rates; electrical jobs and upholstery jobs have the highest cancellation rates. The average rating within category also varies, ranging from 4.57 to 4.87 across categories. As such, we control for the service category fixed effects in our empirical model.

We also observe changes in job count, cancellations and ratings across year. The number of jobs on the platform increased over the years from 2016 to 2019. As the platform grew, the cancellation rates increased as well, perhaps because the early adopters on the platform were more risk-tolerant. On the other hand, the average rating increases over the year, perhaps because the more capable workers stay longer on the platform. This suggests accounting for year fixed effects in our model.

4.2. Job Outcome and Worker Race

Given that SC does not collect worker’s demographic information, to understand the relationship between the job outcomes and the customers’ perception of worker race, we had two coders review the workers’ profile pictures and note their perception of worker race using the pre-defined categories. More details about the coding can be found in Botelho and DeCelles (2023). Given that the majority of workers were coded to be White and our theory is not about a particular minority race but about whether the workers are perceived to be minority or not, we define a worker to be a minority if the worker is not perceived as White by our coders, and White if otherwise.

Table 4 presents the average job outcomes by worker race. The summary statistics are suggestive of minority workers being disadvantaged in both cancellation and rating processes. Overall, minority worker have a higher proportion of jobs canceled. Once a job is finished, there is no significant difference in the rating submission rate between the minority and White workers. However, conditional on rating submission, minority workers receive lower ratings in average. Besides, in most cases, most customers give a 5-star rating, and the ratings below 4 are scarce. This motivates us to model the rating behavior as a binary choice between 5-star and sub-5 star ratings.

4.3. Job Outcomes and the Displayed Rating

Table 4 seems to be suggestive of taste-based discrimination, where minority workers are faced with higher cancellation and lower rating. On the other hand, given that the platform displays the
Can customer ratings be discrimination amplifiers?

Table 4  Job Outcome and Perceived Worker Race

<table>
<thead>
<tr>
<th></th>
<th>All Jobs</th>
<th>White</th>
<th>Minority</th>
<th>Minority Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Workers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cancellation Rate</td>
<td>18.93%</td>
<td>17.60%</td>
<td>21.70%</td>
<td>-4.10%</td>
</tr>
<tr>
<td>Rating Submission Rate</td>
<td>69.06%</td>
<td>69.11%</td>
<td>68.97%</td>
<td>0.14%</td>
</tr>
<tr>
<td>Ratings:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg. Rating</td>
<td>4.76</td>
<td>4.78</td>
<td>4.71</td>
<td>0.07</td>
</tr>
<tr>
<td>Rating=5</td>
<td>85.20%</td>
<td>86.26%</td>
<td>82.88%</td>
<td>3.38%</td>
</tr>
<tr>
<td>% for Ratings below 5:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rating=4</td>
<td>9.58%</td>
<td>9.09%</td>
<td>10.65%</td>
<td>-1.56%</td>
</tr>
<tr>
<td>Rating below 4</td>
<td>5.22%</td>
<td>4.65%</td>
<td>6.46%</td>
<td>-1.82%</td>
</tr>
</tbody>
</table>

Table 5 shows how job outcomes are correlated with the positive ratings. First, job cancellation rate is negatively associated with displayed rating, which suggests that a higher displayed rating could potentially signal worker's quality and lead to lower cancellation. However, for any given range of displayed ratings, minority workers have higher cancellation rate than non-minorities. Thus, even though positive rating may resolve some uncertainty, customers do not give the same credit to minorities and non-minorities for their displayed ratings. Second, job rating is positively

$$pos\_rating = \mathbb{1}_{\{avg\_rating \geq 4.5\}} \times (avg\_rating - 4.5)/(5 - 4.5)$$

11 The average rating is masked if the worker has obtained less than 5 ratings by the time the job is assigned. Once the worker gets at least 5 ratings, the average rating is available to the customer when the job is assigned.

12 This formulation helps account for the the fact that consumers are not very sensitive to ratings below a threshold. We did not find our results to be very sensitive to this cutoff; we choose the 4.5 threshold as it provides an easy to interpret metric of rating from 0 to 1.
Can customer ratings be discrimination amplifiers?

associated with higher displayed rating. This could potentially be driven by two mechanisms. First
the higher displayed rating captures higher worker quality and job quality, which in turns leads
to higher job rating. Or it could also be that customers ratings are impacted by the ratings of
others such that the job rating is higher when the observed worker rating is higher (consistent with
the claim in Park et al. 2021). Regardless of the mechanism, the evidence above suggests that on
average, minorities received lower ratings conditional on their displayed rating.

<table>
<thead>
<tr>
<th>pos_rating range</th>
<th>Cancellation Rate</th>
<th>Average Job Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All Jobs</td>
<td>White</td>
</tr>
<tr>
<td>(4.75, 5]</td>
<td>17.48%</td>
<td>16.69%</td>
</tr>
<tr>
<td>(4.5, 4.75]</td>
<td>19.76%</td>
<td>18.37%</td>
</tr>
<tr>
<td>\leq 4.5</td>
<td>20.98%</td>
<td>18.93%</td>
</tr>
</tbody>
</table>

4.4. Expansion in Minority Gap

Section 4.2 shows the overall minority gap in both cancellation and ratings. In this section, we
explore whether the gap expands across time and discuss potential explanation behind the expan-
sion. To get some insight into how displayed ratings can impact the ratings gap, we exploit the
fact that the platform does not display average ratings until a worker has accumulated five ratings.
Fig. 4a shows the percentage of completed jobs with a non 5-star rating for minority and White
workers when they receive 5 ratings (when the average ratings are not displayed), 10 ratings, 15
ratings, and up until 40 ratings. The graph shows a significant increase in the rating gap between
minority and White workers when ratings are displayed versus not. Remarkably, the ratings for
White workers remain flat, but the percentage of sub-5 star ratings for minority workers increases
over time after ratings are displayed. The ratings between minority and White workers are not sig-
nificantly different in the initial periods. This suggests that displaying ratings does have an impact
beyond worker ability, but also amplifies any potential differential impact between minorities and
non-minorities over time.

Fig. 4b shows the pattern for a subset of workers – workers who receive all 5-stars in the first
5 ratings. Though these workers are more likely to be of higher quality than those who did not
receive the 5-star average for the first five ratings, the minority rating gap still arises and expands over time as the minority worker receives lower ratings in later periods. This suggests that even if starting from similar displayed ratings, the impact of bias from later can impact their ratings and this impact is amplified over time.

Figure 4  Rating Gap between Minority and White Workers

(a) All Workers

(b) Workers with All 5-star before Display

Note: The left plot shows the rating trend for all workers. The right plot shows rating trend for workers who receive all 5-star ratings in the first 5 ratings before ratings are displayed. We select the workers that have received at least 20 ratings in total. The pattern is robust if we plot using the workers who have received at least 40 ratings.

5. Model and Estimation

In this section, we bring the intuition from the analytical model and the model-free evidence to assess its implications by calibrating an empirical model of customer behavior. Our focus is on customer behaviors after the platform’s job assignment. Specifically, we model three customer behaviors: whether to cancel the job, whether to submit a rating, and what rating to submit. After describing the empirical model, we describe our estimation approach. Finally, we clarify the features of the data that help identification of the model’s key parameters.

5.1. The Empirical Model

The model consists of three parts, related to the three customer behaviors we model. The first part models the binary decision of whether to cancel the job after the job is paired with a worker. The second part captures the binary decision to submit a rating to the platform once the job is completed. The third part models the rating of the job given that the customer decides to submit a rating. The model allows for unobserved heterogeneity as latent segments across customers.
Let \( i \) denote the customer. Each customer belongs to a segment \( g \in \{1,...,G\} \). We denote a job with \( j \) and a worker with \( k \). Suppose a customer \( i \) requests a job \( j \), and the job is assigned to worker \( k \). First, we model the decision of job cancellation as a binary logit. \( C_{ij}^{g} \) takes the value of 1 if the customer cancels the job \( j \) conditional on worker \( k \) accepting the job. Let \( X_{ij}^{c} \) denote the observables that impact the decision of whether to cancel the job. We formulate the cancellation decision as a binary logit for each customer segment \( g \) as:

\[
P_{ij}^{cg} = P^{g}(C_{ij} = 1) = \frac{\exp(\beta_{g}^{c}X_{ij}^{c})}{1 + \exp(\beta_{g}^{c}X_{ij}^{c})},
\]

where \( \beta_{g} \) denotes the parameters for segment \( g \).

For each job \( j \) that is not canceled and thus completed, let \( W_{ij} \in \{0,1\} \) denote whether the customer submits the rating to the platform. \( W_{ij} \) is modeled as a binary logit model. For a customer that belongs to the segment \( g \), the probability of submitting a rating given the job is not cancelled is:

\[
P_{ij}^{wg} = P^{g}(W_{ij} = 1|C_{ij} = 0) = \frac{\exp(\delta_{g}^{w}X_{ij}^{w})}{1 + \exp(\delta_{g}^{w}X_{ij}^{w})},
\]

where \( X_{ij}^{w} \) are the observables and \( \delta_{g} \) denotes the parameters for segment \( g \).

Finally, conditional on a rating being given, we model the rating choice as a binary outcome: \( R_{ij} \) takes the value of 1 if the customer gives a rating of 5-star.\(^{13}\) For a customer that belongs to segment \( g \), the probability of giving a 5-star rating given a rating is submitted is

\[
P_{ij}^{rg} = P^{g}(R_{ij} = 1|W_{ij} = 1) = \frac{\exp(\delta_{g}^{r}X_{ij}^{r})}{1 + \exp(\gamma_{g}^{r}X_{ij}^{r})},
\]

where \( X_{ij}^{r} \) are the observables and \( \delta_{g} \) denotes the parameters for segment \( g \).

Let \( q^{g} \) denote the share of consumers that belong to segment \( g \), where \( \sum_{g} q^{g} = 1 \). Then the parameters to be estimated are \( \Theta = \{\Theta_{1},...\Theta_{G}\} \), where \( \Theta_{g} = \{\beta_{g}^{r},\delta_{g}^{r},\gamma_{g}^{r},q^{g}\} \). Let \( J(i) \) denote the set of jobs requested by customer \( i \) and \( S_{i} = \{C_{ij},W_{ij},R_{ij}\}_{j \in J(i)} \) denote the set of observations.

\(^{13}\)We considered a model with rating choice of 1-5 as an ordinal logit model, but since most ratings are 5 or 4, the ordinal logit did not add much incremental value in terms of fit. Hence we model rating choice as a binary choice.
of customer $i$. The likelihood function of an individual customer $i$ that belongs to segment $g$ is given by

$$L^g_i(S_i; \Theta_g) = \left[ \prod_{j \in J(i): C_{ijk}=1} (P_{ij}^g)^{C_{ijk}} \right] \times \left[ \prod_{j \in J(i): C_{ijk}=1, W_{ijk}=0} (1 - P_{ij}^g (1 - P_{ij}^g)) \right] \times \left[ \prod_{j \in J(i): C_{ijk}=1, W_{ijk}=1} (1 - P_{ij}^g (W_{ijk} P_{ij}^g P_{ij}^R R_{ijk} (1 - R_{ijk}))) \right],$$

where the first component accounts for jobs being assigned to a worker and then cancelled by customer $i$, the second component accounts for jobs not canceled and not rated, and the third component accounts for jobs not canceled and rated. By summing over all of the unobserved segments $g \in \{1, ..., G\}$, we obtain the overall likelihood of customer $i$:

$$L_i(S_i; \Theta) = \sum_g q_g L^g_i(S_i; \Theta_g)$$

Then the log-likelihood over all customers is given by

$$\sum_i log(L_i(S_i; \Theta)) = \sum_i log(\sum_g q_g L^g_i(S_i; \Theta_g))$$

5.2. Estimation

We estimate the model using the EM algorithm that iteratively maximizes the expected log-likelihood in Equation

$$\sum_i \sum_g q^g_i log(L^g_i(S_i; \Theta_g)),$$

where $q^g_i$ is defined below as the probability that customer $i$ is of segment $g$ given parameters values $\Theta$, conditional on all of the observed jobs of customer $i$:

$$q^g_i = Pr(g|S_i; \Theta) = \frac{L^g_i(S_i; \Theta_g)}{L_i(S_i; \Theta)}$$

The EM algorithm is implemented in the following steps:

1. Initialize $\Theta_0$. We initialize $q^g = 1/G$. Then we randomly partition the customers into $G$ segments and maximize the three components of the log-likelihood for each segment to get the initial values for $\Theta_g$. 
2. For each customer and each segment, calculate the probability of being in the segment conditional on the customer’s cancellation and rating choices given $\Theta$:

$$q_i^g = P(i \in g|S_i) = \frac{q^g L_i^g(S_i; \Theta_g)}{\sum_{g'} q^{g'} L_i^{g'}(S_i; \Theta_{g'})} \tag{9}$$

3. Update $q^g = \frac{\sum_i q_i^g}{\sum_i \sum_{g'} q_i^{g'}}$.

4. Update the coefficients $\Theta_g$ for each segment $g$ by maximizing the segment likelihood with the updated $q_i^g$ in step 2: $\sum_i q_i^g \log(L_i^g(S_i; \Theta_g))$.

5. Repeat steps 2-4 until convergence.

5.3. Identification

Our empirical strategy to identify discrimination by customer segment is based on how customer cancellation and rating behavior is related to the effect of disadvantaged group membership of the worker. We are able to identify segments because we have panel data on each customer’s jobs and on worker performance at the individual level. Further, since the assignment of workers to jobs (customers) is unrelated to customer characteristics and worker race, this allows us to treat these characteristics as randomly assigned.

At the cancellation stage—after job assignment—the customer observes worker characteristics (number of past jobs, ratings, and perceived race of the worker) and job characteristics (e.g., whether the job requires worker’s credential) and then makes the decision on whether to cancel or not. To the extent that we observe multiple jobs from the same customer, it is possible to identify how customer cancellation behaviors differ by worker and job characteristics, and in particular, the focal variable of interest—whether the worker is a minority. If indeed there is a direct negative main effect of being a minority or a negative interaction of minority with past rating on cancellation, we treat that as evidence of taste-based or statistical discrimination.

At the rating stage, customers also observe the actual performance of the worker; but this is unobserved to researchers. To the extent that there is randomness in worker assignment to customers, worker performance is a random draw around a fixed mean performance. A customer’s
rating for a job will be correlated with the cumulative average of past ratings due to the common effect of skills on rating across time. However, if the strength of this relationship between current and past ratings is discounted for minorities, then we interpret this as evidence of either taste-based or statistical discrimination. In addition, if there is a direct main effect of minority on current ratings that is also evidence of taste or statistical discrimination.

Given our interest in spillovers across customer segments, we allow for unobserved heterogeneity in customers’ cancellation and rating behavior. We leverage the two-sided panel data on workers and customers to identify customer heterogeneity. Specifically, workers perform many jobs and are assigned to different customers; and there is randomness in the assignment of workers to customers—and their ratings and number of jobs evolve over time. Each customer is exposed to both minority and White workers, as well as varying customer characteristics. To the extent that different customer segments show different levels of discrimination—or no discrimination at all—in terms of the differential impact of perceived race on how past ratings impact cancellation and current ratings, and they all are impacted by overall past rating (even with no differences by race), we are able to quantify the spillovers on ratings and job cancellations through past ratings. We quantify the magnitude of the spillovers by simulating how rating gap between minority and White workers will be greater in the presence of rating display.

6. **Empirical Analysis**

We first provide some additional description of the data that we use to estimate the model. We then report the model estimates – showing evidence of discrimination among some latent segments and none in others. Finally, we describe the characteristics of the latent segments to facilitate interpretation.

6.1. **Model Estimates**

Given that there are differences in average cancellation rates and ratings for different service categories, we control for service fixed effects in the model. As such, we focus on the service categories that account for at least 1.5% of the jobs so that we can reliably estimate fixed effects
for all the equations. As discussed earlier, to identify customer’s unobserved heterogeneity in their cancellation and rating behaviors, we need multiple jobs from each customer in the data set, and ideally each customer should have been assigned to both White and minority workers. We therefore include customers who have offered at least 4 jobs in our analysis. After these deletions, our sample for model estimation includes 36215 jobs requested by 5113 customers and accepted by 571 workers.

With customers with at least 4 job requests in our estimation sample, the data sample preserves a mixing of minority workers and White workers within each individual customers. Fig. 5 presents the pattern. For each customer, we count the total number of jobs requested and specify whether the customer has been assigned to only minority workers, only to White workers, or to both types of workers. Among the customers who have requested 4 jobs in the data, 78.45% have been assigned to both types of workers. The ratio goes up to 85.61% for customers who have requested 5 jobs and increases with the customer job count. The plot shows the exposure to both minority and White workers at customer-level.

![Figure 5: Customer Distribution by Number of Jobs Requested and Worker Composition](image)

Table 6 shows the estimates for the consumer model of choice of cancellations, rating submission, and rating choice. We begin with a brief discussion of the explanatory variables in the model. Apart from the fixed effects of service categories and years, we include variables that capture the worker’s past history, including the total job count and the rating displayed. We also include a dummy variable of no rating to indicate if the worker has received less than 5 ratings and the rating is
Can customer ratings be discrimination amplifiers?

Given our primary interest in how consumers differentially make choices for minorities, we include a minority main effect and its interaction with the measure of positive rating, as defined in Section 4.3.

We estimate the model with two, three, and four customer segments respectively. Based on the highest log-likelihood and lowest AIC, we choose the three segment model as the best fitting model. We adopt the three-segment model for our main empirical analysis.

For all three segments, the positive rating variable has a negative sign in the cancellation equation and a positive sign in the rating equation. This indicates that displayed rating, in general, reduces cancellation and is positively correlated with the current job rating, which is also consistent with our theory model. A larger worker job count is negatively correlated with cancellation for segment 1 and 3 and is positive correlated with job rating for segment 1.

For segment 1 (19% of customers), none of the minority coefficients or minority interactions are significantly different from zero. Any differences in their behaviors between minorities and non-minorities is only through observable variables. Hence this segment does not engage in either taste or belief-based (i.e., statistical) discrimination between minorities and Whites. We call this the “unbiased” (or no-discrimination) segment.

For segment 2 (24% of customers), in the cancellation model, the minority coefficient is not significant, while its interaction term with the positive rating is significant. That is, while positive rating reduces cancellation rate for White workers, customers in segment 2 do not give as much credit to minority workers based on the displayed rating. This reflects the $\gamma$ in the analytical model. On the other hand, the minority variable and its interaction term are not significant in the rating equation for this segment. We call this the minority “avoider” segment.

---

14 We have also considered incorporating the interaction of no rating and the minority indicator. However, the parameter of the interaction term was found to be insignificant. Hence, we have opted not to include it in the model.

15 The log-likelihood for two- and three-segment models are -39768 and -39222 respectively. The corresponding AIC values are 79791 and 78826. The four-segment model does not converge due to over-parameterization.
Table 6  Consumer Model Parameter Estimates

<table>
<thead>
<tr>
<th></th>
<th>Segment 1: Neutral Customers</th>
<th>Segment 2: Minority Avoiders</th>
<th>Segment 3: Minority Avoiders &amp; Under-Raters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Segment Probability</td>
<td>0.19</td>
<td>0.24</td>
<td>0.57</td>
</tr>
<tr>
<td>SE</td>
<td>0.02</td>
<td>0.04</td>
<td>0.02</td>
</tr>
<tr>
<td><strong>Cancellation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>logWorkerJobCount</td>
<td>-0.08**</td>
<td>-0.02</td>
<td>-0.06**</td>
</tr>
<tr>
<td>minority</td>
<td>0.14</td>
<td>0.03</td>
<td>0.17*</td>
</tr>
<tr>
<td>noRating</td>
<td>-0.42</td>
<td>-0.05</td>
<td>-0.48**</td>
</tr>
<tr>
<td>posRating</td>
<td>-0.70***</td>
<td>-0.58***</td>
<td>-0.58**</td>
</tr>
<tr>
<td>minority*posRating</td>
<td>0.20</td>
<td>0.61*</td>
<td>-0.17</td>
</tr>
<tr>
<td>Service FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Year FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Submit Rating</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>logWorkerJobCount</td>
<td>-0.13***</td>
<td>-0.07**</td>
<td>-0.04</td>
</tr>
<tr>
<td>minority</td>
<td>0.27</td>
<td>0.14</td>
<td>0.11</td>
</tr>
<tr>
<td>noRating</td>
<td>-0.72*</td>
<td>0.24</td>
<td>0.71**</td>
</tr>
<tr>
<td>posRating</td>
<td>0.02</td>
<td>0.47**</td>
<td>0.65**</td>
</tr>
<tr>
<td>minority*posRating</td>
<td>-0.70</td>
<td>-0.18</td>
<td>-0.18</td>
</tr>
<tr>
<td>Service FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Year FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Rating</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>logWorkerJobCount</td>
<td>0.07*</td>
<td>-0.06</td>
<td>0.04</td>
</tr>
<tr>
<td>minority</td>
<td>-0.12</td>
<td>-0.29</td>
<td>0.12</td>
</tr>
<tr>
<td>noRating</td>
<td>1.20***</td>
<td>0.76</td>
<td>0.35</td>
</tr>
<tr>
<td>posRating</td>
<td>1.16***</td>
<td>1.04*</td>
<td>1.23**</td>
</tr>
<tr>
<td>minority*posRating</td>
<td>0.05</td>
<td>0.17</td>
<td>-0.66**</td>
</tr>
<tr>
<td>Service FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Year FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

N: 36215
log-likelihood: -39222
AIC: 78826

Finally, segment 3 (57% of the customers), discriminates in both the cancellation stage and the rating stage. In the cancellation stage, they are more likely to cancel jobs assigned to minorities regardless of the displayed rating. In the rating stage, compared to consumers in segments 1 and 2, they rate minorities with the same average level of past ratings systematically lower than non-minorities. This reflects the $\delta$ in the analytical model. We call this the “minority avoider and under-rater” segment.

6.2. Latent Segment Characteristics

Table 7 shows the descriptive statistics of cancellation and ratings by customer segment. First, we find that percentage of jobs needing credentials are roughly equal across all three segments—
suggesting no significant differences in the types of jobs requested by these segments. In the cancellation stage, the difference between the minority and White workers is largest for Segment 2—the minority avoider segment. We also find a smaller gap for Segment 3—the minority avoider and under-rater segment. Consistent with our model estimates and our data descriptives, there is no significant difference in rating submission rate between minority workers and White workers for all three segments. Finally, all three segments display rating differences between minorities and non-minorities, and the difference is largest for the neutral segment. This is suggestive of the spillover of rating discrimination to the neutral segment—and the large gap arising from the fact that it has lower baseline of ratings, but is sensitive to positive ratings (which are biased by the minority underrated third segment).

<table>
<thead>
<tr>
<th>Table 7</th>
<th>Customer Segments Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Segment 1</td>
</tr>
<tr>
<td></td>
<td>Neutral Customers</td>
</tr>
<tr>
<td>total job count</td>
<td>5845</td>
</tr>
<tr>
<td>% jobs need credential</td>
<td>40.51%</td>
</tr>
<tr>
<td>Cancellation Rate</td>
<td></td>
</tr>
<tr>
<td>non-minority</td>
<td>18.43%</td>
</tr>
<tr>
<td>minority</td>
<td>21.88%</td>
</tr>
<tr>
<td>Rating Submission Rate</td>
<td></td>
</tr>
<tr>
<td>non-minority</td>
<td>81.17%</td>
</tr>
<tr>
<td>minority</td>
<td>79.60%</td>
</tr>
<tr>
<td>Perc. Rating=5</td>
<td></td>
</tr>
<tr>
<td>non-minority</td>
<td>48.84%</td>
</tr>
<tr>
<td>minority</td>
<td>42.10%</td>
</tr>
</tbody>
</table>

7. Counterfactuals

Using the model estimates, we conduct counterfactual analysis to answer our second and third research questions on (i) to what extents customer discrimination leads to the minority rating and earnings gap; and (ii) how displaying ratings impacts the ratings and earnings gap.

7.1. Quantify the impact of customer discrimination on the minority gap

To study the effect of customer discrimination on minority ratings and earnings gap, we compare the ratings and earnings between a minority worker and a White worker for a sequence of jobs
respectively within each service category. More specifically, we draw a sequence of 300 jobs, with each job randomly being offered by one of the three estimated consumer segments, consistent with the estimated empirical share. For each service category, we apply the corresponding category fixed effect in the estimated equations so as to match with the baseline rates of cancellations, rating submission and ratings. Overall, these choices for the simulation imply that the jobs are all homogeneous except for the customer segment.

We simulate the cancellations, ratings submission and ratings over the 300 jobs first with a White worker assignment and then with a minority worker assignment. For each worker, we repeat the simulation 10 times and take the averages of the ratings and job counts across the 10 sequences of jobs. The rating and job counts for the White and minority worker and the earnings gap for all service categories are reported in Table 8.

<table>
<thead>
<tr>
<th>Service Category</th>
<th>% sub-5 star Rating</th>
<th>Job Count (Earnings)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>White</td>
<td>Minority</td>
</tr>
<tr>
<td>Maintenance</td>
<td>9.0</td>
<td>11.3</td>
</tr>
<tr>
<td>Plumbing</td>
<td>10.8</td>
<td>12.8</td>
</tr>
<tr>
<td>Appliance (Cred.)</td>
<td>11.3</td>
<td>13.0</td>
</tr>
<tr>
<td>Landscaping</td>
<td>16.3</td>
<td>19.5</td>
</tr>
<tr>
<td>Electrical</td>
<td>10.1</td>
<td>12.6</td>
</tr>
<tr>
<td>HVAC</td>
<td>10.7</td>
<td>12.8</td>
</tr>
<tr>
<td>Gutters</td>
<td>11.5</td>
<td>14.9</td>
</tr>
<tr>
<td>Snow</td>
<td>14.0</td>
<td>17.6</td>
</tr>
<tr>
<td>Moving</td>
<td>9.2</td>
<td>11.3</td>
</tr>
<tr>
<td>Upholstery</td>
<td>16.5</td>
<td>19.8</td>
</tr>
<tr>
<td>Appliance (Non Cred.)</td>
<td>6.5</td>
<td>9.6</td>
</tr>
<tr>
<td>Locksmith</td>
<td>8.4</td>
<td>11.1</td>
</tr>
<tr>
<td>Misc.</td>
<td>12.1</td>
<td>16.6</td>
</tr>
</tbody>
</table>

Note: The service categories are listed in descending order by the empirical share in the data.

We report the percentage of sub-5 star rating for the White worker and the minority worker separately as well as the percentage gap between them. The earnings gap is calculated as the 16\footnote{We use the 2019 year fixed effect; hence the estimates should be interpreted as the effect for 2019—the last year of data in our sample. Besides, the coefficient of the logWorkerJobCount variable for segment 2 in the rating equation is small and insignificant, but has a negative sign (opposite of what is theoretically expected). We set this insignificant coefficient to zero to avoid any contamination of the simulated results. Similarly, we set the insignificant coefficient for all minority-related variables to 0.}
percentage difference in total job count for the White worker compared to the minority worker.
In all service categories, the White worker receives both a higher average rating and more jobs.
The differences between the minority and White worker vary across service categories. The overall
ratings gap, which is the average of service-specific earnings gap, weighted by the empirical share
of jobs, is 23%. The overall minority earnings gap is 4%.

7.2. Quantify the impact of customer ratings display on the minority gaps

Here we compare the ratings and earning differences between minority and White workers for a
sequence of jobs when (i) ratings are displayed versus (ii) ratings are not displayed. For the first
case, the simulation is identical to the previous counterfactual. For the second case, the simulation
steps are largely similar but with one difference. When the rating is not displayed, starting from the
6th rating, we need an assumption about what the consumer will impute. To make the imputations
comparable between no-display and the early stages of display, where the platform does not display
ratings, we assume that customers impute the average rating across all jobs in the corresponding
service category in the no-display case.

Table 9 reports the gaps with and without displayed ratings by service category. The second and
the 5th columns show ratings and earnings gaps when ratings are displayed, where the numbers
are identical to those in Table 8. The 3rd and 6th columns show the gaps when ratings are not
displayed, and the 4th and 7th columns show the percentage gap closed when ratings are not
displayed compared to the case when ratings are displayed. Overall, weighted by the proportion of
jobs, not displaying ratings closes 50% of the ratings gap and 18% of the earnings gap. Equivalently,
this suggests that displaying ratings increases 22% of the ratings gap and 115% of the earnings
gap. This shows the discrimination amplification effect of displaying ratings.

8. Conclusion

The growth of online e-commerce and service platforms has made customer rating systems ubi-
quitous. Consumers rely on this feedback to make choices on these platforms, while platforms use
it as means of evaluation and a tool for management of suppliers and service providers. While the
Can customer ratings be discrimination amplifiers?

<table>
<thead>
<tr>
<th>Service Category</th>
<th>% Ratings Gap</th>
<th>% Earnings Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Display</td>
<td>No Display</td>
</tr>
<tr>
<td>Maintenance</td>
<td>27</td>
<td>12</td>
</tr>
<tr>
<td>Plumbing</td>
<td>18</td>
<td>12</td>
</tr>
<tr>
<td>Appliance (Cred.)</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>Landscaping</td>
<td>20</td>
<td>7</td>
</tr>
<tr>
<td>Electrical</td>
<td>24</td>
<td>11</td>
</tr>
<tr>
<td>HVAC</td>
<td>20</td>
<td>12</td>
</tr>
<tr>
<td>Gutters</td>
<td>29</td>
<td>11</td>
</tr>
<tr>
<td>Snow</td>
<td>26</td>
<td>7</td>
</tr>
<tr>
<td>Moving</td>
<td>23</td>
<td>14</td>
</tr>
<tr>
<td>Upholstery</td>
<td>20</td>
<td>9</td>
</tr>
<tr>
<td>Appliance (Non Cred.)</td>
<td>48</td>
<td>14</td>
</tr>
<tr>
<td>Locksmith</td>
<td>32</td>
<td>20</td>
</tr>
<tr>
<td>Misc.</td>
<td>37</td>
<td>12</td>
</tr>
</tbody>
</table>

Note: The service categories are listed in descending order by the empirical share in the data.

conventional wisdom is that these systems can serve to reduce discrimination of minority workers and firms due to the availability of individual level ratings, this paper introduces a counter perspective. The paper proposes that when customer ratings embed the effects of minority discrimination by even a segment of customers, the display of average ratings can amplify discrimination and create discrimination spillovers even among customers who do not discriminate—as the rating now memorializes rating differences due to tastes as differences in “quality” for all future customers. We formalized the idea with a stylistic analytical model, and then empirically investigated the issue in the context of an online labor market platform. We find three segments of customers—a neutral segment (that does not discriminate), a minority avoider segment (that discriminates on cancellations), and a minority avoider and under-rater segment (that discriminates on cancellations and ratings). Overall, we find that customer discrimination contributes 23% to the minority ratings gap, and 4% to the earnings gap. Further, as conjectured, the display of customer ratings causes discrimination spillovers to the neutral non-discriminating group and amplifies discrimination among the other two discriminating groups. Overall, displaying reviews amplifies the discrimination gap; it increase the minority ratings gap by 22% and the earnings gap on the platform by 115%. We note that though the earnings gap of 4% may seem small, to the extent that these types of workers
live at close to subsistence levels, the cumulative impact of the 4% earnings gap in terms of savings and wealth gaps will be much larger.

We conclude with a discussion of limitations and suggestions for future work. Clearly, our work is in the context of one online labor market platform, it would be useful to assess the generalizability and limits of our conjecture on discrimination amplification by studying not only other labor market platforms. It would be interesting to know how the discrimination and impact of the earnings gap associated with displaying ratings may be moderated by particular design features of the platform. For example, customers are anonymous on the platform, that may lead to greater discrimination. On the other hand, customers do not choose from a set of providers, and that leads to fewer opportunities to discriminate. Here work occurs in the home, compared to many services that are done at the place of business. Hence sensitivity to get work done by outgroup workers may be greater in this context than others. Understanding how these design factors impact discrimination and the earnings gap would lead to greater insight. Finally, our research can be extended to e-commerce platforms. Past research has shown evidence of racial and gender discrimination in e-commerce platforms such as eBay (e.g., Ayres et al. 2015, Kricheli-Katz and Regev 2016), so the replicability of discrimination spillover and amplification is relevant in such contexts.

In general, the discrimination literature has focused on average effects, and typically has not considered unobserved heterogeneity. Accounting for unobserved heterogeneity should provide a richer description of discrimination. This is relevant not just in market/marketing settings, but also in settings such as education and criminal justice, where there is much work on bias and discrimination — and there is likely heterogeneity among teachers and police officers in whether they discriminate and the magnitude of discrimination. Not only does accounting for unobserved heterogeneity provide an accurate description, it can also lead to greater acceptance by society of the research findings as it fits the lay notion that not all people discriminate, and equally. In particular, we hope that our work inspires the literature on fairness and biases in machine learning to account for unobserved heterogeneity, and the impacts of cross-segment spillovers and amplification.
Can customer ratings be discrimination amplifiers?

Our modeling approach is also a useful lens to study questions of structural inequity in settings such as education and criminal justice. For example, the approach would be useful to study and quantify how biases/discrimination in early disciplinary actions in school or interactions with the criminal justice system can lead to a “record” that often get used to justify tougher actions against minorities, leading to worse life outcomes. More generally, we hope our case based modeling approach allows managers and scholars not only to measure the presence of biases/discrimination in a particular setting, but also help to quantify how such early discriminatory outcomes, when translated into merit, leads to structural inequity over time—and assessing ways in which such inequity may be mitigated.

Funding and Competing Interests

All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript. The authors have no funding to report.
Can customer ratings be discrimination amplifiers?

References


Appendix

Communications with Customers

Figure A.1 Screenshots of Emails

(a) Confirmation Email after Worker Accepts Job

Hi John,

Meet Marcel, your ServicesConnect Plumbing contractor!

If this is work that will be done inside, please make sure that an adult is home during the service. If it is outside, you do not need to be home.

Your contractor will be contacting you soon to discuss the details of the work. Also, please note that depending on the weather, you might need to reschedule this job.

Marcel will be there on:

WEDNESDAY, OCTOBER 6, 2021 / 1:00 PM - 2:00 PM

Marcel
SC CERTIFIED PLUMBING CONTRACTOR
4.75
(20 JOBS)

If you need to adjust the timing or have questions, please contact Marcel or use the links provided, or use the help desk.

(b) Rating Reminder Email after Job is Completed

Hey John,

Your Plumbing job is complete. Wasn’t that simple? Your receipt is included below.

We are proud to work with our terrific contractors. Please take a second to rate them below!

If you have questions or complaints, please visit our help desk or reply to this email.