Intelligence and Autonomy

October 2016

Discriminating Tastes: Customer Ratings as Vehicles for Bias

Alex Rosenblat
Data & Society, alex@datasociety.net

Karen Levy
Cornell University, karen.levy@cornell.edu

Solon Barocas
Microsoft Research, solon@microsoft.com

Tim Hwang
Data & Society, tim@datasociety.net

Edited by Patrick Davison

Abstract:

Consumer-sourced rating systems are a dominant method of worker evaluation in platform-based work. These systems facilitate the semi-automated management of large, disaggregated workforces, and the rapid growth of service platforms—but may also represent a potential backdoor to employment discrimination. Our paper analyzes the Uber platform as a case study to explore how bias may creep into evaluations of drivers through consumer-sourced rating systems. A good deal of social science research suggests that aggregated consumer ratings are likely to be inflected with biases against members of legally protected groups. While companies are legally prohibited from making employment decisions based on protected characteristics of workers, their reliance on potentially biased consumer ratings to make material determinations may nonetheless lead to disparate impact in employment outcomes. Hence, the mediating role of the rating system opens the door to employment discrimination.

We analyze the limitations of current civil rights law to address this issue, and outline a number of operational, legal, and design-based interventions that might assist in so doing. The analysis highlights how innovative work structures challenge traditional legal frameworks, and require creative design, development, operation, and regulation to ensure that they do not facilitate discriminatory outcomes against historically disadvantaged groups.

This project was partly supported by a generous grant from the John D. and Catherine T. MacArthur Foundation.

We are grateful for the insights and assistance of danah boyd, Stacy Abder, Annabelle Gawer, Benjamin Edelman, Alexandra Mateescu, Idil Ali, Angie Waller, Andrew Selbst, Rachel Goodman and participants in IPP2016: The Platform Society at the Oxford Internet Institute.

Illustration by Alexandra Mateescu

Introduction

The on-demand economy has coalesced around platform-based companies that connect service providers with consumers through an intermediary app or digital matching service. Uber is the most contested success among these: the service is popular and available globally, but faces numerous lawsuits concerning its practices in the United States and abroad (Brown, 2016). These companies' distributed employment systems rely on a range of remote, electronic, and automated management techniques. One significant facet of these systems is the use of consumer-facing ratings for worker evaluation, which hand the task of evaluating workers' performance to consumers (Lee et. al, 2015, p. 1603; Rosenblat & Stark, 2016; Brown et. al, 2016; Raval & Dourish, 2016). This method is baked into the design of the Uber app: consumers (passengers) are prompted to rate Uber drivers, and drivers are prompted to rate passengers, as a component of every service interaction.[1]

The growth of software platforms as workplace infrastructures (Gorbis & Fidler, 2016) prompts us to consider the fit between these new forms of work and the legal protections that attend more traditional workplaces—specifically, laws ensuring nondiscrimination in employment. This paper will use Uber as a case study for how rating systems may redistribute supervisory or managerial roles over worker performance to consumers, and how this redesigns the distribution of potential liability for prospective workplace discrimination. The relevance of this analysis extends to management processes that outsource performance evaluations onto consumers, particularly in

1. The present analysis considers the issue of potential bias in *passengers'* ratings of *drivers*, not *drivers'* ratings of *passengers*. While bias may indeed inflect drivers' ratings of passengers as well, this issue raises distinct and complex legal issues that are outside the

interactive service jobs. It has specific implications for the viability and liability of platform-mediated business models that rely heavily on consumer evaluations to maintain quality control over a large, distributed, and disaggregated workforce.

This paper presents an introduction to Uber, its employment practices, and the formal operations of its driver rating system. The first section explains how this rating system is designed to work as one managerial component of Uber's broader platform. The second section offers a critique of the rating system based on an analysis of Uber's corporate policies and research into the system's effect on individual drivers. Both of these sections draw on Rosenblat's ongoing, qualitative fieldwork—a combination of ethnographic work conducted in online forums and through interviews with drivers that began in December 2014 and continues through the present (September 2016). In the third section, we examine the status of the driver rating system relative to the legal framework surrounding employment discrimination. We conclude by offering a list of potential future interventions, and a discussion of how this case study might lay the groundwork for the assessment of other instances of consumer-ratings-based employment determinations in the ever-growing on-demand economy.

Part I : How The Uber Rating System Works

Uber was founded in 2009; as of April 2016, the company managed 450,000 drivers who are active each month in the United States (Uber Newsroom, 2016a). It operates in 208 cities in North America alone (Uber, 2016) and 68 countries globally (Uber Newsroom, 2016b). Uber driver retention rates are low, and its rapid expansion has thus far relied on a constant flood of new drivers: slightly more than half of drivers on-boarded in the U.S. in 2013 remain active (having completed at least one trip in the previous 6 months) on the platform a little over a year later, according to Uber's own data (Hall & Krueger, 2015, p. 16).

These dynamics raise the question: how can a constant flood of new workers be adequately supervised by a platform-employer? Uber's primary answer is its rating system, which is a scalable solution to maintaining quality control over a far-flung and fluctuating workforce.[2]

After every Uber-mediated ride, passengers are prompted to rate drivers on a 1- to 5-star scale, and are given the option to add specific comments

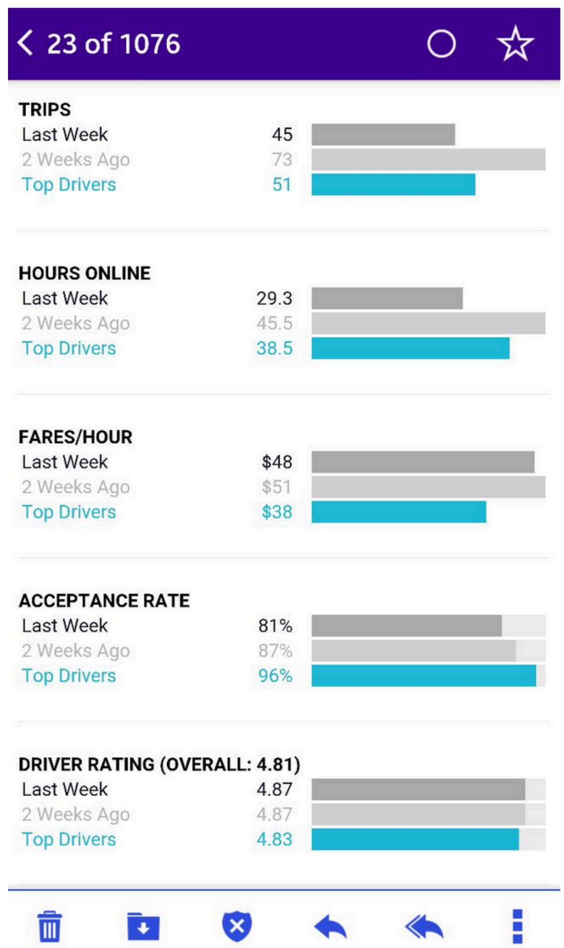

Fig 1. A sample performance evaluation received by a driver.

2. Two additional, important metrics for driver evaluations are their ride acceptance rates, and their cancellation rates.

on driver performance.[3] In Uber's system, this consumer feedback generates instantaneous evaluations that allow Uber to track worker performance and intervene with poor performers. These ratings serve as the basis for deactivation notices or suggestions for improvement to under-performing drivers, such as "Never ask for a 5-star review, but focus instead on providing an excellent experience," or "Riders count on Uber for a comfortable, relaxing experience. They prefer for drivers not to promote other businesses during the trip." Drivers are not shown which passenger gave them which rating (to protect passenger privacy), but the total tallies of rated trips, the driver's average rating, and the total number of 5-star trips are displayed to each driver. In order to remain active on the system, drivers must meet an average rating target that hovers around 4.6 out of 5 stars. Uber's policy is that drivers who fall below regional performance targets risk deactivation (temporary suspension or permanent termination) from the system.[4]

Typically, a driver's overall rating reflects an average of his or her last 500 rated trips, although drivers have received deactivation warnings when the average rating for only their last 25 or 50 trips dipped too low (Rosenblat, 2015). While the rating system alerts Uber to drivers who are under-performing, it also provides a context though which Uber can communicate desired behaviors to its drivers. This can come in the form of generic notices that list a series of errant or

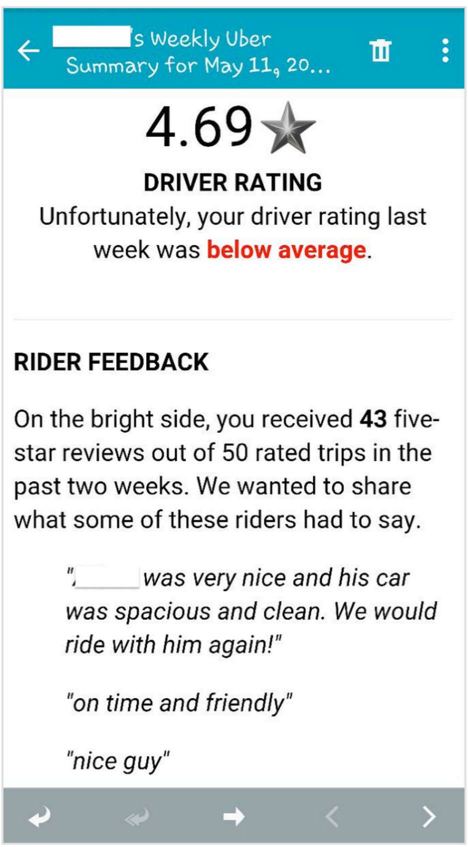

Fig 2. Example of weekly rating message sent to drivers.

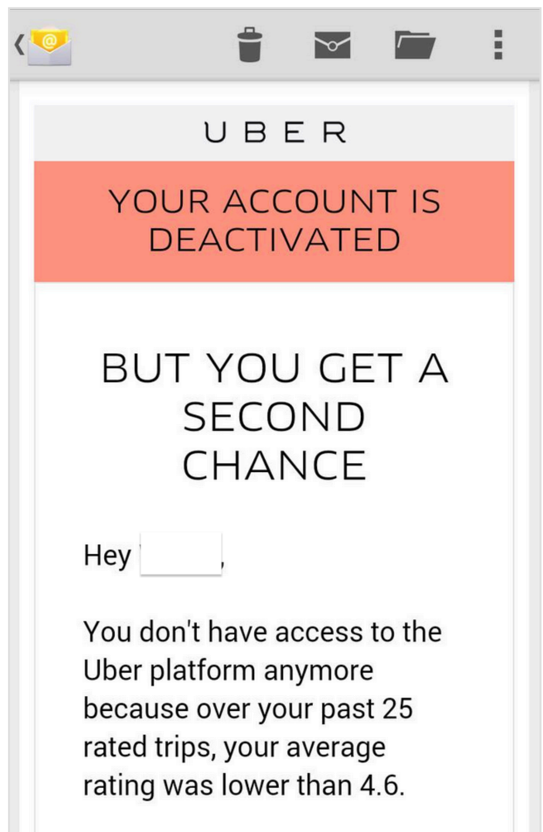

Fig 3. A sample deactivation notice.

3. Uber alters its platform periodically and for different user groups; this description of the rating process is the general current method, but may vary for some users based on experimentation or evolution over time.
4. Drivers are compared with other drivers in their local market (Uber does not publicly define the boundaries of those markets), rather than with all drivers globally. This, sensibly, maintains quality control through localized expectations. Rating targets also vary by Uber service. For example, uberX and uberXL might require a minimum rating of 4.6/5, but UberBlack and UberSUV in the same market might require a minimum of 4.7/5.

desirable behaviors, framed as "Problems Reported: There were a few things riders in your city commonly reported. Here are some tips on how to improve," or that reference specific user comments, such as "You received a 'Talks Too Much' complaint." Drivers who receive this notice are directed to a website that gives them detailed advice on rider interactions, such as "If they don't seem to be engaging in conversation, then silence could be key here."

Drivers routinely receive weekly performance evaluations that highlight their overall rating, their rating for the last week, their rating for the last two weeks, and the ratings of "top drivers" or "top partners" for comparison. In these notices, a driver's rating is highlighted at the very top of his or her notice, in a large font, and is designed to emphasize its crucial importance in the driver's performance evaluation, often with phrases such as "Unfortunately, your driver rating last week was below average."

An advantage of this method for worker evaluation can be that the system encourages accountability, similar to the ways in which businesses are rated on Yelp or products are rated on Amazon. This view is evident even among drivers, some of whom explain that the reasons Uber drivers maintain a friendly demeanor, offer candy and snacks, keep their vehicles clean, and generally perceive that they provide better service than traditional taxis is substantially because of their concern for their rating (Campbell, 2015). The view that the rating system holds drivers, and to a lesser extent, passengers, accountable for good behavior is also widespread in media coverage of Uber's service (Ondraskova, 2015; Motroc, 2015; Smith IV, 2015). Drivers also acknowledge that Uber is able to influence how drivers behave, particularly in relation to regular taxis because ratings prompt them to keep cleaner vehicles and to be polite in passenger interactions, as a function of the rating system (Rosenblat & Stark, 2016, p. 3775). Crucially, the indirectness of Uber's control over driver behavior is important, from Uber's position, for maintaining the argument that drivers are best classified as independent contractors because their employers are limited in the degree that they may control independent contractors' work practices (Rosenblat & Stark, 2016; Kessler, 2016).

Fig. 4. A sample explanatory flier in the backseat of an Uber car, which was posted to a driver forum online.

Part II : How The Uber Rating System Impacts Drivers

In Uber's driver-rating model, consumers are empowered to act, in part, as middle-managers of workers (Stark & Levy, 2015), both through the design of the app and in the evaluation[5] functions they perform (Rosenblat & Stark, 2016, p. 3772). Ratings, as a reflection of consumer preferences, allow companies to institutionalize those consumer preferences if they use them as direct assessments of worker performance.

5. Ratings are one of three main metrics that act as performance targets: high ride acceptance rates (such as 80% or 90%) and low cancellation rates (such as 5%) are the other two (Rosenblat & Stark, 2015, p. 11).

The reputations that workers develop on platforms through rating systems (on Uber and elsewhere in the on-demand economy) directly impact workers' earnings and, particularly, opportunities for higher-paid work. Take for example Uber's incentive-based wage structures. Uber sometimes offers select drivers guaranteed hourly pay at higher rates, such as $22/hr, if they opt-in or "RSVP" to the guarantee. The conditions for receiving this guarantee follow a typical template: accept 90% of ride requests, complete one trip per hour, be online for at least 50 minutes of every hour, and maintain a specified high rating during those trips. While the criteria Uber uses to select drivers who are "invited" to participate in higher earning shifts isn't disclosed by the company, drivers are required to maintain a high rating during the "guarantee" periods that they participate in, or they lose the guaranteed amount. (When contacted by the researchers, Uber declined to disclose the criteria by which drivers are selected and invited to participate in hourly guarantees).[6] The rating system therefore determines not just the basis for firing, but also the qualification for higher-paying wages.

While the driver rating system is designed to mediate accountability among riders, drivers, and the company (Brown et. al, 2016), its implementation can have other consequences for drivers. Drivers in a previous study of Uber drivers (Rosenblat & Stark 2016) expressed frustration and anxiety about their ratings – which inevitably seemed to decline at some point – because drivers were often not able to identify what, if anything, had changed in their performance. Passengers are not generally educated on Uber's rating system and may presume that 4 out of 5 stars is a good rating, even though such a score is actually a "failing grade" for drivers (Rosenblat & Stark, 2016, p. 3775; Raval & Dourish, 2016, p. 5). It is likely that passengers' ratings of drivers tend toward extremes – 5 stars or 1 star, rather than an intermediate rating (Hu, Zhang, and Pavlou, 2009) – and that Uber's cutoff threshold is therefore set at a point that seems unreasonably high, even though it may reflect a lack of precision in *consumers'* ratings. Some drivers make attempts to educate passengers on the realities of driver ratings in conversation, or by nudging them with explanatory fliers in their backseat (Rosenblat & Stark, 2016, p. 16).

Many drivers express that they aren't always sure what they are being rated on, and many have tried to compensate for anticipated negative ratings by offering snacks, water, or a phone-charger cord (Rosenblat & Stark, 2016, p. 3775; Raval & Dourish, 2016). The uniformity of this behavior may stem partly from Uber's training videos, which explicitly recommend that 5-star-aspiring drivers provide bottled water or phone chargers (Uber Driver Training Video, 5:54-6:30). Drivers perform other emotional labor for good ratings: they offer to adjust music, the temperature, evaluate whether the passenger wants to engage in or disengage from conversation, and in some cases, find something (anything) to apologize for (see also Hochschild, 2003).

Because the Uber system is designed and marketed as a seamless experience (Uber Newsroom, 2015b), and coupled with confusion over what driver ratings are for, any friction during a ride can cause passengers to channel their frustrations with the Uber system as a whole into the ratings that primarily impact an individual driver's employment eligibility. Some drivers observe that they receive low ratings in response to a variety of things outside of their control, including: surge pricing; GPS or navigation malfunctions; the passenger's misplacement of their own location pin for pick-up; holding passengers in compliance with both Uber's rules and local laws, such as not taking more passengers than there are seatbelts in the vehicle (Rosenblat & Stark, 2016, p. 3775).

6. It's possible driver ratings are part of the selection criteria that determines who is "invited" into higher earnings. Thus, the rating system can be a tool for producing tiered wages for drivers that could theoretically produce wage discrepancies for similar performances by drivers if the ratings contain negative bias towards select drivers with protected-class characteristics.

In some markets, Uber has recognized that drivers receive lower ratings when prices surge (see Figure 5). Surge pricing means that the base rate is multiplied by a factor, such as 1.5–9.5×, determined algorithmically, according to Uber, based on the levels of supply (drivers) and demand (passengers) (Uber Help, 2016). Uber has informed drivers in some markets that “ratings on high surge trips will also not be taken into consideration” (see Fig. 5). However, surge seems to be the only exception to how ratings are weighted, and the criteria for what constitutes a “high surge” (as compared to a normal surge) are not transparent.

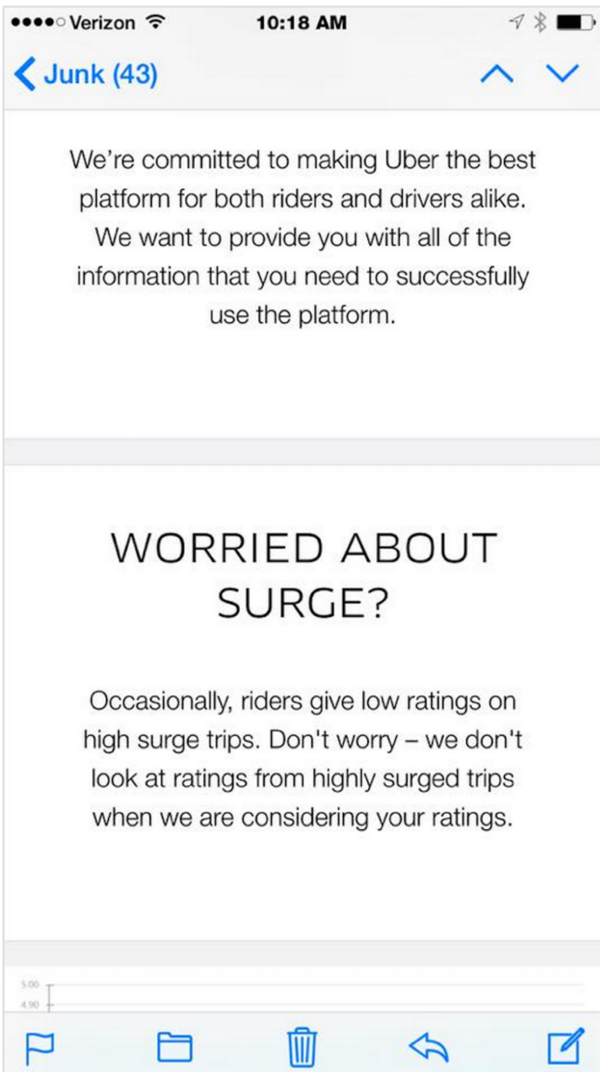

Fig 5: A message received by a driver from Uber.

Bias in Ratings: Evidence from Social Science

Consumer-sourced ratings like those used by Uber are highly likely to be influenced by bias on the basis of factors like race or ethnicity. If a platform bases material employment determinations on such ratings, these systems – while appearing outwardly neutral – can operate as vehicles through which consumer bias can adversely impact protected groups.

Without analysis of (and access to) both ratings data and information about worker characteristics, we cannot determine unequivocally whether consumer-sourced ratings (e.g., passengers’ ratings of Uber drivers) are, in the aggregate, racially biased. But we have ample reason to believe that there is a strong likelihood they would be so. A plethora of social science research has established that racial and gender bias commonly “creeps into” ratings of all sorts. Much of this research concerns two relevant topics: consumer behavior in online marketplaces, and performance evaluations of workers in managerial settings. We focus on these topics because platform-based, consumer-sourced ratings represent the confluence of the two. We describe each briefly below.

In online markets, consumer behavior often exhibits bias based on the perceived race of another party to the exchange; this bias often manifests via lower offer prices and decreased response rates. Researchers have a long-standing interest in online markets because these platforms offer a convenient way to conduct field experiments, but also because evidence continues to mount suggesting that bias persists even in such mediated settings (Shohat & Musch, 2003). For example, one study of iPod sales on Craigslist found that, when photos of the iPod for sale showed the item in a black person's hand, the listing garnered fewer offers to purchase the iPod, and at lower prices, than when a white person's hand was shown (Doleac et al., 2013). An experiment involving baseball card auctions on eBay found a very similar pattern of bias (Ayres et al., 2015), as did a study of Prosper.com, a peer-to-peer online lending platform (Pope & Sydnor, 2011).

More recently, a study of Airbnb, an online marketplace for short-term housing rentals, found that guests with African-American names were about 16% less likely to be accepted as rentees than guests with characteristically white names (Edelman et al., 2016). A complementary study focused on hosts on Airbnb found that Asians earned 20% less than their white equivalents in Oakland, California (Wang et al., 2015).

The dynamics of implicit and explicit bias have also been addressed in social science research about managerial performance evaluations of workers. A wealth of studies demonstrate that racial, gender, and nationality biases impact managerial ratings of workers (Castilla, 2008; Mobley, 1982; Elvira and Town, 2001; Kraiger and Ford, 1985). This work suggests that multiple mechanisms can account for how bias can influence evaluations. Supervisors may render a higher level of scrutiny in evaluating workers with protected-class characteristics (Stauffer & Buckley, 2005). Homophily (shared characteristics) between managers and workers may positively influence managers' ratings of worker performance, suggesting that workers perceived as being different from the evaluator may compare unfavorably (Castilla, 2011). In Uber's case, any biases held by passengers may be funneled through the ratings model feedback mechanism and could have a disproportionately adverse impact on drivers who, for example, are people of color. Passengers might implicitly rate minority drivers less charitably if, for instance, their self-presentation fails to emulate perceived white, middle-class norms (Rogers, 2015). Whether riders are less generous with or more critical of drivers who happen to be members of a protected class is an open empirical question which researchers outside Uber are not well positioned to answer.

Through a rating system, consumers can directly assert their preferences and biases in ways that companies would be prohibited from doing directly.[7] In effect, companies may perpetuate bias without being liable for it, as the grounds for firing or "deactivating" a particular driver may be derived from a large corpus of individual ratings, whose discriminatory character is currently impossible to verify or oversee by researchers external to the company.

Part III : Legal Status of The Uber Rating System

Uber's rating system may present a facially neutral[8] route for discrimination to "creep in" to employment decisions, because customers may have a tendency to give systematically biased ratings to drivers based on protected characteristics like a driver's race, and because Uber bases employment decisions (including termination) in large part on customer ratings.

7. Thanks to Dr. Benjamin Edelman for an illuminating discussion on this topic.
8. "Facially neutral" refers to a protocol or process that does not consider a protected class membership explicitly, like gender, but which nevertheless has the effect of harming members of these groups.

It is important to note that Uber is somewhat unique in that drivers' continued employment is directly tied to customer ratings; this is in contrast to other platforms, such as Airbnb and eBay, in which user ratings provide information to other market participants (or possibly influence search rankings) upon which they may decide with whom to transact, but do not necessarily result in termination.

While managers may be similarly discriminatory in their employment decision-making in more traditional contexts, platforms that make *their own* employment decisions on the basis of customers' ratings (rather than leaving platform-based market participants to potentially discriminate against one another directly) do not fall as cleanly under existing discrimination law. As scholars have established, the legal protections against discrimination usually available to workers (under Title VII of the Civil Rights Act of 1964) may be difficult to apply when customer-sourced ratings drive employment determinations (Wang, 2016). Thus, the discriminatory harms that emerge in the customer rating context may be less easily addressed than those in a manager-employee context.

Title VII suits may be brought under either a disparate treatment or disparate impact theory. In the former, the claimant must demonstrate that the employer possesses a discriminatory intent or motive. In the latter, a facially neutral employment practice may be challenged because it causes a substantial adverse impact to a protected group, regardless of the employer's intent. Title VII jurisprudence clearly establishes that employee scoring, testing, and other forms of evaluation may serve as the basis for a successful Title VII claim, if such evaluation disparately impacts a protected group, and cannot be justified by the employer.

In the case of Uber's driver rating system, the applicable analysis will likely be disparate impact. We assume that Uber is not knowingly and purposefully catering to the race-based preferences of riders, and is not otherwise motivated by discriminatory intent.

As a preliminary matter, it appears unlikely that a platform like Uber would be able to claim a defense under Section 230 of the Communications Decency Act ("CDA 230"), which grants immunity to platforms for content created by third-parties. Uber might make the argument here that discrimination is the result of content generated purely by riders on the platform, thereby excusing the platform itself from liability.

However, under CDA 230, platforms do not receive this immunity if they are "information content providers," that is, that they are "responsible in whole or in part, for the creation or development" of infringing material. Under the rule in *Fair Housing Council of San Fernando Valley v. Roommates.com, LLC*, affirmative solicitation of discriminatory information is sufficent to exclude immunity. In that case, Room-mates.com provided dropdown menus for users to provide information about gender, sexual orientation, and other information. Plaintiffs argued successfully in that case that CDA 230 did not apply because of the active role that Room-mates.com played in "providing content" which was the basis for discrimination. This is the case here: by establishing the bounds of the rating system, Uber plays an active role in eliciting the content which is infringing, even though it does not create the content itself.

The central legal question, then, is whether platforms should be liable for making employment decisions on the basis of facially neutral, but potentially discriminatory, consumer-sourced ratings. We suggest that a disparate impact claim on this basis would likely be unsuccessful, due to three formidable, independent hurdles in the path of prospective plaintiffs. This raises the concern that existing federal law may make it difficult for workers who are indirectly subject to discriminatory employment practices through consumer ratings to receive a remedy.

Hurdle 1: Employment Status

The first challenge regards the legal classification of Uber drivers as employees or independent contractors. Title VII protections attach to employees but not independent contractors.[9] Uber currently classifies its drivers as independent contractors, although this status is being challenged by a pending employment misclassification class-action lawsuit in California (Gibson, Dunn, & Crutcher, 2015) that alleges they should be classified as employees. If the relationship between the employer and the independent contractor indicates that the former exerts significant control, as Uber drivers allege in the lawsuit, the independent contractor may be considered an "employee," thus triggering the protections of Title VII (Rubinstein, 2012, p. 617; see also *Salamon v. Our Lady of Victory Hosp.*). This burden may vary depending on the platform. Although Title VII may not apply at present to Uber's independent contractors, and the scope of its applicability relies heavily on the outcome of the employee misclassification lawsuit, the issues we highlight are applicable to other companies that use Uber as a template and imitate its management structures, like the rating system. These companies may have different contractual arrangements with its employees that make it more or less likely that employment status presents a formidable barrier to bringing a claim. In general, the unsettled nature of labor classification with respect to platform-based companies imposes a significant hurdle on prospective plaintiffs bringing a Title VII suit premised on discriminatory consumer-sourced ratings.

Hurdle 2: Job Relatedness

Secondly, even if litigants were classified as employees, Uber may be able to defend its consumer rating system as a business necessity. Under Title VII doctrine, an employer may avoid liability for an employment practice that creates a disparate impact if they are able to meet the burden of showing that the practice is "job related for the position in question and consistent with business necessity."

Uber's case for job-relatedness is straightforward: the ratings are a means of evaluating the quality of a given driver. The company would also be able to mount a further case arguing for the "business necessity" of its potentially discriminatory rating system. Under the existing arrangement, Uber riders rate Uber drivers; drivers with sufficiently low ratings are then suspended or fired, a process labeled "deactivation." While such a management technique creates the opportunity for Uber to refuse to retain drivers who have been subject to discriminatory assessment by riders, it is precisely this distributed rating system which allows Uber to manage a large, geographically distributed, and transitory population of 1.1 million workers worldwide.

Hence, Uber could potentially assert that the scalability of its model constitutes a business necessity that demands a consumer-sourced ratings system. The adequacy of this assertion would be a factual question litigated in a potential Title VII suit. If the employer is able to show

9. However, some state laws may be construed as extending antidiscrimination protections to contractors. See, e.g., Minn. Stat. 363A.17(3).

job relatedness, the burden turns back to the plaintiff to show that less discriminatory alternatives are possible to achieve this business goal.

There are, in principle, many other ways of making employment decisions that could potentially be less discriminatory than the one Uber has adopted, such as the direct administration of periodic driver competency tests, or the installation of specialized hardware to monitor drivers. However, these solutions would arguably shift the dispersed costs of evaluating driver performance from a large pool of riders to a concentrated cost shouldered by Uber. This might impose significant costs on the platform and limit its ability to provide its core service; Uber, then, might argue that it would be unable to realize its fundamental business objective using an alternative evaluation system.

Hurdle 3: Data Availability

A final concern is pragmatic. The ability to bring a Title VII claim successfully relies on a plaintiff's ability to demonstrate the discriminatory impact of the driver rating system in the first place. As described above, the plaintiff would also need to show the presence of less discriminatory alternatives in the scenario where Uber successfully argues at trial that its rating system is justified by its job relatedness. In both cases, the plaintiff lacks access to the needed data to effectively argue these points. This asymmetry of data and data-gathering capacity thus presents a third hurdle to the application of Title VII in addressing the discriminatory impact that might emerge from tying consumer ratings to employment decision-making on these platforms.

Information asymmetry always presents a challenge for plaintiffs in civil rights cases, and it imposes a similar burden here. Practically speaking, it would be very challenging for anyone other than Uber to do the analysis required to investigate disparate impact of protected-class drivers, much less evaluate the impact of alternative designs. Further, the platform is very likely in the best position to implement tests to monitor other factors that might be missed by consumer ratings, such as the longevity of drivers on the platform, the type of vehicle being driven, or otherwise.

However, even in a situation where full data about protected statuses and ratings were available to the plaintiff, it is important to note that a simple statistical analysis would not be able to account for unobserved characteristics that might give rise to disparate ratings, as well as the ambiguous nature of ratings themselves. Ratings are intentionally subjective, and aimed at capturing a customer's general level of satisfaction with a product or service. Because of their generality and subjectivity, there is no *correct* or most accurate rating for a particular interaction; indeed, the customer's experience is "treated as sovereign" (Wang 2016), creating a necessarily ambiguous, general, and subjective metric without clear benchmarks for satisfaction. Put another way, to consider biased ratings to be *in error* overstates the case. If ratings are intended to be truly subjective, even biased judgments that accurately reflect a consumer's general level of satisfaction cannot be understood to be erroneously rendered. And to further complicate the matter, even aside from matters of bias, membership in protected classes may correlate with other aspects of the customer experience that might strike us as more palatable bases for ratings (e.g., language barriers).

Cumulative Burden

The cumulative burden of these three hurdles on plaintiffs seeking redress is striking. In effect, these hurdles render Title VII an ineffective means of ending discriminatory employment practices that may be perpetuated through distributed rating systems like the one seen in platforms like Uber. Given that this is the case, this paper concludes with a set of interventions and research avenues that explore the possible means by which traditional protections granted to workers may be better extended to a new technological environment.

Part IV : Proposed Interventions

The analysis above suggests that it may be extremely difficult, under the current structure of Title VII, for prospective plaintiffs to mount a successful legal challenge against discriminatory employment impact arising from consumer-driven ratings systems. These difficulties are poised to more significantly impede workers' rights as the "gig economy" continues to expand within the United States, and as more platforms pattern themselves on the successes of companies like Uber. Airbnb—which, together with Uber, has become a leading symbol of the "gig economy"—has similarly faced accusations of discrimination on its platform (Edelman, Luca, & Svirsky, 2016).

Based on our analysis of bias on Uber's platform, we propose a set of potential interventions to allow the protections of Title VII to more effectively extend into this new labor environment, and to limit the bias that might affect consumer ratings and the employment decisions that depend upon them. The following proposals are intended as jumping-off points for further exploration, and as a means of laying out potential alternatives to address the threat of employment discrimination in the consumer rating-driven on-demand economy.

Category 1: Establish baseline statistics.

The collection (and possible publication) of descriptive statistics about ratings and employment outcomes among different groups of drivers is an essential first step to determine whether discrepancies on the basis of protected characteristics exist in the first place.

Track patterns in ratings, employment outcomes, and correlation with protected-class status.

A platform could gather data about workers' demographic characteristics, the ratings workers receive from consumers, and employment outcomes. In Uber's case, this could include tracking whether drivers that belong to a protected class are more likely than others to receive low ratings from consumers, whether Uber itself is more likely to issue warnings of potential deactivation to these drivers, or whether they are, in fact, deactivated. Separately, the company could investigate which drivers are recommended for reactivation classes and how those drivers fare when they return to the platform.

Publish disclosure statements about these patterns.

Internal monitoring of employment outcomes vis-à-vis protected class categories could produce its own benefits (e.g., prompting platforms to address apparent, but potentially avoidable disparities in employment decisions, aside from and prior to any legal challenges). However, voluntary or compelled public disclosure of such statistics would serve additional purposes. In the same way that technology companies' recent practice of releasing diversity and inclusion reports is intended to bring public pressure and competitive dynamics to bear on the task of increasing the representations of women and minorities in the workforce, public disclosure regarding disparities in customers' ratings and the related employment decisions could spur companies to more aggressively seek out potential solutions. Such disclosure could, however, also serve potential litigants' interests in that it would make it easier to identify problematic patterns of activity in the first place—and establish a prima facie case of disparate impact in the course of litigation.

Category 2: Evaluate and adjust for data quality.

As described in Part III, the inherent and intentional subjectivity of ratings makes it difficult to conceptualize what a data "quality" adjustment might look like. Even the suggestion that implicit or explicit consumer biases ought not inflect their ratings of workers – or at least, that platforms ought to account and correct for the likely presence of such biases – represents a complex normative judgment, and we must acknowledge that adjustments to correct for bias in this context are therefore more normatively laden than adjustments made to correct for systematic error (e.g., sampling bias) in other data analysis contexts. Despite this ambiguity, certain interventions might be used to "validate" or adjust biased ratings in ways that minimize the impact of potential discrimination.

Validate ratings with behavioral data.

One plausible alternative might be to implement more rigorous checks on biased consumer ratings. Rather than taking ratings at face value, direct measurements of behavior could be used to validate ratings drivers receive from their passengers. If ratings are tied to specific performance criteria – such as driving at an appropriate speed – this could be estimated through the sensor data produced by accelerometers, GPS data, and gyrometers in the drivers' smartphones; video or audio recordings of worker-customer interactions might be used in certain contexts as well. If these data sources do not corroborate ratings data, the ratings might be adjusted or discarded. Uber launched a pilot project in Houston, Texas to track driver movements, which could prospectively expand to improve the signal accuracy of passenger feedback (Sawers, 2016). It has since deployed a method for measuring "safe driving" by using data from driver phones to flag issues like smooth braking and acceleration. More granular data about behavioral activity could reveal whether, say, a driver is speeding and therefore "deserving" of a lower rating. However, such behavioral data will necessarily only capture certain measurable aspects of the customer experience, rather than "thick data" about the experience as a whole—and importantly, the collection of these sorts of data to corroborate or contradict passenger ratings entails more invasive surveillance of drivers' work activities, which may introduce a host of additional legal and ethical concerns.

Weight ratings to account for potential bias.

Another approach would involve statistically weighting ratings data to account for the likelihood of bias on the basis of protected-class membership. This intervention could take different forms (see, for example: Dellarocas, 2000; Whitby, Jøsang, & Indulska, 2004). Most directly, if evidence of bias is found or assumed, the composite ratings of workers could be adjusted upward if they belong to protected classes. Another variation would be to assign lower weight to (or discard) ratings provided by the most biased *raters* (for instance, look for the greatest statistical disparities between ratings assigned by a particular rater to workers inside and outside protected groups, according to matched comparisons based on other observable attributes). In practice and in certain contexts, data sparsity may be an impediment to reliable implementation on this intervention, as well as complications introduced by unobservable characteristics.

Category 3: Design user interfaces to minimize implicit bias.

Design constraints might be used to minimize the role of bias in the ratings process, either by providing raters with *less* information or by gathering *more* information in the face of suspect ratings.

Increase the reporting burden on customers.

Platforms could raise the reporting burden for consumers who give low ratings—for instance, by requiring them to specify the reason for a low rating (e.g., speeding, uncleanliness). The location-based website Nextdoor, which provides a platform for people to report criminal activity (among other things) in a neighborhood, recently implemented design changes in response to complaints that these reporting functions were becoming vehicles for racial profiling by users. Nextdoor took a number of steps to discourage racial profiling via design: for instance, users are specifically prompted not to identify suspicious people solely by race, and if users do specify the race of a suspicious person, they are required to include at least two other identifying characteristics (e.g. hairstyle, clothing; O'Donovan, 2016). One could imagine a platform like Uber selectively increasing the reporting burden along one (or more than one) of several axes: e.g., requiring extra reporting for low ratings; requiring extra reporting for drivers from protected classes; requiring extra reporting from passengers whose rating patterns suggest the possibility of biased assessments. By increasing the reporting burden on users for determinations that are likely to be inflected by bias, platforms may be able to increase users' reflection on the criteria that drive their ratings; in addition, the information collected might provide specific guidance to the driver or platform on the reason for the low rating, which might inform a concrete change in behavior.

Reduce the information available to raters.

Conversely, a platform could withhold information from customers, such as eliminating workers' profile names in anticipation that names without salient racial associations will elicit less biased customer ratings, or by minimizing or eliminating the use of photographs of the person to be evaluated. Such strategies for mitigating against bias and preventing discrimination have a long history in hiring and housing, and are well supported by scholarship on their potential in the gig economy, specifically on Airbnb (Edelman, 2016). Such an approach may be of limited utility on a platform like Uber, in which the rater and ratee interact in person before the rating occurs; however, on other platforms in which interactions are more attenuated, it might serve such purposes. Reflecting on earlier research documenting that Airbnb users exhibited discriminatory bias in deciding who to host, Edelman (2016) has argued that the platform could limit the opportunity for users to discriminate by withholding guests' photos from hosts prior to confirming the booking. Airbnb has stated that it plans to explore ways to reduce the prominence of photos on its website, but remains committed to including them in the initial encounter between potential hosts and guests, arguing that they are necessary for establishing trust and community.

In some contexts, removing indicators of the characteristics against which users discriminate may, counterintuitively, create additional risks for members of those groups. For instance, if indicators are removed prior to a face-to-face interaction, some users may be placed in particularly vulnerable situations that spark harassment or even violence against them from customers with biased sentiments. Therefore, platforms need to be sensitive to how the degree of information available to both parties before an exchange is likely to impact social interactions between them; it may be salient to consider issues like the duration and nature of service provision.

Category 4: Resituate the use of ratings within organizational structures.

Biased ratings are problematic if they are relied upon to terminate or otherwise materially affect employment outcomes for workers, but companies might rethink the use of ratings, or use them alongside other evaluation techniques, to avoid such difficulties. These interventions imply reliance on alternative workplace evaluation processes, which could present challenges to the scalability of platform-based management.

Decouple ratings from employment determinations.

Customer-sourced ratings may serve useful business purposes even divorced from material employment outcomes. Hence, platforms might collect ratings data from consumers but *not* use it for purposes of evaluating workers; for instance, ratings might be maintained as part of the user experience on a platform, or as a generalized barometer of customer satisfaction. They might be used to inform a worker or the platform about her/his performance, but not formally fed into workplace evaluation processes (which would need to be based on alternative methods). This would eliminate the relationship between ratings and employment that give rise to concern under Title VII.

Implement more robust in-person escalation.

Rather than using consumer ratings as a substitute for more traditional worker evaluations, they might be used in tandem with a more traditional evaluation system. One option might be to leverage trained human evaluators, who could investigate and provide a direct assessment of a worker who receives low consumer-sourced ratings. (As we suggest in Part III, such an intervention might run contrary to a platform's business model premised on the cost savings that come from deferring evaluations of workers to consumers.)

Allow buyers and sellers to find one another.

Platforms can assume a more passive role in helping supply find demand. Rather than actively matching, in Uber's case, drivers and riders, the platform could function more like an open market, allowing its users to rely on ratings to make decisions about whether to transact with other parties. Drivers could receive information about nearby riders; riders, in turn, could learn about nearby drivers; both parties could then decide with whom they would like to transact—and at what price. This would turn Uber into something much more similar to eBay or Airbnb, potentially relieving the company of having to make any employment decisions of its own. As an intermediary, Uber and similarly structured platforms could escape the reach of discrimination law. Though this intervention is not aimed at actually reducing bias in consumers' ratings (and the discriminatory outcomes to which they may lead), it could reduce the risk of liability for intermediary platforms.

Category 5: Alter legal frameworks.

The interventions described thus far attempt to address the potential discriminatory impact of consumer ratings by altering the structure or procedures of the platform itself. But other interventions could be addressed by changing the structure of Title VII. These interventions could enable those harmed by discriminatory ratings to more effectively exercise rights under the law in light of the hurdles to doing so described earlier in the paper.

Reclassify workers.

As discussed, one immediate and daunting hurdle facing workers attempting to bring a challenge against a discriminatory rating system is the fact that the legal classification of workers as independent contractors precludes them from leveraging the protections of Title VII. One approach may be to formally reclassify on-demand economy workers as employees, or as a class of workers who are protected by law that would subject them to protections against discrimination. This route will require a careful assessment of whether rating systems are a business necessity generally, and specifically in which form or iteration.

Modify pleading requirements.

Recognizing the challenges in demonstrating a disparate impact on employment and in proving less discriminatory alternatives in this new technological context, one approach may be to lower the pleading requirements for claims brought against these types of platforms. This would increase the likelihood that a litigant would be able to survive preliminary motions and leverage the process of discovery to level the playing field with the platform. Along these lines, legal scholars (Wang, 2016; Zatz, 2009) have suggested ways that anti-discrimination law might better account for discrimination arising from outside the employer-worker dyad (e.g., from customer ratings), in light of the delegation of management responsibility to customers and service providers that arises due to the use of such evaluation mechanisms.

We anticipate that these approaches implemented independently may be necessary but insufficient to readily address consumer-sourced ratings bias. Certain approaches might work best in concert with one another (e.g., increasing the reporting burden on raters for suspect ratings *combined with* validation based on behavioral data collection). In addition, any intervention or set of interventions needs to be sensitive to the specific technical and organizational contexts within which a platform is situated. Bias impacts platforms in different ways, and each case thus requires context-specific analyses and modes of intervention (Edelman, 2016).

Conclusion

The discrimination issues raised in this article are relevant to a broader selection of companies than Uber alone. While Uber is used as the case study here, all firms that leverage customer feedback, particularly those that belong to the on-demand economy, risk assuming the biases of their customers in their worker evaluations (Fuller & Smith, 1991). Uber's case throws into sharper relief the consideration we should give to the role and responsibility of companies that are positioned as intermediaries. The need to exercise quality control over a large disaggregated workforce may, however, permit the continued use of rating systems under existing employment discrimination law, even in cases where doing so has a manifest disparate impact on members of protected classes. Uber and similarly structured companies could argue that consumer preferences as expressed in star ratings are "job related" factors that companies are well justified in considering in employment decisions. They could further defend the use of rating systems as a "business necessity," given the scale of their business—with no obvious alternative method for achieving the same business goal.

These issues are potentially latent in any automated system that employs an ad hoc, distributed labor force regulated largely by consumer feedback. As Uber-like models continue to multiply, employment discrimination may become hotly contested political ground, joining existing debates over whether or not workers should be classified as employees or contractors. Models leveraging consumer ratings as feedback systems for guiding autonomous mechanisms of worker control should be considered in the context of this emerging risk as much as they are seen in light of the enormous economic value they enable.

Employer liability for consumer preferences should depend on how much the system is designed to, or in effect does, rely on consumer feedback to determine the employability of workers. "Business necessity" as the pivot of the legal analysis under Title VII in these situations is significant because it spurs a discussion of what less discriminatory alternatives could and should be in these business models. Each of the proposed interventions will distribute costs and benefits across all players in the system: riders, drivers, and the platform itself.

The issue of driver ratings presents challenges in terms of determining the extent of discriminatory practices, and proposing effective changes to the discriminatory practices given the complicated technical, legal, and social realities of the system described in this paper. Maintaining fair labor practices under these conditions will require creative thinking about how to design, develop, operate, and regulate these platforms.

Works Cited

Ayres, I., M. Banaji, and C. Jolls. 2015. “Race effects on eBay.” The RAND Journal of Economics 46, no. 4 (2015): 891-917.

Barocas, S., and A. Selbst. 2015. “Big Data’s Disparate Impact.” California Law Review, Vol. 104, 2016.

Beinstein, A. 2016. “How Uber Engineering Increases Safe Driving With Telematics.” Uber Engineering. June 29. https://eng.uber.com/telematics/.

Belzer, A., and N. Leong. 2016. “The New Public Accommodations.” 105 Georgetown Law Journal (2017), Forthcoming. Available at SSRN: http://ssrn.com/abstract=2687486.

Bhuiyan, J. 2015. “Why Is Uber New York Funneling Thousands Of Drivers To This Training Class?” BuzzFeedNews. March 29. http://www.buzzfeed.com/johanabhuiyan/why-is-uber-new-york-funneling-thousands-of-drivers-to-this#.ymmW8GxQQ.

Brown, K.V. 2016. “Uber is facing a staggering number of lawsuits.” Fusion. January 1. http://fusion.net/story/257423/everyone-is-suing-uber/.

Brown, B., M. McGregor, M. Glöss, A. Lampinen. 2016. “On-Demand Taxi Driving: Labour Conditions, Surveillance, and Exclusion.” Paper presented at the 2016 The Internet, Policy & Politics Conferences, Oxford, UK. http://ipp.oii.ox.ac.uk/sites/ipp/files/documents/McGregor_Uber%2520paper%2520Sept%25201%2520PDF.pdf.

Campbell, H. 2015. “My Rating System For Uber and Lyft Passengers.” The Ride Share Guy. August 31. http://therideshareguy.com/my-rating-system-for-uber-and-lyft-passengers/.

Cantor, R.L. 1999. “Consumer Preferences for Sex and Title VII: Employing Market Definition Analysis For Evaluating BFOQ Defenses.” University of Chicago Legal Forum, Vol. 1999, Issue 1, Article 12. http://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=1275&context=uclf.

Dellarocas, C. 2000. “Immunizing online reputation reporting systems against unfair ratings and discriminatory behavior.” In Proceedings of the 2nd ACM Conference on Electronic Commerce. 150-157.

Diaz v. Pan Am. World Airways, Inc. 442 F2d 385 (5th Cir 1971).

Doleac, J.L., and L.C.D. Stein. “The visible hand: Race and online market outcomes.” The Economic Journal 123, no. 572 (2013): F469-F492.

Edelman, B., M. Luca, and D. Svirsky. 2016. “Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment.” American Economic Journal: Applied Economics, Forthcoming. http://www.benedelman.org/publications/airbnb-guest-discrimination-2016-01-06.pdf.

Edelman, B. 2016. “Responses to Airbnb’s Report on Discrimination.” September 19, 2016. http://www.benedelman.org/news/091916-1.html.

Elvira, M.M., and R. Town. 2001. “The Effects of Race and Worker Productivity on Performance Evaluations.” Industrial Relations 40 (4): 571–90.

Fair Housing Council of San Fernando Valley v. Roommates.com, LLC, 521 F.3d 1157 (9th Cir. 2008).

Feldman, M., S. Friedler, J. Moeller, C. Scheidegger, and S. Venkatasubramanian. 2015. “Certifying and removing disparate impact.” July 16. http://arxiv.org/abs/1412.3756.

Fuller, L., and V. Smith. 1991. “Consumers’ Reports: Management by Customers in a Changing Economy.” 5 WORK, EMPLOYMENT & SOCIETY 1, 2.

Gibson, Dunn, and Crutcher LLP. 2015. “DEFENDANT UBER TECHNOLOGIES, INC.’S OPPOSITION TO PLAINTIFFS’ MOTION FOR CLASS CERTIFICATION.” August 6. http://www.scribd.com/doc/271072643/BRIEF-Uber-Opposition-to-Motion-for-Class-Certification-1.

Grover, S.S. 1999. “The Business Necessity Defense In Disparate Impact Discrimination Cases.” Faculty Publications, Paper 19. http://scholarship.law.wm.edu/cgi/viewcontent.cgi?article=1038&context=facpubs.

Hochschild, A.R. 2003. The Managed Heart: Commercialization of Human Feeling (2nd ed.). Berkeley, CA: University of California Press.

Kane, K. 2015. “The Big Hidden Problem Problem With Uber? Insincere 5-Star Ratings.” Wired. March 19. http://www.wired.com/2015/03/bogus-uber-reviews/.

Kessler, S. 2016. “Why A New Generation Of On-Demand Businesses Rejected The Uber Model.” Fast Company. March 29. https://www.fastcompany.com/3058299/why-a-new-generation-of-on-demand-businesses-rejected-the-uber-model.

Kraiger, K., and J.K. Ford. "A meta-analysis of ratee race effects in performance ratings." Journal of applied psychology 70.1 (1985): 56.

Leidner, R. 1999. “Emotional Labor in Service Work.” Annals of the American Academy of Political and Social Science.

Lee, M.K., D. Kusbit, E. Metsky, and L. Dabbish. 2015. “Working with Machines: The impact of algorithmic, data-driven management on human worker.” In Proceedings of the 33rd Annual ACM SIGCHI Conference. Seoul, South Korea: ACM, 1603–1612. http://dl.acm.org/citation.cfm?id=2702548.

Luca, M., and G. Zervas. 2013. “Fake It Till You Make It: Reputation, Competition and Yelp Review Fraud.” https://consumermediallc.files.wordpress.com/2014/09/ssrn-id2293164.pdf.

Mobley, W.H. 1982. “Supervisor and Employee Race Effects on Performance Appraisals: A Field Study of Adverse Impact and Generalizability.” Academy of Management Journal 25 (3): 598–606.

Motroc, G. 2015. “Ratings system makes UberX ‘superior’ experience to taxis.” Australian National Review. http://www.australiannationalreview.com/ratings-system-uberx-superior-experience-taxis/.

Newcomer, E. 2015. “Uber Raises Funding at $62.5 Billion Valuation.” Bloomberg Technology. Decemeber 3. http://www.bloomberg.com/news/articles/2015-12-03/uber-raises-funding-at-62-5-valuation.

Nunley, J.M., M.F. Owens, and R.S. Howard. 2011. “The effects of information and competition on racial discrimination: Evidence from a field experiment.” Journal of Economic Behavior & Organization, Vol. 80, Issue 3: 670-679. http://www.sciencedirect.com/science/article/pii/S0167268111001739.

Ondraskova, D. 2015. “Despite Opponents’ Attacks, Uber is Just What America Needed.” Townhall. December 2. http://townhall.com/columnists/danniondraskova/2015/12/02/despite-opponents-attacks-uber-is-just-what-americans-need-n2088042/page/full.

Raiser, LLC. 2015. “Technology Services Agreement.” December 11. https://s3.amazonaws.com/uber-regulatory-documents/country/united_states/RASIER%20Technology%20Services%20Agreement%20Decmeber%2010%202015.pdf.

Raval, N. and P. Dourish. 2016. “Standing Out from the Crowd: Emotional Labor, Body Labor, and Temporal Labor in Ridesharing.” Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing. New York: ACM, 97-107. https://dl.acm.org/citation.cfm?id=2820026.

Rogers, B. 2015. “The Social Costs of Uber.” The University of Chicago Law Review Dialogue 82:85. https://lawreview.uchicago.edu/sites/lawreview.uchicago.edu/files/uploads/Dialogue/Rogers_Dialogue.pdf.

Rosenblat, A. 2015. “The Future of Work: For Uber Drivers, Data Is the Boss.” Pacific Standard Magazine. September 10. http://www.psmag.com/business-economics/the-future-of-work-for-uber-drivers-data-is-the-boss.

Rosenblat, A., and L. Stark. 2016. “Algorithmic Labor and Information Asymmetries: A Case Study of Uber’s Drivers.” International Journal Of Communication, 10, 27. http://ijoc.org/index.php/ijoc/article/view/4892/1739.

Rubinstein, M.H. 2012. “Employees, Employers, and Quasi-Employers: An Analysis of Employees and Employers Who Operate In The Borderline And Between An Employer-And-Employee Relationship.” http://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=1411&context=jbl.

Salamon v. Our Lady of Victory Hosp., No. 06-1707-cv, 2008 WL 2609712, at *1061 (2d Cir. Jan. 16, 2008).

Sawers, P. 2016. “Uber’s, Using gyrometer and accelerometer data from smartphones to verify user feedback.” Venture Beat. January 16. http://venturebeat.com/2016/01/26/ubers-using-gyrometer-and-accelerometer-data-from-smartphones-to-verify-user-feedback/.

Shohat, M., and J. Musch. 2003. "Online auctions as a research tool: A field experiment on ethnic discrimination." Swiss Journal of Psychology/Schweizerische Zeitschrift für Psychologie/Revue Suisse de Psychologie 62, (2): 139.

Smith IV, J. 2015. “Uber Drivers: The Punishment For Bad Ratings Is Costly Training Courses.” The Observer. February 3. http://observer.com/2015/02/uber-drivers-the-punishment-for-bad-ratings-is-costly-training-courses/.

Stark, L. and K. Levy. 2015. “The Consumer as Surveillor.” Paper presented at the 2015 Privacy Law Scholars Conference, June 2–3, Berkeley, CA.

Stauffer, J.M., and M.R. Buckley. 2005. “The Existence and Nature of Racial Bias in Supervisory Ratings.” 90 J. OF APPLIED PSYCHOLOGY 586.

Sweeney, L. 2013. “Discrimination in Online Ad Delivery.” January 28. http://ssrn.com/abstract=2208240.

Uber. 2015a. “How Are Ratings Calculated?” https://help.uber.com/h/66ce3340-aa1f-4357-b955-027ef50441d3.

Uber. 2015b. “Ratings Are A Two-Way Street.” http://sanfrancisco.ubermovement.com/resources/ratings-are-a-two-way-street/.

Uber. 2016. “Find a City.” https://www.uber.com/cities/.

Uber B.V. 2015. “Services Agreement.” https://s3.amazonaws.com/uber-regulatory-documents/country/united_kingdom/Uber+BV+Driver+Terms+-+UK+Preview.pdf.

Uber Help. 2016. “What Is Surge Pricing?” https://help.uber.com/h/6c8065cf-5535-4a8b-9940-d292ffdce119.

Uber Melbourne. 2016. “Training Video.” https://drive.uber.com/melbourne/how-can-we-help/how-to-uber/training-video/.

Uber Movement. 2016. “Get Started.” http://sanfrancisco.ubermovement.com/get-started/.

Uber Newsroom. 2015a. “New Survey: Drivers Choose Uber for its Flexibility and Convenience.” https://newsroom.uber.com/driver-partner-survey/.

Uber Newsroom. 2015b. “Enabling Seamless Pickups through Color Coding.” https://newsroom.uber.com/us-washington/enabling-seamless-pickups-through-color-coding/.

Uber Newsroom. 2016a. “Growing and growing up.” April 21. https://newsroom.uber.com/growing-and-growing-up/.

Uber Newsroom. 2016b. “Celebrating Cities: A New Look and Feel for Uber.” February 2. https://newsroom.uber.com/celebrating-cities-a-new-look-and-feel-for-uber/.

Uber South Florida. 2016. “Help Center: Deactivated Drivers.” http://ubersouthflorida.com/for-deactivated-partners/.

Wang D., S. Xi, and J. Gilheany. 2015. “The Model Minority? Not on Airbnb.com: A Hedonic Pricing Model to Quantify Racial Bias against Asian Americans.” Technology Science. September 1. http://techscience.org/a/2015090104/.

Wang, L. 2016. “When the Customer Is King: Employment Discrimination as Customer Service.” U. of Pittsburgh Legal Studies Research Paper No. 2016-01. http://ssrn.com/abstract=2657758.

Whitby, A., Jøsang, A., and Indulska, J. 2004. “Filtering out unfair ratings in bayesian reputation systems.” In Proc. 7th Int. Workshop on Trust in Agent Societies. Vol. 6: 106-117.

Workplace Fairness. 2016. “Your Rights Independent Contractors.” https://www.workplacefairness.org/independent-contractors#13.

Zatz, N. 2016. “Is Uber Wagging the Dog With Its Moonlighting Drivers?” onlabor. January 19. http://onlabor.org/2016/01/19/beyond-misclassification-gig-economy-discrimination-outside-employment-law/.