Policy and Internet

Discriminating Tastes: Uber's Customer Ratings as Vehicles for Workplace Discrimination

Alex Rosenblat, Karen E.C. Levy, Solon Barocas, and Tim Hwang

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 avenue for employment discrimination that negatively impacts members of legally protected groups. We analyze the Uber platform as a case study to explore how bias may creep into evaluations of drivers through consumer-sourced rating systems, and draw on social science research to demonstrate how such bias emerges in other types of rating and evaluation systems. 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 a disparate impact in employment outcomes. 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.

KEY WORDS: platforms, data, discrimination, bias, inequality, ratings, sharing economy, algorithm

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. These companies' distributed employment systems rely on a range of remote, electronic, and automated-management techniques. Uber is both one of the most successful and most controversial of these platforms: The service is popular and available globally, but faces numerous lawsuits concerning its practices in the United States and abroad (Brown, 2016).

One significant feature of distributed employment systems, including Uber's, is the use of consumer-facing ratings for worker evaluation, which hands the task of evaluating workers' performance to consumers (Brown, McGregor, Glöss, & Lampinen, 2016; Lee, Kusbit, Metsky, & Dabbish, 2015, p. 1603; Raval & Dourish, 2016; Rosenblat & Stark, 2016). This method is baked

doi: 10.1002/poi3.153

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 article uses Uber as a case study to examine how rating systems may redistribute supervisory or managerial roles over worker performance to consumers, and how this may impact potential liability for prospective workplace discrimination. The relevance of this analysis extends to management processes that outsource performance evaluations onto consumers, particularly in 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.

Although recent scholarship and policy documents have explored the potential for data-driven decision making to discriminate, as well as the limits of existing law to deal with such problems (Barocas & Selbst, 2015; Federal Trade Commission, 2016; Kim, 2017; White House, 2014, 2016), they have not considered consumer reviews as a source of potential bias, nor the distinct challenges that such behavior poses for addressing discrimination through standard legal mechanisms. The primary concern in the literature to date has been the potential for automated decision making to inherit the biases expressed in the human decisions that serve as examples from which computers are instructed to learn a decision procedure. Consumer ratings pose what might seem like a similar problem, insofar as employment decisions that take such ratings into account run the risk of inheriting consumers' biases—but, as this article demonstrates, platforms that are purposefully structured to make consumer ratings the key input to automated employment decisions present unique legal and practical difficulties.

This article offers an overview of Uber, its employment practices, and the formal operations of its driver rating system. The first section explains how the rating system is designed to work as one managerial component of Uber's 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 effects 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. 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, 2016b). It operates in 208 cities in North America (Uber, 2016) and 68 countries globally (Uber Newsroom, 2016a). 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 United States 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, 2016, 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 strategy is its rating system, which is a scalable solution to maintaining quality control over a far-flung and fluctuating workforce.

After every Uber-mediated ride, passengers are prompted to rate drivers on a one- to five-star scale, and are given the option to add specific comments on driver performance.[2] 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 underperforming drivers. 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 five-star trips are displayed to each driver. To remain active on the system, drivers must meet an average rating target that hovers around 4.6 out of five stars. Uber's policy is that drivers who fall below regional performance targets risk deactivation (temporary suspension or permanent termination) from the system.[3] 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 their last 25 or 50 trips has dipped too low (Rosenblat, 2015).

While the rating system alerts Uber to drivers who are underperforming, it also provides a context through which Uber can communicate desired behaviors to its drivers. This can come in the form of notices that list a series of errant or 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 2 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, to emphasize its crucial importance in the driver's performance evaluation, often

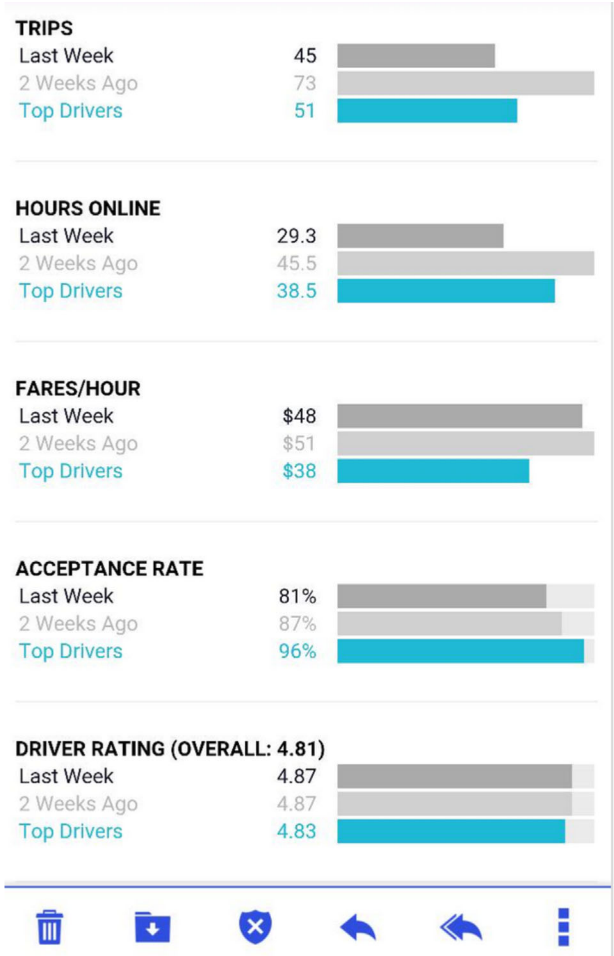

Figure 1. A Sample Performance Evaluation Received by a Driver.

with phrases such as "Unfortunately, your driver rating last week was below average" (Figures 1 and 2).

An advantage of this method for worker evaluation can be that the system encourages accountability; a similar justification underlies rating systems across platforms, from Amazon to Yelp. 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 a better service than traditional taxis is substantially because of their concern for their rating (Brown et al., 2016; 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 (Motroc, 2015; Ondraskova, 2015; Smith, 2015). Drivers acknowledge that Uber is able to influence how drivers behave,

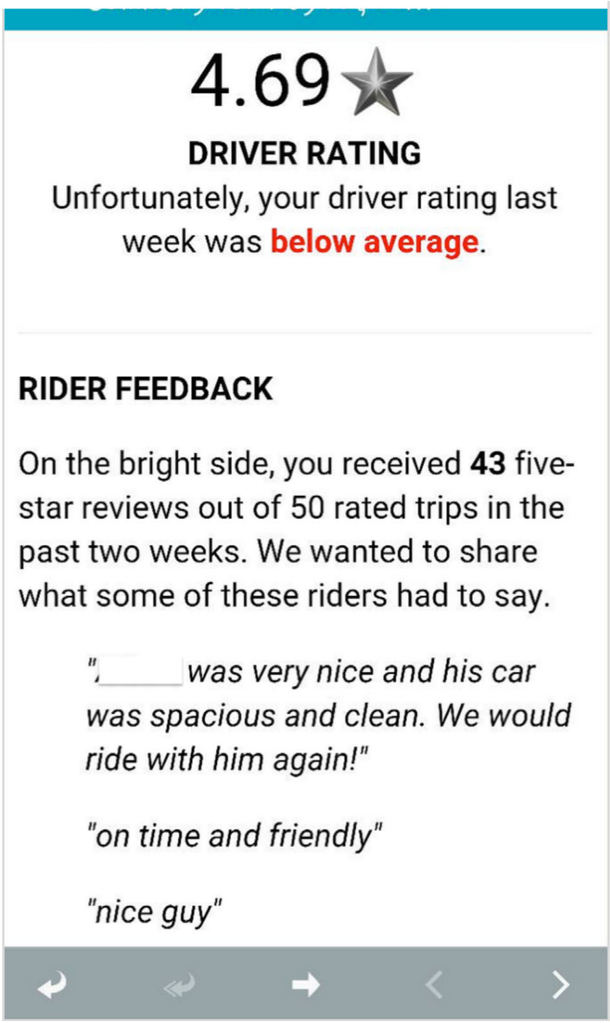

Figure 2. Example of Weekly Rating Message Sent to Drivers.

sometimes by asserting the superiority of Uber to regular taxis, such as through cleaner vehicles or more polite interactions, as a function of the rating system (Rosenblat & Stark, 2016, p. 3775). For Uber, the indirectness of Uber's control over driver behavior is also important for maintaining the argument that drivers are best classified as independent contractors, rather than employees (Kessler, 2016; Rosenblat & Stark, 2016).

Part II. How the Uber Rating System Impacts Drivers

In Uber's driver-rating model, consumers act as middle managers of workers (Stark & Levy, 2015), both through the design of the app and in the evaluation[4] functions they perform (Rosenblat & Stark, 2016, p. 3772). Ratings, as a reflection of consumer preferences, allow companies to institutionalize those preferences by using them as assessments of worker performance.

The reputations that workers develop on platforms through rating systems (on Uber and elsewhere in the on-demand economy) can directly impact their earnings and 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 if they opt-in or "RSVP" to a guarantee. The conditions for receiving this guarantee follow a typical template: accept 90 percent 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 is not 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).[5] The rating system, therefore, determines not just the basis for termination, but also qualification for higher wages (Figures 3 and 4).

Although the driver rating system is designed to mediate accountability among riders, drivers, and the company, its implementation can have other consequences for drivers. In a previous study (Rosenblat & Stark, 2016), Uber drivers 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 four out of five stars is a good rating, even though such a score is actually a "failing grade" for drivers (Raval & Dourish, 2016, p. 101; Rosenblat & Stark, 2016, p. 3775). It is likely that passengers' ratings of drivers tend toward extremes—five stars or one star, rather than an intermediate rating (Hu, Zhang, & 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. 3775).

Many drivers express that they are not always sure what they are being rated on, and have tried to compensate for anticipated negative ratings by offering snacks, water, or a phone-charger cord (Raval & Dourish, 2016; Rosenblat & Stark, 2016, p. 3775). 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; Uber Melbourne, 2016). Drivers also 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, 2015), and coupled with confusion over what driver ratings are for, any friction during a ride can cause passengers to channel their frustrations

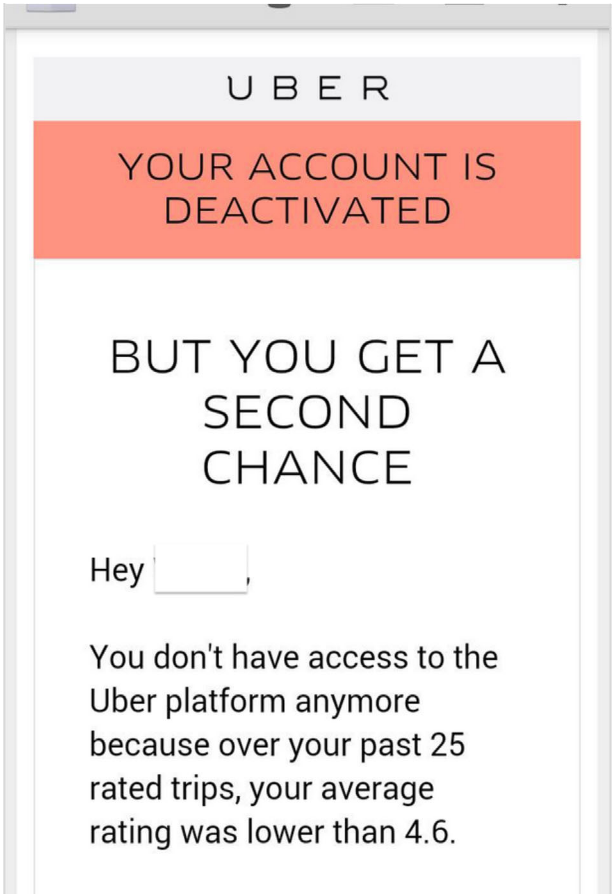

Figure 3. A Sample Deactivation Notice.

with the Uber system as a whole into the ratings that impact an individual driver. 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; and holding passengers in compliance with both Uber's rules and local laws, such as not taking more passengers than there are seat belts in the vehicle (Rosenblat & Stark, 2016, p. 3775).

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; according to Uber, the multiplier is algorithmically determined 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 Figure 5). However, surge seems to be the only revealed exception to how ratings

Figure 4. A Sample Explanatory Flier in the Backseat of an Uber Car, Which Was Posted to a Driver Forum Online.

are weighted, and the criteria for what constitutes a "high surge" (as compared to a normal surge) are not transparent.

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 there is a strong likelihood that 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

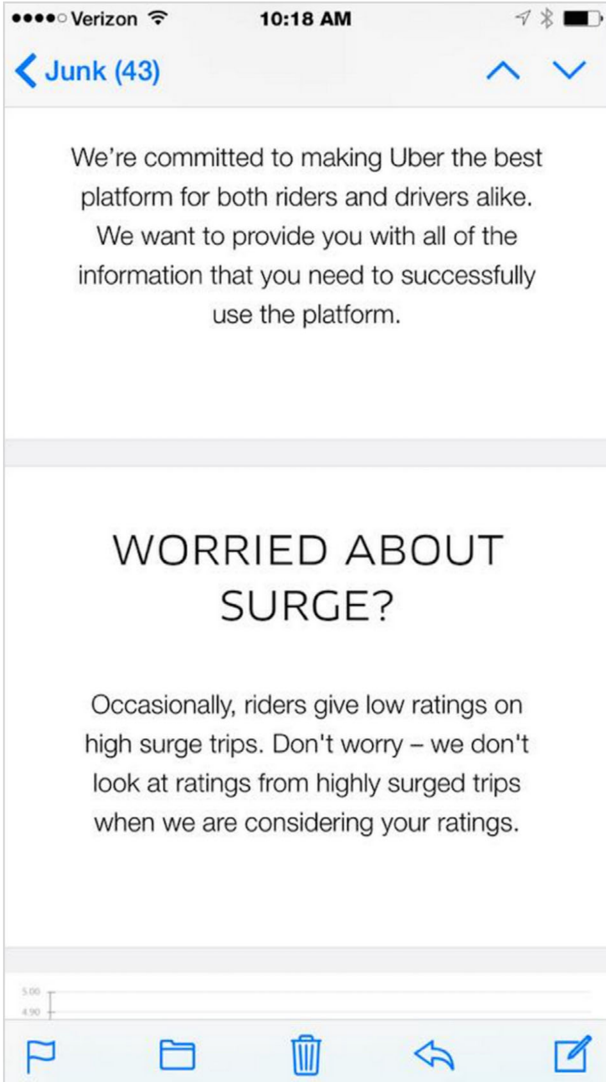

Figure 5. A Message Received by a Driver From Uber.

fewer offers to purchase the iPod, and at lower prices, than when a white person's hand was shown (Doleac & Stein, 2013). An experiment involving baseball card auctions on eBay found a very similar pattern of bias (Ayres, Banaji, & Jolls, 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 percent less likely to be accepted as rentees than guests with characteristically white names (Edelman, Luca, & Svirsky, 2016). A complementary study that focused on hosts on Airbnb found that Asian hosts earned 20 percent less than similarly situated white hosts in Oakland, California (Wang, Xi, & Gilheany, 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; Elvira & Town, 2001; Kraiger & Ford, 1985; Mobley, 1982). 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 that researchers outside Uber are not well positioned to answer.

Part III. Legal Status of the Uber Rating System

Uber's rating system may, thus, present a facially neutral[6] route for discrimination to "creep in" to employment decisions. 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 systemically biased consumer ratings. Importantly, Uber is somewhat distinct in that drivers' continued employment is directly tied to customer ratings; this stands 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.

Platforms that make employment decisions on the basis of customer ratings (rather than leaving platform-based market participants to potentially discriminate against one another directly) do not fall cleanly under existing discrimination law. The legal protections against discrimination usually available to U.S. 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, discriminatory harms that emerge in the customer rating context may be less easily addressed than those in a more traditional 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 an 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, CDA 230 is inapplicable because users of the platform are not engaged in illegal conduct; the ratings generated by customers (even if biased) are not in themselves discriminatory acts. Rather, it is Uber's use of riders' ratings to make employment decisions about drivers that gives rise to discrimination.[8]

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 CA (Gibson & Crutcher, 2015) that alleges they should be classified as employees. If the relationship between an employer and a worker indicates that the former exerts significant control over the latter's performance of work tasks, as Uber drivers allege in the lawsuit, the worker 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.*, 2008). The extent of 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: Business Necessity and Less Discriminatory Alternatives

Second, even if litigants were classified as employees, Uber may still be able to defend its consumer rating system. Under Title VII, 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." In its rulings, the Supreme Court has conflated these two elements into a single requirement, noting that "[t]he touchstone is business necessity. If an employer practice ... cannot be shown to be related to job performance, the practice is prohibited" (*Griggs v. Duke Power Co.*, 1971). While there exist a handful of different legal tests that courts apply to validate a practice as a business necessity, most relevant for the purposes of this analysis is criterion validation, which "requires the defendant to establish a statistically significant correlation between good performance on a test and good performance on the job according to some accepted criterion" (Rutherglen, 1987, p. 1317).

The adequacy of this assertion would be a factual question litigated in a potential Title VII suit. Courts "[refuse] to accept bare or 'commonsense'-based assertions of business necessity and instead [require] some level of empirical proof that challenged hiring criteria accurately predicted job performance" (*El v. Southeastern Pennsylvania Transportation Authority*, 2007, p. 240). Uber should be able to successfully meet this burden based on data it is already collecting from its platform: it seems probable that consumer ratings will bear some correlation in aggregate with a range of different job performance variables such as rider satisfaction, driver safety, or successful trip completion.

If the employer is able to show business necessity, 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.

But even here, Uber may have a defense. While in general the costs of eliminating a discriminatory practice are ignored, courts have suggested that a narrow exception exists where "the expense would threaten the very survival of the defendant's business" (Grover, 1996, p. 398). 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 form of rating system which allows Uber to manage a large, geographically distributed, and transitory population of 1.1 million workers worldwide.

Many less discriminatory alternatives 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 prevent the platform from providing 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 in Hurdle 2 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 others.

However, even in a situation in which 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 aim to capture 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. Though normatively—and in light of the very real harms of discriminatory judgments based on customer bias—we might want to design systems to mitigate against biased consumer ratings, this is a different endeavor than trying to correct for error therein. Put another way, to consider biased ratings to be somehow "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. They render Title VII an ineffective means of ending discriminatory employment practices that may be perpetuated through Uber's rating systems and similar mechanisms on other online platforms. Given the limited utility of existing law, this article concludes with a set of interventions and research avenues that explore other possible means by which traditional protections granted to workers may be better extended to online platforms' novel employment practices.

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 discrimination 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.[10]

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. Importantly, the following proposals are intended as a set of provocations for further reflection rather than recommended policy prescriptions, 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,[11] the ratings workers receive from consumers, and employment outcomes (and a government could obligate it to do so, as, e.g., the Equal Employment

Opportunity Commission does with certain employers). In Uber's case, this could include tracking whether drivers who 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 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 standard data analysis. Despite this ambiguity, certain interventions might be used to "validate" or adjust biased ratings in ways that minimize any potential discriminatory impact.

Validate Ratings With Other Data About User Behavior. 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, TX 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 behavior 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, e.g., 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 of this intervention, as well as complications introduced by unobservable characteristics. Even adjustments that lead to some kind of nondiscriminatory "bottom line" may not fully insulate an employer from liability for a particular practice that has a disparate impact on a protected class (see *Connecticut v. Teal*, 1982).

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 assessments.

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, etc.). 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 by redesigning the reporting interface: 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: for example, 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 them 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 be 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).

Restructure the Intermediary's Role to 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 or termination decisions of its own. As such an intermediary, Uber and similarly structured platforms could escape the reach of discrimination law.

Importantly, this intervention is aimed at repositioning the role of the platform with respect to its users to reduce the risk of liability for employment discrimination, rather than reducing bias in consumers' ratings (and the discriminatory outcomes to which they may lead). In practice, such an intervention may not operate to ameliorate those harms—or might even exacerbate them. However, we note it here as it is within the range of solutions companies could adopt to further insulate themselves from responsibility for bias in users' interactions.

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 article.

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 antidiscrimination 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.

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 sharp 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 differently across all players in the system: riders, drivers, and the platform itself.

Distributed rating systems on online platforms offer much promise, but also potentially create new avenues to discriminatory outcomes. Maintaining fair labor practices under these conditions will require creative thinking about how to design, develop, operate, and regulate such platforms.

Alex Rosenblat, M.A., Data & Society Research Institute, New York [alex@datasociety.net].

Karen E.C. Levy, J.D., Ph.D., Department of Information Science, Cornell University, Ithaca, New York.

Solon Barocas, Ph.D., Microsoft Research, New York; Department of Information Science, Cornell University, Ithaca, New York.

Tim Hwang, J.D., Microsoft Research, New York; Data & Society Research Institute, New York.

Notes

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 scope of our inquiry, in part due to the fact that distinct legal regimes apply. For related discussion, see Belzer and Leong (2016).
2. 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.
3. 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.
4. Ratings are one of three main metrics that act as performance targets: high ride acceptance rates (such as 80 percent or 90 percent) and low cancellation rates (such as 5 percent) are the other two (Rosenblat & Stark, 2016, p. 3761).
5. It is 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 toward select drivers with protected-class characteristics.
6. "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.
7. Thanks to Benjamin Edelman for an illuminating discussion on this topic.
8. The situation here stands in contrast to cases like *Fair Housing Council of San Fernando Valley v. Roommates.com, LLC* (2008), which established that affirmative solicitation of discriminatory information is sufficient to override immunity for platforms. In that case, Roommates.com provided dropdown menus for users to provide information about gender, sexual orientation, and other information, in ways that violated the Fair Housing Act. Plaintiffs argued successfully in that case that CDA 230 did not apply because of the active role that Roommates.com played in "providing content" which was the basis for discrimination. The Uber case is distinguishable both because Uber is unlikely to qualify as a "content provider" under the terms of the statute, and because they are not facilitating unlawful action by the site's *users*, as was the case in Roommates.com—rather, it is the platform's *own* actions as an employer that might run afoul of discrimination law.
9. However, some state laws may be construed as extending antidiscrimination protections to contractors. See, for example, Minn. Stat. 363A.17(3) available at https://www.revisor.mn.gov/statutes/?id=363a.17. In addition, some courts have implemented an "economic realities" test that considers in part the degree to which workers depend economically on the business to which they provide services (see, e.g., *Real v. Driscoll Strawberry Assocs., Inc.*, 1979). Under all applicable tests, however, the degree to which the employer controls the "means and manner of the worker's performance" is the most important factor (*Spirides v. Reinhardt*, 1979).
10. Airbnb—which, together with Uber, has become a leading symbol of the "gig economy"—has similarly faced accusations of discrimination on its platform (Edelman et al., 2016).
11. Uber appears to do this already (Vaccaro, 2017): "According to Uber's own national data, as of 2014, about 40 percent of drivers were white, 19.5 percent were black, 17.7 percent Hispanic, and 16.5 percent Asian."

References

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

Barocas, S., and A. Selbst. 2015. "Big Data's Disparate Impact." *California Law Review* 104: 2016.

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

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

Brown, K.V. 2016. "Uber Is Facing a Staggering Number of Lawsuits." *Fusion*. January 1. http://fusion.net/story/257423/everyone-is-suing-uber/.

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/.

Castilla, E.J. 2008. "Gender, Race, and Meritocracy in Organizational Careers." *American Journal of Sociology* 113 (6): 1479–526.

Castilla, E.J. 2011. "Bringing Managers Back In: Managerial Influences on Workplace Inequality." *American Sociological Review* 76 (5): 667–94.

Connecticut v. Teal, 457 U.S. 440 (1982).

Dellarocas, C. 2000. "Immunizing Online Reputation Reporting Systems Against Unfair Ratings and Discriminatory Behavior." In *Proceedings of the 2nd ACM Conference on Electronic Commerce*. New York: ACM, 150–7. http://dl.acm.org/citation.cfm?id=352889.

Doleac, J.L., and L.C.D. Stein. 2013. "The Visible Hand: Race and Online Market Outcomes." *The Economic Journal* 123 (572): F469–92. http://onlinelibrary.wiley.com/doi/10.1111/ecoj.12082/full.

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

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.

El v. Southeastern Pennsylvania Transportation Authority, 479 F.3d 232 (3rd Circuit, 2007).

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).

Federal Trade Commission. 2016. "Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues." FTC Report. https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf.

Fuller, L., and V. Smith. 1991. "Consumers' Reports: Management by Customers in a Changing Economy." *Work Employment & Society* 5 (1): 1–12. https://www.researchgate.net/publication/238437205_Consumers'_Reports_Management_by_Customers_in_a_Changing_Economy.

Gibson, D., 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.

Gorbis, M., and D. Fidler. 2016. "Design It Like Our Livelihoods Depended on It." *Medium*. April 4. https://medium.com/the-wtf-economy/design-it-like-our-livelihoods-depend-on-it-e1b6388eb752.

Griggs v. Duke Power Co., 401 U.S. 424 (1971).

Grover, S.S. 1996. "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.

Hall, J., and A.B. Krueger. 2016. "An Analysis of the Labor Market for Uber's Driver-Partners in the United States." NBER Working Paper No. w22843. http://www.nber.org/papers/w22843.

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

Hu, N., J. Zhang, and P.A. Pavlou 2009. "Overcoming the J-Shaped Distribution of Product Reviews." *Communications of the ACM* 52 (10): 144–7. http://dl.acm.org/citation.cfm?id=1562800.

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.

Kim, P.T. 2017. "Data-Driven Discrimination at Work." *58 Wm. & Mary L. Rev. 857*. https://litigation-essentials.lexisnexis.com/webcd/app?action=DocumentDisplay&crawlid=1&doctype=cite&docid=58+Wm.+%26+Mary+L.+Rev.+857&srctype=smi&srcid=3B15&key=c1d342281062842 2e6e08bfd15f152ce.

Kraiger, K., and J.K. Ford. 1985. "A Meta-Analysis of Ratee Race Effects in Performance Ratings." *Journal of Applied Psychology* 70 (1): 56–65. https://scholars.opb.msu.edu/en/publications/a-meta-analysis-of-ratee-race-effects-in-performance-ratings-4.

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, 1603–12. http://dl.acm.org/citation.cfm?id=2702548.

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/.

O'Donovan, C. 2016. "Nextdoor Rolls Out Product Fix It Hopes Will Stem Racial Profiling." *BuzzFeedNews*. August 24. https://www.buzzfeed.com/carolineodonovan/nextdoor-rolls-out-product-fix-it-hopes-will-stem-racial-pro?utm_term=.kfZ5GORyd#.yxQ6GRM1p.

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.

Pope, D.G., and J.R. Sydnor. 2011. "What's in a Picture? Evidence of Discrimination From Prosper.com." *Journal of Human Resources* 46 (1): 53–92.

Raval, N. and P. Dourish. 2016. "Standing Out From the Crowd: Emotional Labor, Body Labor, and Temporal Labor in Ridesharing." In *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.

Real v. Driscoll Strawberry Associates, Inc., 603 F.2d 748 (9th Cir. 1979).

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: 3758–84. 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.

Rutherglen, G. 1987. "Disparate Impact Under Title VII: An Objective Theory of Discrimination." *73 Va. L. Rev. 1297*. https://litigation-essentials.lexisnexis.com/webcd/app?action=DocumentDisplay&crawlid=1&doctype=cite&docid=73+Va.+L.+Rev.+1297&srctype=smi&srcid=3B15&key=4a99031e92258cc796206544769bc0bd.

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, J., IV. 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/.

Spirides v. Reinhardt, 613 F.2d 826 (D. C. Cir. 1979).

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." *Journal of Applied Psychology* 90 (3): 586–91. https://utpb.influuent.utsystem.edu/en/publications/the-existence-and-nature-of-racial-bias-in-supervisory-ratings.

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

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 Newsroom. 2015. *Enabling Seamless Pickups Through Color Coding*. https://newsroom.uber.com/us-washington/enabling-seamless-pickups-through-color-coding/.

Uber Newsroom. 2016a. *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 Newsroom. 2016b. *Growing and Growing Up*. April 21. https://newsroom.uber.com/growing-and-growing-up/.

Vaccaro, A. 2017. "MIT Professor Working With Uber to Address Racial Bias." *Boston Globe*. February 15. https://www.bostonglobe.com/business/2017/02/15/mit-professor-working-with-uber-address-racial-bias/MMCyaxlBTofPPVr7gojXOI/story.html.

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." *Virginia Journal of Social Policy and Law* 23: 249–304.

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*. 6: 106–17. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.60.1789.

White House. 2014. Big Data: Seizing Opportunities, Preserving Values. https://obamawhitehouse.archives.gov/sites/default/files/docs/20150204_Big_Data_Seizing_Opportunities_Preserving_Values_Memo.pdf.

White House. 2016. Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf.

Zatz, N. 2009. "Managing the Macaw: Third-Party Harassers, Accommodation, and the Disaggregation of Discriminatory Intent." *Columbia Law Review* 109: 1357–439.