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 be "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 matches service workers with customer jobs. Using a model of customer's job cancellations and rating choice 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 the non 5-star ratings for the average minority worker by 24% and decreases the earnings by 4%. Displaying ratings amplifies discrimination and increases the minority rating gap by 80% and the minority earnings gap by 28% 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 growth of online commerce platforms (e.g., Amazon, eBay, Taobao) and digitally mediated marketplaces (e.g., Airbnb, Uber, TaskRabbit) has provided customers with no shortage of options to leave feedback; a five-star or thumbs up/thumbs down evaluation system is available on nearly every type of transaction. Customers increasingly rely on such peer feedback in a wide range of service sectors such as travel (e.g., Tripadvisor), restaurants (e.g., Yelp), healthcare (e.g., Healthgrades), education (e.g., Coursera), law (e.g., Avvo), and finance (e.g., RateMyInvestor). Further, consumer-sourced rating systems have become a dominant method of worker evaluation and semi-automated management of large-disaggregated workforces on gig economy digital platforms such as TaskRabbit, Uber and Upwork (Rosenblat et al. 2016). Ostensibly this is a "win-win" for customers, service providers, and platforms; the general expectation is that rating systems will increase customer confidence in transacting on the platform, leading to greater volume of exchanges, providing more value to customers and greater earnings for workers and platforms.

The conventional wisdom is that online rating systems can also reduce discrimination against disadvantaged groups in online platforms (Cui et al. 2020). The idea is that by providing metrics about the performance and behavior of individual workers based on ex-post customer experiences, they can reduce statistical discrimination—where customers discriminate among groups (e.g., race, gender) based on their beliefs about differences in group performance.[1] But, a growing empirical literature has documented racial and gender biases in customer ratings on online platforms (e.g., Hannák et al. 2017, Botelho and Gertsberg 2022). Some have expressed concern that due to such biases, customer rating systems may be potential "vehicles for bias," where workers from

[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 on whether statistical discrimination arises from accurate or inaccurate beliefs of differences between groups (Bordalo et al. 2016). Lang and Spitzer (2020) however argue that inaccurate group belief based discrimination is better understood as taste based discrimination—when the evidence against the inaccurate beliefs are clear.

underrepresented or protected groups (e.g., racial/ethnic minorities, women) will more likely face earnings/employment discrimination (Rosenblat et al. 2016, Hanrahan et al. 2018).[2]

In this paper, we propose that customer rating systems may be more than vehicles for discrimination, but could be potentially "discrimination amplifiers" in a market. We formalize and then empirically investigate the idea that when customer ratings embed consumer choices based on taste or statistical discrimination by even a segment of customers, the display of aggregate ratings to consumers can amplify the discrimination over time and create discrimination spillovers even among customers who otherwise do not discriminate. Our proposed mechanism is as follows: to the extent that customers view ratings as metrics of worker quality, displaying the average ratings "memorializes" differences in ratings that may have partially arisen due to discriminatory tastes or beliefs as differences in "quality." As such, it impacts the beliefs about the "quality" of the rated individuals for future customers. Hence even if the difference in average ratings is due to discriminatory tastes or beliefs, all future customers will be impacted by these ratings in terms of how they rate or engage with minority groups. It can amplify the discrimination for those who already do and create discrimination spillovers to customers who otherwise may not have engaged in statistical or taste-based discrimination. This increased divergence in ratings will translate to increased divergence of minority earnings from those of Whites, if consumers prefer to have their jobs done with workers having higher ratings. Thus even if a minority and White worker were ex-ante identical in terms of quality, displaying customer ratings will systemically increase the divergence between minority and White workers in ratings and earnings.

We first formalize the intuition underlying our idea with a simple, stylized analytical model. We show that displaying summary ratings of customers causes (i) the discrimination of one segment to spill over to customers that otherwise do not discriminate; and (ii) reinforces and amplifies

[2] In contrast to our focus on discriminatory impact of reviews, there is a large literature on other problems related to the use of review based rating scores; such concerns include review manipulation (e.g., Mayzlin et al. 2014), the impact of social influence independent of actual experiences (e.g., Aral 2013) and selection on when and what to write about an experience (e.g. Chakraborty et al. 2022).

the discrimination by the discriminating segment. To the extent, that the resulting inequity is not related to the "true" minority worker performance quality, this disadvantage due to customer rating systems is an example of how minorities may face "systemic" or "structural" disadvantages in the marketplace, even if the platform had no such intent to disadvantage minorities.

Next we explore the issue empirically in the context of an online labor market platform that matches customers with service workers. There are several challenges to demonstrating the idea in a rich empirical setting. Relative to research that documents evidence discrimination in ratings on average, quantifying how discriminatory customer ratings translate to earnings inequity for service providers over time is more challenging. This is because, such translation requires modeling the stage at which discrimination occurs and how the discrimination at each stage impacts the dynamic evolution of ratings and earnings inequity over time. There are several aspects to consider in building and estimating such a model. First, even if there is discrimination on average, not all customers discriminate or exhibit similar levels of discrimination. Hence it is important to allow for unobserved heterogeneity in both the presence and level of discrimination by customers. Second, discrimination can manifest itself in different ways. Some customers prefer not to engage with minority groups and are more likely to cancel appointments when matched with such service providers. Others may engage with minority providers, but may be less likely to provide them reviews. Yet others may provide reviews, but may rate minority groups lower. Each of these discriminatory actions will have differential impacts on how the rating metrics evolve over time.

To address these issues, we develop a model that is flexible enough to capture discrimination at different stages of the consumer's journey and also allows for unobserved heterogeneity. We leverage the two-sided panel structure of the data arising from both customers and workers in our empirical setting for identification of unobserved heterogeneity. Specifically, customers offer multiple jobs on the platform that are assigned to both minority and non-minority workers, while workers perform a large number of jobs over time, assigned to various customer segments and accompanied with various levels of displayed metrics such as ratings, and number of past jobs.

We then use the estimates of the model and counterfactual analysis to address three sets of research questions. First, do customers on the platform discriminate against minorities? If yes, is there heterogeneity in how they discriminate (e.g., through discriminatory ratings or job cancellations)? Second, how much does customer discrimination impact the ratings and earnings of minority workers relative to White workers on the platform? Third, does display of customer ratings serve as discrimination amplifiers? If yes, how much does the display of ratings amplify the ratings gap for those who discriminate and create spillovers for those who do not discriminate? How much does it contribute to the earnings gap between minority and white workers?

Our key results are as follows: Our model estimates reveal that there are three segments; while one segment does not discriminate against minority workers, two segments discriminate, and in different ways. One discriminates by cancelling jobs assigned to minorities, while the other exhibits bias both at the cancellation and rating stages. Counterfactual analysis shows that customer discrimination contributes 24% 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; it increases the minority ratings gap by 80% and the minority earnings gap on the platform by 28%.

Though our paper focuses on the empirical setting of online ratings, the core idea of the paper that memorialization of discrimination-tainted metrics as representing "merit" or "quality" can lead to discrimination spillovers and amplification and greater inequity has broader implications beyond marketing settings into areas such as education and criminal justice. For example, (Owens 2022, Owens and McLanahan 2020) show that teacher biases can lead to greater minority student suspensions and expulsions. However when they become part of a student academic record (or the implicit impressions about a student), this can have longer term negative impact on further punishments, and the student's own performance. Similarly, early disparities in enforcement of laws with respect to drug use between minorities and whites create differential criminal records, with future implications for how the same illegal behaviors would be treated in the criminal justice systems due to “history.” In the context of the Indian caste system, and how it impacts admission to elite colleges in India, Subramanian (2019) argues that by relabeling differences arising from structural disadvantages of a group as a difference in contemporaneous observable “merit,” advantaged groups may not only justify their advantages, but also cast any effort at correcting structural inequities as contradictory to the goals of a merit-based society. But the conjecture is hard to test empirically due to data challenges and the time lag in measurable impact. Our modeling approach can also be seen as a way to formalize, and provide evidence for such a conjecture in a setting, where “history” and the evolution of merit and outcomes are fully observed.

From a modeling perspective, the paper contributes to the broader interest in issues of fairness in machine learning and algorithmic discrimination. In particular, the paper advances the important role of unobserved heterogeneity among agents who practice discrimination. Much work in the literature on bias and machine learning, discrimination demonstrate “average bias,” or condition bias on observables. However, we show that unobservable heterogeneity in bias dynamically impacts the “observable metrics” presented to customers and needs to be accounted for in machine learning. To the extent that “observable metrics” used by the system are tainted by past discrimination, it embeds structural inequities and disparities and perpetuates and amplifies them.

Further from a social perspective, a focus on average discrimination, without acknowledging unobserved heterogeneity leads to reactance against acceptance of the results, as many people may either feel that they are merit-oriented, not discriminatory, and even if discriminatory, not at all stages. By allowing for unobserved heterogeneity in discrimination, our modeling approach not only reflects a view of the market that is more true to reality, but also lead to lower reactance against the research findings around discrimination. The modeling approach that allows for interaction between discrimination and “merit” provides a valuable lens to look at panel data when choices may involve both discrimination, but also control variables that evolve over time, and include effects of lagged discrimination.

The rest of this paper is organized as follows: §2 positions the paper with respect to the literature on racial discrimination and the earnings gap. §3 presents a stylized analytical model that formalizes our intuition about how rating systems with bias can impact minorities through bias spillovers and bias accentuation. §4 describes the empirical setting and the data. §5 describes the empirical model of customer cancellations and ratings. §6 describes the results—descriptive analysis, model estimates and findings. §7 describes the counterfactual results. §8 concludes.

2. Related Literature

We position the paper against the broader literature on racial discrimination and the earnings gap. We organize the literature on racial discrimination involving consumers in Table 1 along three broad streams: (i) discrimination of consumers by firms/workers; (ii) discrimination of suppliers by consumers; and (iii) discrimination of workers by consumers/firms. We note that the literature in many of these areas is vast, and the references we provide are merely exemplars for the categories.

From a marketing perspective, the literature on the discrimination of consumers is obviously relevant. Within the literature on discrimination of consumers, there has been extensive research on discrimination in the housing (e.g., Yinger 1995) and credit (e.g., Blanchflower et al. 2003) markets. The literature on discrimination in consumer product markets (e.g., Ayres and Siegelman 1995, Edelman et al. 2017) is relatively small compared to the housing and credit markets. There is also a literature on discrimination of suppliers. On Ebay, prices for products are lower for blacks (Ayres et al. 2015) and females (Kricheli-Katz and Regev 2016). Similarly, black and female hosts obtained lower revenues than White and male hosts for similar properties on Airbnb (Marchenko 2019). In marketing, 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

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

Our focus in this paper is on worker (service provider) discrimination arising from consumer actions—an issue of particular relevance to marketers in various service settings. The literature here is relatively limited. Lynn and Sturman (2011) find that customers rate servers of the same race more favorably on promptness and attentiveness than those of a different race. Brewster and Lynn (2014) find that both white and black restaurant customers discriminate against black servers by tipping them less than their white co-workers. They find no evidence that this "Black tip penalty" results from inter-racial differences in service skills. Bar and Zussman (2017) find customer discrimination against Arab workers in the Israeli market for labor-intensive services. Jewish customers prefer firms employing Jewish, rather than Arab customers (for personal safety reasons), and these in turn adversely affect Arab hiring by firms and the prices in the market; firms with Arab workers charge lower prices for similar services. Overall, these results directly support Becker's predictions that customer preferences drive labor market discrimination. Kuppuswamy and Younkin (2020) find that consumer preferences drive the lack of minority hiring in Hollywood movies. Botelho and DeCelles (2023) find that customers cancel more often at a labor market platform when matched with minority customers than when matched with White customers. In crowdfunding, Younkin and Kuppuswamy (2016) find that projects of African American founders are funded less; and prospective funders rate these projects worse.

Such discrimination in the workplace leads to a significant lifetime earnings gap and reduced wealth accumulation for minorities. The combination of these biases leads to negative outcomes for minorities. The literature notes that while racial differences in social origins and socioeconomic characteristics play a role in creating wealth disparities (Altonji and Doraszelski 2005) , more than 70% of the race gap in wealth levels persists after controlling for a range of demographic, marital and socioeconomic factors (Oliver and Shapiro 2013, Shapiro et al. 2004). Overall, this literature concludes that differential socioeconomic resources are insufficient to explain the wealth and earnings gap. Reviews of the literature find that 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 than compared to Whites. (Pager and Shepherd 2008, Tomaskovic-Devey et al. 2005).

In contrast to these common regression-based approaches that seek to quantify the minority earnings gap across the economy or a large industry sector, the current paper uses the case study approach from industrial organization to build a process model of consumer choices that allows for discrimination in a particular empirical context and uses counterfactual analysis to quantify the earnings gap due to a firm's design/policy decisions (e.g., display of online ratings). The approach is particularly useful to marketers and firms to specifically consider design decisions that can mitigate discriminatory outcomes within a particular institutional setting.

3. A Stylized Analytical Model

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

Assume that there are two workers in the market, who are identical on all dimensions (skill, quality of service etc.), except that they belong to one of two observably different groups $w = \in \{A, D\}$. 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 an 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] We refer to the advantaged group as A and the disadvantaged group as D.

We assume a unit mass of customers who are of two types $c = \{N, P\}$, where N denotes customers who are neutral in that they do not discriminate in their behavior towards the two groups and 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. Each customer offers one job in each period; may choose to cancel the job when assigned a worker, and gives a rating on each job they accept at the end of the period. The partial customer may discriminate either at the job acceptance stage by canceling or at the rating stage.

We denote the rating for worker $w \in \{A, D\}$ by customer $c \in \{N, P\}$ at time t as r_t^{wc}. We start with the setting where the customers do not cancel jobs. 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: $r_t^w \equiv (r_t^{wN})^{1-\alpha}(r_t^{wP})^{\alpha}$.[5] 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).

[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 and statistical discrimination.

[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.

Suppose customers rate workers based on the work quality of the current period and past worker ratings (if the ratings are displayed). We assume that the workers have the same work quality of 1 and that the rating r^{wc} is the product of the work quality and the displayed ratings. Since neutral customers have neither differences in tastes nor beliefs between the two groups, when customers do not have any information about worker quality, they rate both advantaged and disadvantaged workers identically. But the partial customer discounts the rating of the disadvantaged customer relative to the advantaged customer by a factor of $1-\delta$, where δ is the degree of rating discrimination. Therefore, for the first job, given there is no rating displayed, both types customers rate the workers by work quality. So $R_1^{Ac} = r_1^{Ac} = 1 \forall\ c \in \{N, P\}$, $R_1^{DN} = r_1^{DN} = 1$, and $R_1^{DP} = r_1^{DP} = 1-\delta$. Starting from the second period, when customers are shown the geometric mean of past ratings of the worker R_{t-1}^w, neutral customers give a rating of R_{t-1}^w, and partial customers give a rating of $(1-\delta)R_{t-1}^w$. On the other hand, if the past rating is not displayed, we assume that the neutral customers always give a rating of 1 to both types of workers, while the partial customer gives a rating of 1 to workers in the advantaged group and a rating of $1-\delta$ to workers in the disadvantaged group.

In the setting where there are job cancellations, We assume that the customer's decision to accept/cancel a job upon being matched with the worker depends on the worker's displayed rating, which is a signal of worker quality. In each period t, the neutral customer accepts a job with probability R_{t-1}^w for workers $w \in \{A, D\}$. Thus if the cumulative rating $R_{t-1}^w = 1$, the job is always accepted. 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 with a lower probability R_{t-1}^D by discounting the worker's cumulative rating, where $0 < \gamma < 1$ and $1-\gamma$ indicates the level of acceptance discrimination.

We first consider the case where all jobs are accepted and there are no job cancellations. Subsequently, we extend the analysis to include job cancellations.

3.1. Case A: No job cancellations

We compare how ratings evolve when ratings are displayed or not displayed.

i. Ratings not displayed: Here the ratings of the advantaged and disadvantaged workers are straightforward. Neutral and partial customers give a rating of 1 to advantaged workers in all periods; i.e, $r_t^{AN} = r_t^{AP} = 1, \forall t$. Partial customers give a rating of 1 to advantaged workers but a rating of $1-\delta$ to disadvantaged workers in all periods, i.e., $r_t^{AP} = 1$ and $r_t^{DP} = 1-\delta$, $\forall t$. The geometric mean rating for the advantaged worker is $r_t^A = 1$, and that for the disadvantaged worker is $r_t^D = (1-\delta)^\alpha$ in each period t, but the summary ratings have no impact on other customers as they are only known to the platform, but not displayed to customers. The cumulative geometric mean ratings up to period t for the advantaged and disadvantaged worker are $R_t^A = 1$ and $R_t^D = (1-\delta)^\alpha$ respectively.

ii. Ratings displayed: Again, the neutral and partial customers give a rating of 1 to the worker of advantaged group in all periods; i.e, $r_t^{AP} = r_t^{AN} = 1, \forall t$. In the first period, when they have no information on past rating, neutral customers give a rating of 1 i.e., $r_1^{DN} = 1$, partial customers give a rating of $1-\delta$ to the disadvantaged worker, i.e, $r_1^{DP} = 1-\delta$. Then for period t, the rating for the disadvantaged worker from the partial customers is given by

$$r_t^{DP} = (1-\delta)R_{t-1}^D \tag{1}$$

where R_{t-1}^D is the displayed rating after the minority worker finishes jobs in period $t-1$. The neutral customers give a rating of $r_t^{DN} = R_{t-1}^D$. The geometric mean of the rating for the current period is $r_t^D = ((1-\delta)R_{t-1}^D)^\alpha (R_{t-1}^D)^{1-\alpha} = (1-\delta)^\alpha R_{t-1}^D$. The cumulative geometric mean ratings at the end of period t is given by

$$R_t^D = (R_{t-1}^D)^{\frac{t-1}{t}} (r_t^D)^{\frac{1}{t}} = (R_{t-1}^D)^{\frac{t-1}{t}} ((1-\delta)^\alpha R_{t-1}^D)^{\frac{1}{t}} = (1-\delta)^{\frac{\alpha}{t}} R_{t-1}^D$$

Given we assume $R_0^D = 1$ so that $r_1^{DN} = 1$, $r_1^{DP} = 1-\delta$, and $R_1^D = r_1^D = (1-\delta)^\alpha$, the recursive equation above can be reduced to

$$R_t^D = (1-\delta)^{\alpha \sum_{i=1}^{t} \frac{1}{i}}$$

The power term on $(1-\delta)^\alpha$ is the harmonic series, which is divergent. So R_t^D approaches 0 as t gets large. The neutral and partial customer ratings also approach 0 as t gets large.

Since the displayed ratings cause future customers to interpret partiality of the partial customer as a metric of quality, it creates a spillover of the partiality onto neutral customers and makes the partial customer even more partial.

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 (δ) and the size of the partial customer segment (α); i.e., $\partial r_t^{DN}/\partial\delta < 0$, $\partial r_t^{DN}/\partial\alpha < 0$.*

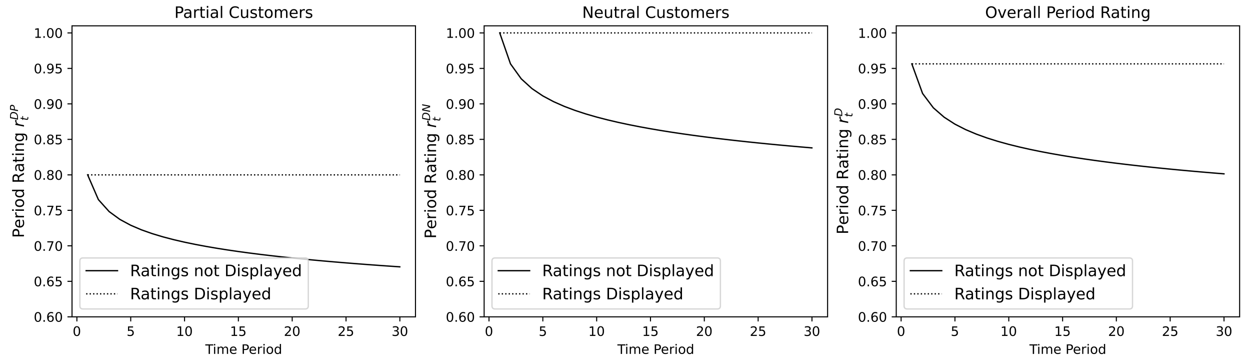
In contrast, when summary ratings are not displayed, the neutral customer's rating for the disadvantaged worker is not impacted by the rating partiality and the size of the partial customer segment; i.e., $\partial r_t^{DN}/\partial\delta = 0$, $\partial r_t^{DN}/\partial\alpha = 0$.

Proposition 2 (Discrimination Amplification)**.** *When ratings are displayed, the rating of disadvantaged worker decreases over time for both neutral and partial customers i.e., $\partial r_t^{DN}/\partial t < 0$ and $\partial r_t^{DP}/\partial t < 0$. Over time, both the neutral and partial customer's rating becomes less favorable to the disadvantaged worker, but the partial customer always remains less favorable by a factor of $1-\delta$.*

In contrast, when the ratings are not displayed, the rating of the disadvantaged worker is constant over time for both neutral and partial customers i.e., $\partial r_t^{DN}/\partial t = 0$ and $\partial r_t^{DP}/\partial t = 0$ $\partial r_t^{DN}/\partial\alpha < 0$.

Figure 1 graphically shows the results of proposition 2 to illustrate how displaying ratings amplify discrimination for the neutral and partial segments and overall. The first plot shows the rating of the partial customers when the worker's cumulative rating is not displayed and when the rating is displayed. The downward sloping trend of the rating in the latter case shows that displayed rating further amplifies the impact of partiality. The second plot shows the ratings from the neutral customers, where the rating is not different relative to the advantaged worker when the cumulative ratings are not displayed. However, when the ratings are displayed, the ratings of the neutral customers of the disadvantaged workers are negatively affected by partial customers.

Figure 1 Illustration of Prop. 2: Ratings for the Disadvantaged Worker.



Note: $\delta = 0.2$ and $\alpha = 0.2$.

3.2. Case B: Customers cancel some jobs based on ratings

Again, we compare the cases where rating are not displayed versus displayed. Here the cancellation depends on the displayed rating (a signal of worker quality). To keep track of the workers' earnings, we denote the number of jobs requested by customer type c and completed by worker type w in period t as n_t^{wc}. Note that n_t^{wc} 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 $n_t^w \equiv n_t^{wN} + n_t^{wP}$. The total number of jobs completed at the end of period t is given by $N_t^w \equiv \sum_{t'=1}^{t} n_{t'}^w$.

i. Ratings not displayed:

Neutral customers start with a prior of $R_0^w = 1, \forall w \in A, D$ while partial customers start with a prior of $R_0^A = 1$ and $R_0^D = 1 - \delta$. Given this prior, neutral and partial customers give a rating of 1 to advantaged workers in all periods; i.e, $r_t^{AN} = r_t^{AP} = 1, \forall t$. The geometric mean rating for the advantaged worker $r_t^A = 1, \forall t$. The cumulative geometric mean ratings up to period t for the advantaged worker is $R_t^A = 1$. Hence all jobs are accepted for the advantaged worker.

Neutral and partial customers give a rating of 1 and $1 - \delta$ respectively to disadvantaged workers in all periods, i.e., $r_t^{DN} = 1$, $r_t^{DP} = 1 - \delta$, $\forall t$. As ratings are not displayed, partial customers accept jobs when matched with the disadvantaged worker with probability $\gamma(1 - \delta)$, reflecting their rating and acceptance discrimination. For the disadvantaged worker, the total job count for period t is given by

$$n_t^D = 1 - \alpha + \alpha\gamma(1 - \delta) \tag{2}$$

The cumulative job count up to period t is therefore

$$N_t^D = t(1-\alpha+\alpha\gamma(1-\delta))$$

ii. Ratings displayed: Table 2 shows the expected job count (after accounting for cancellation probabilities) and the rating as a function of cumulative rating from both types of customers for each worker:

Table 2 **Expected job count and rating at period** t

Worker w	Customer c	Job Count (n_t^{wc})	Rating (r_t^{wc})
A	N	$(1-\alpha)R_{t-1}^A$	R_{t-1}^A
A	P	αR_{t-1}^A	R_{t-1}^A
D	N	$(1-\alpha)R_{t-1}^D$	R_{t-1}^D
D	P	$\alpha\gamma R_{t-1}^D$	$(1-\delta)R_{t-1}^D$

It is trivial to see that, similar to case 1, the advantaged worker always receives a rating of 1 from both types of customers and thus receives one job from each period. That is, $n_t^A = 1$, $r_t^A = 1$, $R_t^A = 1$, $\forall t$. For the disadvantaged worker, based on Table 2, the total job count for period t is given by

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

The average rating for period t is

$$r_t^D = (R_{t-1}^D)^{\frac{1-\alpha}{1-\alpha+\alpha\gamma}}((1-\delta)R_{t-1}^D)^{\frac{\alpha\gamma}{1-\alpha+\alpha\gamma}} = (1-\delta)^{\frac{\alpha\gamma}{1-\alpha+\alpha\gamma}}R_{t-1}^D \tag{4}$$

Then the cumulative job count can be written as

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

The cumulative rating can be shown to be

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

Propositions 1 and 2 continue to hold for the case where customers cancel jobs based on ratings. However, as the partial customers are the ones that withdraw from the rating score by cancellations, the rate of ratings decline over time is slower. This can be seen by comparing equation (4) with equation (1) in the no-cancellation case because the term $0 < \frac{\alpha\gamma}{1-\alpha+\alpha\gamma} < \alpha$. Since r_t^D in the cancellation case is always higher, R_t^D will also be higher. This leads us to proposition 3.

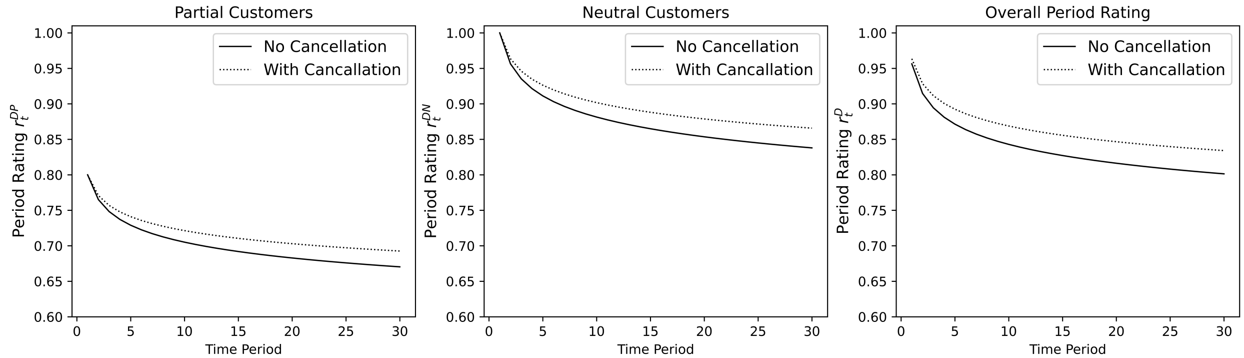
Proposition 3 (Discriminatory Spillover and Amplification with Cancellations)**.** *In the presence of cancellations, when ratings are displayed, the results from proposition (1) and (3) continue to hold. Further, the overall ratings and ratings of the neutral and partial segments are higher than in the no cancellation case. Also, the rate of decline in ratings over time is smaller than with the no cancellation case.*

Figure 2 shows the job rating and cumulative job count for the disadvantaged worker by the two types of customers for a chosen set of parameters. Similar to case 2, for the partial customers, the rating difference between disadvantaged and advantaged workers increases over time. For the neutral customers, there is still the spillover effect from the partial customers where the rating trend is downward sloping compared to the case where rating is not displayed. However, under this setting, given that the partial customers cancel more jobs from the disadvantaged worker compared to the neutral customers, the partial customer's ratings have a lower impact on the worker's overall rating. Therefore, when ratings are displayed, the job rating is higher when cancellations are impacted by ratings than when they are not.

Comparing equations (2) and (3), we see that cumulative ratings impact the neutral customer segment of size $1-\alpha$ when ratings are displayed, but not otherwise. Further $0 < R_{t-1}^D < 1-\delta$. 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 4.

Proposition 4 (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.*

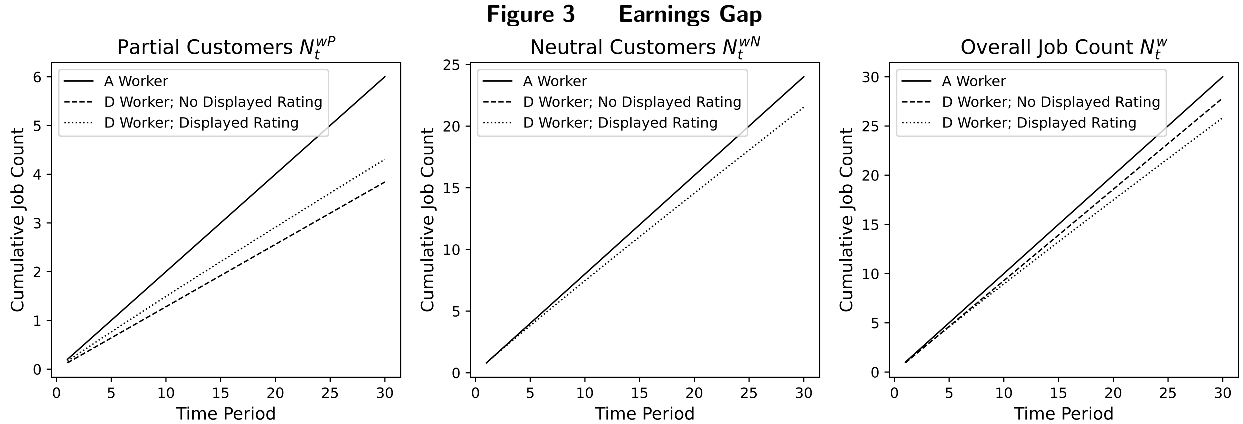
Figure 2 Comparison of Job Rating with and without Cancellation.



Note: $\alpha = 0.2$, $\delta = 0.2$ and $\gamma = 0.8$.

Figure 3 shows the corresponding earnings gap when cancellation is based on ratings. Compared to the advantaged worker, the accumulated earnings gap from the partial customers continues to widen across time. Moreover, when cancellation depends on displayed rating, the spillover effect is also carried over to the job count, where there is earning gap between the two workers even from neutral customers.

Figure 3 Earnings Gap



Note: $\alpha = 0.2$, $\delta = 0.2$ and $\gamma = 0.8$.

4. Empirical Setting

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 matches customers to workers to perform a variety of home-service jobs across a pre-defined list of service categories, such as appliance installation and repair, electrical work, maintenance services, and plumbing. SC mediates the entire transaction between customers and workers, including completing the match between customers and workers, marketing, payment, scheduling assistance, and worker verification.

Most workers on SC are small-business entrepreneurs who focus on certain types of jobs that they complete on SC. To join SC, workers must apply and complete SC's screening process. Screening includes an interview, verification of an individual's skills, and a criminal background check. Customers are reminded that workers must be verified and screened to join SC's platform. This verification process should help attenuate concerns around large variations in worker quality.

A customer creates a job on SC's website or app using a survey to explain what job needs to be completed. Jobs listed on SC typically require a few hours of work; they are not large-scale construction projects. After a customer provides the job's specifications, a minimum cost and an hourly rate are generated. These costs are constant within service category and do not differ by worker nor can they be negotiated by the customer or worker. The customer is then prompted to select among available days of the week and times of day for the job to be completed.

After a job is submitted, SC assigns workers to customers using an algorithm. The algorithm prioritizes a small set of workers mostly based on number of jobs the worker has completed and worker's average rating. The small set of workers has 15 minutes to accept the job on a first come, first served basis. If no one accepts the job, the job is open to all workers on a first come, first served basis. Importantly, the algorithm is blind to worker demographics and customer characteristics, as they are not inputs.[6] Since the customer has no control over this matching process, the job assignment can be considered as random with regard to the relationship between perceived worker race and customer characteristics.

Once a worker accepts a job, the customer receives a message providing the worker's name, photo, average rating to date, and number of jobs completed to date. A customer can cancel the

[6] SC does not collect customer demographics.

job at any time before the worker indicates that they are physically on their way. After a job is completed, customers are sent an email asking them to rate their experience with the worker from 1 to 5 stars.[7]

5. Model and Estimation

As our focus is on customer behaviors after the platform's job assignment, we model three customer behaviors: job cancellation, whether to submit a rating, what rating to offer. In this section, we begin by introducing a model of the three customer behaviors. We then describe the estimation approach. Finally, we clarify the features of the data that helps identification of the model's key parameters around discrimination.

5.1. Model of Job Cancellation and Rating

The model consists of three parts, related to three behaviors we model. The first part models the binary decision of whether to cancel the job that has been assigned by the platform to a worker. The second 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_{ijk} takes the vaue of 1 if the customer cancels the job j conditional on worker k accepting the job. Let X^c_{ijk} 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:

$$P^{Cg}_{ijk} = P^g(C_{ijk} = 1) = \frac{exp(\beta^g X^c_{ijk}))}{1 + exp(\beta^g X^c_{ijk}))},$$

where β^g denotes the parameters for segment g.

[7] Fig. A.1 in the appendix shows the two messages sent to the customers.

For each job j that is completed, let $W_{ijk} \in \{0,1\}$ denote if the customer submits the rating to the platform. W_{ijk} is modeled as a binary logit model. For a customer that belongs to the segment g, the probability of writing a review given the job is not cancelled is:

$$P_{ijk}^{Wg} = P^g(W_{ijk} = 1 | C_{ijk} = 0) = \frac{exp(\delta^g X_{ijk}^w)}{1 + exp(\delta^g X_{ijk}^w)},$$

where X_{ijk}^w are the observables and δ^g denotes the parameters for segment g.

Finally, conditional on a rating being given, we model the rating choice as a binary outcome: R_{ijk} takes the value of 1 if the customer gives a rating of 5-star.[8] For a customer that belongs to segment g, the probability of giving a 5-star rating given a rating is submitted is

$$P_{ijk}^{Rg} = P^g(R_{ijk} = 1 | W_{ijk} = 0) = \frac{exp(\delta^g X_{ijk}^r)}{1 + exp(\gamma^g X_{ijk}^r)},$$

where X_{ijk}^r are the observables and δ^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 $\Theta_g = \{\beta^g, \delta^g, \gamma^g, q^g\}$. Let $J(i)$ denote the set of jobs requested by customer i and $S_i = \{C_{ijk}, W_{ijk}, R_{ijk}\}_{j \in J(i)}$ denote the set of observations of customer i. The likelihood function of an individual customer i that belongs to segment g is given by

$$\begin{aligned} L_i^g(S_i; \Theta_g) = & \underbrace{\Bigg[\prod_{j \in J(i): C_{ijk}=1} (P_{ijk}^{Cg})^{C_{ijk}} \Bigg]}_{\text{for jobs canceled}} \times \underbrace{\Bigg[\prod_{j \in J(i): C_{ijk}=1, W_{ijk}=0} (1 - P_{ijk}^{Cg})^{1-C_{ijk}} (1 - P_{ijk}^{Wg})^{(1-W_{ijk})} \Bigg]}_{\text{for jobs not canceled but not rated}} \\ & \times \underbrace{\Bigg[\prod_{j \in J(i): C_{ijk}=1, W_{ijk}=1} (1 - P_{ijk}^{Cg})^{1-C_{ijk}} (P_{ijk}^{Wg})^{W_{ijk}} (P_{ijk}^{Rg})^{R_{ijk}} (1 - P_{ijk}^{Rg})^{1-R_{ijk}} \Bigg]}_{\text{for jobs not canceled \& rated}}, \end{aligned} \tag{5}$$

where the first component accounts for jobs being allocated to a worker and then cancelled by customer i, the second component accounts for jobs not canceled and not rated, and the third

[8] 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.

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:

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

Then the log-likelihood over all customers is given by

$$\sum_i log(L_i(S_i;\Theta)) = \sum_i log(\sum_g q_g L_i^g(S_i;\Theta_g)) \tag{6}$$

5.2. Estimation

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

$$\sum_i^N \sum_g^G q_i^g log(L_i^g(S_i;\Theta_g)), \tag{7}$$

where q_i^g is defined below as the probability that customer i is of segment g given parameters values Θ, conditional on all of the observed jobs of customer i:

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

The EM algorithm is implemented in the following steps:

1. Initialize Θ_0. We intialize $q^g = 1/G$. Then we randomly partition the customers into G segments and maximize the three components of the log-likelihood for each segment to get the intial values for Θ_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 Θ:

$$q_i^g = P(i \in g|S_i) = \frac{q^g L_i^g(S_i;\Theta_g)}{\sum_{g'} q^{g'} L_i^{g'}(S_i;\Theta_{g'})} \tag{8}$$

3. Update $q^g = \frac{\sum_i q_i^g}{\sum_i \sum_{g'} q_i^{g'}}$.
4. Update the coefficients Θ_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 a 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 customers' 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, allows us to treat these characteristics as randomly assigned.

At the cancellation stage (after the job assignment), the customer observes worker characteristics (number of past jobs, ratings, and 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–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 or statistical discrimination.

At the rating stage, the 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 a random draw around a fixed mean performance. A customer's rating for a job will be correlated with 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 interprets this as evidence of either taste 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 customers, we allow for unobserved heterogeneity in customer's 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 segments show different levels and types of discrimination (or no discrimination at all) in terms of differential impact of 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 ratings for minorities will be greater across all segments in the absence of discrimination.

6. Empirical Analysis

We begin with a descriptive analysis of the empirical setting to both aid in understanding of the empirical setting and help with empirical specification of the model. We then report the model estimates—showing evidence of discrimination among some latent segments and none in others. We then describe the characteristics of the latent segments to facilitate interpretation.

6.1. Descriptive Analysis

We obtained data on all jobs from one metro area in North America during the four year period from 2016-19 in which SC operates for the purposes of this analysis. This consisted of 70169 jobs requested by 26610 unique customers and assigned to 586 workers. Given that there are differences in average cancellation rates and ratings for different service categories, 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 used a threshold of at least 4 jobs offered by a customer to include in our analysis. Finally, we deleted the 11 female workers in the data as the platform workers are primarily male. Additionally, given that we use the worker's photo to code a customer's perception of race, we excluded jobs where the worker's photo was unavailable or not of an individual. After these deletions, our sample for model estimation includes 36803 job requested by 5345 customers and accepted by 552 workers.

SC categorizes service jobs into several categories, including maintenance, plumbing, appliance, electrical etc. Given our threshold of 1.5% of the overall jobs for each service category, there are 12 service categories. These are listed in Table 3. Of the 12 categories, 11 categories can be classified as those where jobs require worker credentials. For example, electrical and HVAC requires credentials, while snow removal does not. However, appliance is an exception as some appliance jobs require credentials, while others do not. Therefore, we split appliance into two categories, based on whether credentials are needed. The cancellation rates and average ratings vary across service categories.

Table 3 Job Count, Cancellation and Rating by Service Type

Service Category	Job Count	%Count	%Cancel	%Jobs w/ Rating	Avg. Rating
Job requires credential					
Plumbing	5460	15.14%	14.21%	63.21%	4.76
Appliance	3923	10.88%	13.13%	63.57%	4.73
Electrical	3290	9.12%	14.71%	63.95%	4.79
HVAC	2397	6.65%	21.32%	60.74%	4.74
Job does not requires credential					
Maintenance	7556	20.95%	19.39%	61.16%	4.77
Landscaping	3641	10.10%	17.69%	61.19%	4.58
Gutters	2253	6.25%	14.11%	62.23%	4.69
Snow	2220	6.16%	16.62%	61.04%	4.61
Moving	1727	4.79%	18.24%	64.74%	4.79
Upholstery	1208	3.35%	28.81%	54.14%	4.62
Appliance	923	2.56%	19.28%	60.89%	4.82
Locksmith	774	2.15%	11.24%	71.32%	4.85
Misc. Outdoors	688	1.91%	20.49%	64.10%	4.66

Table 4 presents statistics on the number of jobs, cancellations and ratings by 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. This suggests accounting for year fixed effects in our model.

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

Table 4 Job Count, Cancellation and Rating across Years

Year	Job Count	%Count	%Cancel	%Jobs w/ Rating	Avg. Rating
2016	3934	10.69%	13.22%	68.68%	4.61
2017	7645	20.77%	15.25%	70.35%	4.68
2018	10827	29.42%	17.58%	62.71%	4.74
2019	14397	39.12%	18.77%	55.64%	4.79

the workers' profile pictures and note their perception of the worker's race using the pre-defined categories: Black, East Asian, Hispanic, South Asian, Unclear, White. More details about the coding can be found in Botelho and DeCelles (2023). Given that the majority of workers were perceived 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 minority if the worker is not perceived as White by our coders, and non-minority if otherwise.

With customers with at least 4 job requests in our estimation sample, the data sample preserves a mixing of minority workers and non-minority workers within each individual customers. Fig. 4 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 non-minority workers or both types of workers. Among the customers who have requested 4 jobs in the data, 78.38% have been assigned to both types of workers. The ratio goes up to 85.71% for customers who have requested 5 jobs and increases with the customer job count. The plot shows the exposure to both minority and non-minority workers at customer-level.

Figure 4 Customer Distribution by Number of Jobs Requested and Worker Composition

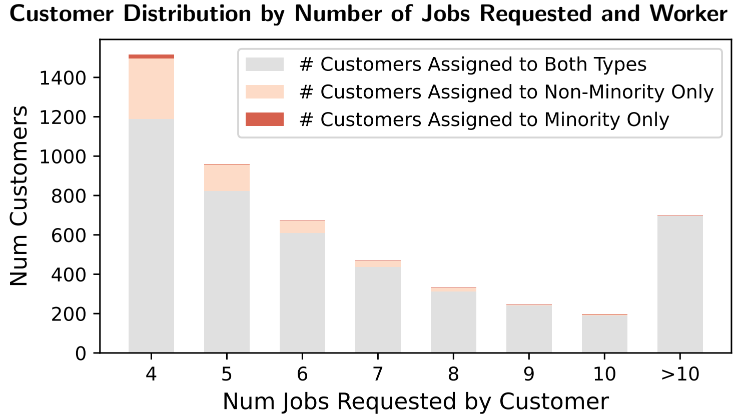


Table 5 presents the average job outcomes and the job outcomes by worker race. The summary statistics are suggestive of minority workers being disadvantaged in both cancellation and rating processes. Overall, minority worker have a higher proportion of jobs canceled. Once a job is finished, there is no significant difference in the rating submission rate between the minority and non-minority 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 non-5-star ratings.

Table 5 Descriptive Statistics

	All Jobs	Non-Minority	Minority	Minority Gap
% Workers		65.70%	34.30%	31.40%
Cancellation Rate	17.06%	15.79%	19.72%	-3.92%
Rating Submission Rate	75.02%	75.16%	74.71%	0.45%
Ratings:				
Avg. Rating	4.73	4.75	4.68	0.07
Rating=5	83.31%	84.36%	80.98%	3.38%
% for Ratings below 5				
Rating below 5	16.69%	15.64%	19.02%	-3.38%
Rating=4	10.76%	10.29%	11.81%	-1.52%
Rating=3	3.05%	2.78%	3.66%	-0.88%
Rating=2	1.09%	0.93%	1.45%	-0.52%
Rating=1	1.78%	1.64%	2.10%	-0.46%

Table 5 seems to be suggestive of the taste-based discrimination, where the minorities workers are faced with higher cancellation and lower rating. On the other hand, given that the platform displays the workers' average rating to the customers, [9] we explore if the displayed rating resolves customer uncertainty and mitigates the gap between minority and non-minority worker. For each job, we calculate the real-time average rating to be displayed for the worker (if the worker has received at least 5 ratings). Given that 89% of the displayed average rating are above 4.5, we use 4.5 as the cutoff and define the metric *positive rating* as the measure of displayed rating. Specifically,

[9] 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 more than 5 ratings, the average rating is available to the customer when the job is assigned.

we take the difference between the displayed average rating and the cutoff of 4.5 and normalize it into $[0,1]$ scale. For ratings below 4.5, the metric takes the value of 0:

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

Table 6 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 could potential resolves uncertainty, customers do not give the same credit to minorities and non-minorities even conditional on displayed ratings. Second, job rating is positively associated with higher displayed rating. This could potentially be driven by two mechanisms. Higher displayed rating captures higher worker quality and job quality, which in turns leads to higher job rating. Meanwhile, customers can be inclined to follow others' rating so that the job rating is higher when the observed worker rating is higher. However, regardless of the mechanism, minorities still get lower ratings conditional on the displayed rating.

Table 6 Job Outcomes by Positive Ratings

pos_rating	**Cancellation Rate**			**Average Job Rating**		
range	All Jobs	Non-Minority	Minority	All Jobs	Non-Minority	Minority
$[0, 0.25)$	18.68%	17.25%	21.11%	4.57	4.59	4.54
$[0.25, 0.5)$	17.36%	15.69%	19.59%	4.71	4.71	4.7
$[0.5, 0.75)$	16.2%	15.8%	17.58%	4.8	4.81	4.75
$[0.75, 1]$	15.72%	14.12%	20.75%	4.84	4.85	4.78

6.2. Model Estimates

Table 7 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 (as described in §6.1). We also include a dummy variable of *no rating* to indicate if the worker has received less than 5 ratings and the rating is not displayed to the customers. Given our primary interest of how consumers differentially make choices for minorities, we include a minority main effect and its interaction with the positive rating. Finally, we also include an interaction of minority and the dummy variable for whether the service needs credential.[10] In this way, we may see if the job credential also resolves the job uncertainty and mitigates the minority gap.

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.[11] 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), the minority coefficient is significant only for the cancellation equation. 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 non-minority workers, customers in segment 2 do not give as much credit to minority workers based on the displayed rating. This reflects the γ in the analytical model. On the other hand, the minority variable and

[10] The main effect of the credential dummy is already absorbed in service category fixed effects.

[11] The log-likelihood for two and three models are -39555 and -39009 respectively. The corresponding AIC values are 79365 and 78400. The four-segment model does not converge due to over-parameterization.

Table 7 Consumer Model Parameter Estimates

	Segment 1: Neutral Customers	**Segment 2**: Minority Avoiders	**Segment 3**: Minority Avoiders &Under-Raters
Segment Probability	0.19	0.24	0.57
SE	0.02	0.04	0.02
Cancellation			
logWorkerJobCount	-0.08*	-0.01	-0.05**
minority	0.16	0.04	0.17*
noRating	-0.47	-0.08	-0.55***
posRating	-0.69***	-0.58***	-0.58***
minority*posRating	0.19	0.61*	-0.16
Service FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Submit Rating			
logWorkerJobCount	-0.14**	-0.08**	-0.04
minority	0.26	0.15	0.11
noRating	-0.70*	0.30	0.71**
posRating	-0.02	0.48**	0.64***
minority*posRating	-0.68	-0.19	-0.19
Service FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Rating			
logWorkerJobCount	0.08*	-0.06	0.03
minority	-0.12	-0.28	0.12
noRating	1.21***	0.76	0.34
posRating	1.15***	1.03*	1.24***
minority*posRating	0.05	0.18	-0.67**
Service FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
N: 36803			
log-likelihood: -39009			
AIC: 78400			

its interaction terms are not significant in the rating equation for this segment. We call this the minority "avoider" segment.

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 δ in the analytical model. We call this the minority "avoider and under-rater" segment.

6.3. Latent Segment Characteristics

Table 8 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 required by these segments. In the cancellation stage, the difference between the minority and non-minority 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 non-minority 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 highly sensitive to positive ratings (which are biased by the minority underrated third segment).

Table 8 Customer Segments Description

	Segment 1 Neutral Customers	**Segment 2** Minority Avoiders	**Segment 3** Minority Avoiders & Under-Raters
total job count	5799	8053	22208
% jobs need credential	40.56%	41.26%	42.30%
Cancellation Rate			
non-minority	17.82%	14.30%	15.81%
minority	21.34%	19.75%	19.26%
Rating Submission Rate			
non-minority	81.11%	27.55%	91.19%
minority	79.55%	28.40%	90.25%
Perc. Rating=5			
non-minority	48.70%	90.92%	91.58%
minority	42.43%	87.31%	89.22%

7. Counterfactuals

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

7.1. Customer discrimination and the minority gap

To study the effect of customer discrimination on minority ratings and earnings gap within each service category, we compare the ratings and earning differences between minority and non-minority workers for a sequence of jobs within each service category given (i) the model estimates; and (ii) setting all minority and minority interaction coefficients in the model to zero. 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.[12] 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 non-minority worker assignment and then with a minority worker assignment. For each worker, we repeat the simulation for 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 non-minority and minority worker and the earnings gap for all service categories are reported in Table 9.

The second column shows the proportion of overall jobs from each service category in the raw data. We report the percentage of non-5-star rating for non-minority worker and minority worker separately as well as the percentage gap between them. The earnings gap is calculated as the percentage difference in total job count for the non-minority worker compared to the minority worker. In all service categories, the non-minority worker receives both a higher average rating and more jobs. The differences between the minority and non-minority worker vary across service categories. The overall ratings gap, which is the average of service-specific earnings gap, weighted by the proportion of jobs is 24%. The overall minority earnings gap is 4%.

[12] 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. Finally, 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.

Table 9 Minority Rating and Earnings Gap by Service Category

Service Category	% Job	% Non-5-Star Rating			Job Count (Earnings)		
		Non-Minority	Minority	%Gap	Non-Minority	Minority	%Gap
Maintenance	21	9.06	11.61	28	247	235	5
Plumbing	15	10.86	12.82	18	260	252	3
Appliance (Cred.)	11	11.30	13.00	15	262	254	3
Landscaping	10	15.99	19.28	21	254	246	4
Electrical	9	10.35	12.70	23	256	247	3
HVAC	7	10.80	12.79	18	240	228	5
Gutters	6	11.53	14.89	29	263	255	3
Snow	6	13.84	17.38	26	257	248	4
Moving	5	9.12	11.86	30	249	237	5
Upholstery	3	16.65	19.96	20	220	205	7
Appliance (Non Cred.)	3	6.54	9.74	49	247	236	5
Locksmith	2	8.25	11.23	36	268	259	3
Misc. Outdoors	2	12.13	16.70	38	243	231	5

7.2. Customer ratings display and the minority gap

Here we compare the ratings and earning differences between minority and non-minority workers for a sequence of jobs when (i) ratings are not displayed versus (ii) when ratings are displayed. The simulation steps are identical to the previous section except for one difference. When the rating is not displayed, 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. For the display case, we make the same assumption about imputation, until the first 5 ratings are generated, because the platform does not display ratings until then.[13]

Table 10 reports the gaps with and without displayed ratings by service category. The top panel shows the gaps in the probability of cancellation, which indicates the earnings gap. The bottom panel shows the gaps in the probability of giving a 5-star rating. The second column of each table shows the average rating for the service category in the data. For segment 1, where the customers are neutral, the gap is larger when ratings are displayed for all service categories for both cancellation and the rating probabilities. This suggests that displayed rating drives bias spillover

[13] We do not make different imputations for minority and non-minority workers, as we did not find significant minority interaction coefficients for *No Rating* in the model estimation.

Table 10 Minority Gap with and without Displayed Rating

Service Category	Gap in Probability of Cancellation (in %)								% ↑ in Gap
	Segment 1		Segment 2		Segment 3		All Jobs		
	No D.	Disp.	No D.	Disp.	No D.	Disp.	No D.	Disp.	
Maintenance	0.08	0.60	5.44	7.65	3.03	3.21	3.04	3.78	24
Plumbing	0.05	0.45	4.08	5.85	2.49	2.55	2.41	2.94	22
Appliance (Cred.)	0.04	0.21	3.32	5.09	2.49	2.35	2.22	2.60	17
Landscaping	0.05	0.68	1.38	4.80	3.09	3.27	2.09	3.14	50
Electrical	0.07	0.53	4.39	5.67	2.60	2.75	2.55	3.03	19
HVAC	0.07	0.44	4.80	7.20	3.42	3.46	3.11	3.79	22
Gutters	0.04	0.46	3.27	5.97	2.50	2.56	2.21	2.99	35
Snow	0.04	0.65	1.91	5.13	3.02	3.18	2.18	3.16	45
Moving	0.08	0.52	5.44	7.27	3.05	3.26	3.06	3.70	21
Upholstery	0.07	1.24	3.03	7.53	4.01	4.86	3.02	4.81	60
Appliance (Non Cred.)	0.09	0.63	6.06	7.65	2.96	3.20	3.16	3.78	20
Locksmith	0.03	0.31	4.13	4.86	2.12	2.24	2.21	2.50	13
Misc. Outdoors	0.07	0.89	3.84	8.05	3.11	3.31	2.70	3.99	48
Service Category	**Gap in Probability of 5-Star Rating (in %)**								**% ↑ in Gap**
	Segment 1		Segment 2		Segment 3		All Jobs		
	No D.	Disp.	No D.	Disp.	No D.	Disp.	No D.	Disp.	
Maintenance	0.09	1.07	0.00	0.16	2.45	3.12	1.41	2.01	43
Plumbing	0.08	1.19	0.00	0.10	2.59	3.36	1.48	2.15	45
Appliance (Cred.)	0.07	0.57	0.00	0.08	2.25	2.99	1.29	1.82	41
Landscaping	0.07	2.22	0.00	0.44	1.51	4.59	0.87	3.13	261
Electrical	0.08	1.04	0.00	0.16	2.73	3.31	1.56	2.12	35
HVAC	0.09	0.83	0.00	0.10	2.64	3.48	1.52	2.16	42
Gutters	0.09	1.64	0.00	0.41	2.46	3.97	1.41	2.66	88
Snow	0.07	2.11	0.00	0.29	1.94	4.56	1.12	3.06	174
Moving	0.11	1.01	0.00	0.08	2.61	3.21	1.50	2.03	35
Upholstery	0.08	2.75	0.00	0.45	1.77	3.95	1.02	2.87	182
Appliance (Non Cred.)	0.08	0.72	0.00	0.12	2.78	3.29	1.59	2.03	28
Locksmith	0.04	0.68	0.00	0.05	2.87	3.24	1.64	1.98	21
Misc. Outdoors	0.09	1.78	0.00	0.42	2.46	4.47	1.41	2.98	111

from other segments. Similarly, the pattern can be seen for segment 2, where the minority gap in rating probability is higher for the segment when ratings are displayed, even though this segment does not discriminate on ratings. For segment 3, where customers are both minority avoiders and minority under-raters, when the rating is displayed, the minority gap in cancellation and rating is higher for all service categories. Moreover, the overall minority gaps for cancellation and rating are larger with ratings displayed, as can be seen from the last column of each table. The overall increase in earnings gap (weighted by proportion of jobs in each service category) is 28%. Similarly, the overall increase in the ratings gap is 80%. This shows the discrimination amplification effect of displayed ratings.

8. Conclusion

The growth of online e-commerce and service platforms have made customer rating systems ubiquitous. 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 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 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 memorialize 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 24% 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 24% and the earnings gap on the platform by 28%.

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 police officers in whether they discriminate and the magnitude of discrimination. Not only does accounting for unobserved heterogeneity provide a 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.

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.

References

Altonji JG, Doraszelski U (2005) The role of permanent income and demographics in black/white differences in wealth. *Journal of Human Resources* 40(1):1–30.

Aral S (2013) The problem with online ratings. *MIT Sloan Management Review* .

Arrow KJ (1973) The theory of discrimination. *Discrimination in Labor Markets*, 1–33 (Princeton NJ: Princeton University Press).

Ayres I, Banaji M, Jolls C (2015) Race effects on ebay. *The RAND Journal of Economics* 46(4):891–917.

Ayres I, Siegelman P (1995) Race and gender discrimination in bargaining for a new car. *The American Economic Review* 304–321.

Bar R, Zussman A (2017) Customer discrimination: evidence from israel. *Journal of Labor Economics* 35(4):1031–1059.

Becker G (1957) The economics of discrimination 1st ed (university of chicago press, chicago).

Bertrand M, Mullainathan S (2004) Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. *American economic review* 94(4):991–1013.

Blanchflower DG, Levine PB, Zimmerman DJ (2003) Discrimination in the small-business credit market. *Review of Economics and Statistics* 85(4):930–943.

Bodvarsson ÖB, Partridge MD (2001) A supply and demand model of co-worker, employer and customer discrimination. *Labour Economics* 8(3):389–416.

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

Botelho TL, DeCelles KA (2023) The customer cancellation gap: The drivers of racial/ethnic disparities in on-demand work. *Working Paper* 1–55.

Botelho TL, Gertsberg M (2022) The disciplining effect of status: Evaluator status awards and observed gender bias in evaluations. *Management Science* 68(7):5311–5329.

Brewster ZW, Lynn M (2014) Black–white earnings gap among restaurant servers: A replication, extension, and exploration of consumer racial discrimination in tipping. *Sociological Inquiry* 84(4):545–569.

Chakraborty I, Kim M, Sudhir K (2022) Attribute sentiment scoring with online text reviews: Accounting for language structure and missing attributes. *Journal of Marketing Research* 59(3):600–622.

Cui R, Li J, Zhang DJ (2020) Reducing discrimination with reviews in the sharing economy: Evidence from field experiments on airbnb. *Management Science* 66(3):1071–1094.

Edelman B, Luca M, Svirsky D (2017) Racial discrimination in the sharing economy: Evidence from a field experiment. *American economic journal: applied economics* 9(2):1–22.

Hannák A, Wagner C, Garcia D, Mislove A, Strohmaier M, Wilson C (2017) Bias in online freelance marketplaces: Evidence from taskrabbit and fiverr. *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*, 1914–1933.

Hanrahan BV, Ma NF, Yuan CW (2018) The roots of bias on uber. *arXiv preprint arXiv:1803.08579* .

Kricheli-Katz T, Regev T (2016) How many cents on the dollar? women and men in product markets. *Science advances* 2(2):e1500599.

Kuppuswamy V, Younkin P (2020) Testing the theory of consumer discrimination as an explanation for the lack of minority hiring in hollywood films. *Management Science* 66(3):1227–1247.

Lang K, Spitzer AKL (2020) Race discrimination: An economic perspective. *Journal of Economic Perspectives* 34(2):68–89.

Lynn M, Sturman M (2011) Is the customer always right? the potential for racial bias in customer evaluations of employee performance. *Journal of Applied Social Psychology* 41(9):2312–2324.

Marchenko A (2019) The impact of host race and gender on prices on airbnb. *Journal of Housing Economics* 46:101635.

Mayzlin D, Dover Y, Chevalier J (2014) Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review* 104(8):2421–2455.

Oliver M, Shapiro T (2013) *Black wealth/white wealth: A new perspective on racial inequality* (Routledge).

Owens J (2022) Double jeopardy: Teacher biases, racialized organizations, and the production of racial/ethnic disparities in school discipline. *American Sociological Review* 87(6):1007–1048.

Owens J, McLanahan SS (2020) Unpacking the drivers of racial disparities in school suspension and expulsion. *Social Forces* 98(4):1548–1577.

Pager D, Shepherd H (2008) The sociology of discrimination: Racial discrimination in employment, housing, credit, and consumer markets. *Annu. Rev. Sociol* 34:181–209.

Phelps ES (1972) The statistical theory of racism and sexism. *The american economic review* 62(4):659–661.

Rosenblat A, Levy K, Barocas S, Hwang T (2016) Discriminating tastes: Customer ratings as vehicles for bias. *Data & Society* 1–21.

Shapiro TM, et al. (2004) *The hidden cost of being African American: How wealth perpetuates inequality* (Oxford University Press, USA).

Subramanian A (2019) *The caste of merit: Engineering education in India* (Harvard University Press).

Tomaskovic-Devey D, Thomas M, Johnson K (2005) Race and the accumulation of human capital across the career: A theoretical model and fixed-effects application. *American Journal of Sociology* 111(1):58–89.

Yinger J (1995) *Closed doors, opportunities lost: The continuing costs of housing discrimination* (Russell Sage Foundation).

Younkin P, Kuppuswamy V (2016) Is the crowd colorblind? founder race and performance in crowdfunding. *Academy of Management Proceedings*, volume 2016, 11665 (Academy of Management Briarcliff Manor, NY 10510).

Zhang S, Mehta N, Singh PV, Srinivasan K (2021) Frontiers: Can an artificial intelligence algorithm mitigate racial economic inequality? an analysis in the context of airbnb. *Marketing Science* 40(5):813–820.

Appendix. Communications with Customers

Figure A.1 Screenshots of Emails

(a) Confirmation Email after Worker Accepts Job

(b) Rating Reminder Email after Job is Completed