

Article The Rating Game: The Discipline of Uber's User-Generated Ratings

Ngai Keung Chan

Cornell University, USA nc478@cornell.edu

Abstract

On-demand labor platforms are built around user-generated rating systems that entail consumers' monitoring of workers as a tool for algorithmic labor management. The user-generated ratings can directly determine workers' employability. Workers are thus under pressure to score well, though they may lack knowledge about the performance metrics. Using in-depth interviews with Uber drivers, this article examines how Uber's rating system as an engine of anxiety shapes drivers' work practices through processes of surveillance and normalization. Specifically, drivers are prompted to play the rating game, a game that pressures them toward the normalization of platform surveillance and self-discipline for maximizing their ratings. The internalization of ratings is channeled through drivers' reactive practices in the rating game—that is, drivers' modification of their behavior in response to being evaluated and monitored. This article considers the socio-technical conditions under which drivers are subject to the platform surveillance and disciplinary power.

Introduction

In January 2017, *Saturday Night Live* aired a sketch entitled "Five Stars" in which Dave and his Uber driver Petros endeavored to dazzle one another to ensure a five-star rating to improve their 3.9/5.0 ratings. Dave and Petros were not only anxious about their ratings but they also judged each other by their respective ratings. To improve their ratings, they attempted to engage in a "friendly conversation," or what Hochschild (1983) terms "emotional labor," based on their stereotypical assumptions about one another's race. Finally, they realized how the rating system had pressured them to engage in the contrived conversation they had entered into. In contrast to the skit, Uber might have already deactivated a driver's account when the average rating—an average of the 500 most recently rated rides—drops below a minimum threshold (e.g., 4.6/5.0 stars)¹ (Rosenblat 2018).

Yet, the interactions between Dave and Petros in the sketch, I argue, speak to a wider concern about the surveillant relationship between the platform, the workers, and the consumers. Today, on-demand labor platforms are built around user-generated rating systems that entail consumers' monitoring of workers as a tool for algorithmic labor management (Gandini 2018; Stark and Levy 2018). Such systems enable Uber and other on-demand labor platforms to transform individual workers' performance into quantifiable metrics

¹ Uber (2018) claims that the minimum average rating is location-specific.

that determine these workers' employability (Gandini 2018). While technologically mediated surveillance has long existed in the workplace (Ajunwa, Crawford, and Schultz 2017; Ball 2010), rating systems facilitate the surveillance of drivers' performance by rendering consumers "middle managers" whose evaluations can determine drivers' employability (Gandini 2018; Rosenblat 2018). As the sketch "Five Stars" shows, Uber's rating system not only prompts drivers and passengers to rate one another's performance but it also incites drivers' reactive practices, or what Anderson (2016) calls "monitored performance" for maintaining a high rating. Ratings become "engines of anxiety" as drivers' practices are motivated by the consciousness of being monitored and "the fear of falling in rank that dominates the consciousness of those subject to them" (Espeland and Sauder 2016: 4).

Based on in-depth interviews with Uber drivers.² I introduce the notion of the rating game to describe how Uber's rating system shapes drivers' expectations of service encounters and work practices through processes of surveillance and normalization. Playing the rating game pressures drivers toward accepting the normalization of platform surveillance and exercising self-discipline in order to maximize their ratings. I use the term "rating game" for two reasons. First, it highlights the gamified nature of user-generated ratings that stimulate drivers' engagement with platform surveillance (Cohen 2016; Scheiber 2017). Second, it acknowledges Burawoy's (1979) game metaphor in the labor process. By pinpointing the micro-interactions that produce worker consent to exploitation, Burawoy (1979) found that factory workers played a game of "making out" wherein workers had to achieve an acceptable level of production and competed against others. The rules of the game disciplined workers to normalize continuous monitoring through rewards and punishments. Similarly, Uber drivers are continuously under pressure to score well while playing the rating game. What distinguishes Burawoy's game from the rating game is that the latter's rules are imposed and policed by the platform through algorithms and information asymmetries (Anderson 2016; Rosenblat 2018). Put simply, the rules of the rating game vary across different labor platforms, based on platform-specific policies and affordances. For example, care work platforms may make the reputational metrics of workers visible to potential clients (Ticona and Mateescu 2018), whereas Uber passengers cannot see a specific driver's rating before a ride. Thus, we may observe various kinds of platform surveillance and workers' reactive practices, though there is a proliferation of rating systems across on-demand labor platforms. The analogy of the rating game attends to how workers interpret the socio-technical systems where they work and encounter ratings.

Analyzing drivers' work as a game is not to romanticize algorithmic feedback-driven labor management but to highlight the disciplinary power of ratings through the lens of drivers' reactive practices. The analysis focuses on the goals workers are compelled to pursue and the factors that incite their reactions. As the opaque platform surveillance becomes integral in drivers' daily work, drivers must navigate the algorithmic platform on their own and modify their behavior in reaction to being evaluated. Hence, an exploration of drivers' responses to ratings can help us understand the discipline and normalization of Uber's rating system.

Uber's Five-Star Rating System and Its Surveillance Functions

Before discussing Uber drivers' practices in the rating game, it seems important to address how and why Uber's rating system operates. This helps unpack the socio-technical conditions under which Uber legitimizes its surveillance over drivers. In 2018 there were three million Uber drivers worldwide (Bhuiyan 2018). Uber manages the workforce through algorithms (Lee et al. 2015; Rosenblat 2018). The production and distribution of ratings are one of the disciplinary techniques that are enacted through surveillance (Chan and Humphreys 2018; Rosenblat 2018; see also Sauder and Espeland 2009). Through the rating system, drivers and passengers can rate one another on a five-point scale after each ride. Drivers and passengers play the dual roles of the watchers and the watched in the rating game, but Uber mainly uses ratings to

² Data for this paper come from a larger qualitative study that examined Uber drivers' work practices. From December 2017 to May 2018, I interviewed 16 Uber drivers who operated in eight cities across the United States. I recruited participants during my rides and on social media. Men were overrepresented in the sample (n = 13). The interviews lasted between 30 minutes and two hours. I assigned pseudonyms to the interviewees to protect their identities.

manage drivers. When a driver's average rating from their last 500 rated rides falls below the minimum threshold Uber may issue a warning or deactivate the driver's account.

Echoing Sauder and Espeland's (2009) Foucauldian analysis of law school rankings, Uber's rating system enables remote platform surveillance by transforming categorical differences among drivers into numerical, user-generated ratings that can be easily circulated. Each driver is an individual case that can be described, judged, and compared through a number (Foucault 1995). Ratings evince the idea that drivers' performance during each ride is comparable in accordance with the rules of the rating game. In its "community guidelines," Uber (2018) equates a high rating with "good" and "professional" services. Ratings impose a normative standard for service encounters; for example, Uber suggests drivers should convey "friendliness" and offer amenities to passengers (Chan and Humphreys 2018). Drivers' ratings thus represent how well they play the rating game wherein they need to conform to the standard. Those who do not conform to the standard implicit in ratings risk the platforms' punishment.

Additionally, Uber converts the surveillance functions of ratings into a seemingly legitimate form of accountability. In the 2015 Federal Trade Commission workshop on the sharing economy, Uber made explicit its design rationale for the rating system:

This rating system does three critical things: it (1) incentivizes high quality service, (2) establishes accountability, and (3) promotes courteous conduct, and helps to mitigate the discrimination that is all too common in traditional for-hire transportation. (Uber 2015: 5)

Uber frames the continuous monitoring as a boon for both drivers and passengers, as ratings seemingly could hold both parties accountable and alleviate potential biases. The rating system thus promotes "a safe and respectful work environment for drivers" (Uber 2015: 5). Uber's (2018) community guidelines describe the ratings as a "two-way system" that "holds everyone accountable for their behavior." Uber's narrative reaffirms its classification of drivers as independent contractors (Rosenblat 2018) and it distributes the duties of surveillance to the end users: the platform seemingly is not responsible for generating ratings but invites drivers and passengers to engage in participatory surveillance over one another. Although Uber uses ratings as a disciplinary mechanism, the ostensibly participatory character of ratings appears to neutralize its governing role. Surveillance, which entails the collection, analysis, and distribution of performance metrics, is thus considered a necessary precondition for workers' benefits and technological innovation (Cohen 2016).

Uber has recently experimented with a rewards program called "Uber Pro," in eight US cities, aiming to improve the retention of drivers on the platform (Parekh and Wiezbowski 2018). As Uber describes this program,

Uber Pro is designed to recognize drivers' commitment and high-quality service. To unlock rewards, *drivers need to maintain at least a 4.85 star rating and a low cancellation rate*. Drivers earn points on every trip during fixed 3-month periods, and the points they earn in one period go toward unlocking status (and more rewards) in the next. (Parekh and Wiezbowski 2018; italics added)

This program, which began in November 2018, classifies drivers into four categories: partners, gold, platinum, and diamond. As drivers earn points to unlock a higher status, they can obtain more monetary rewards. At the time of writing, the effects of the program remain unclear. Yet, drivers seem to praise this rewards program as it accounts for drivers' feedback on the rating system and provides monetary rewards to recognize drivers' efforts (Cradeur 2018). Meanwhile, following Cohen's (2016) discussion of gamified surveillance, the rewards program may mobilize drivers to further participate in meticulous platform surveillance. Expectedly, after the launch of the rewards program both rewards and punishment (i.e., the risk of being deactivated) would work in tandem to motivate drivers to play the rating game.

The Discipline and Normalization of Uber's Rating System

Uber's rating system as a form of surveillance technology is not automatically deployed effectively; instead, it is experienced by drivers in their mundane everyday activities. As I discuss below, the interviewees were conscious of the platform surveillance and developed tactics to maintain a good rating; namely, through the allocation of resources and the performance of emotional labor. Hence, ratings "create a structure in which individuals themselves, without coercion, end up exercising self-discipline and self-restraint, in order to be in conformity with the 'norm' and the perceived expectations of the 'watchers'" (Manokha 2018: 226).

Consistent with existing studies on Uber drivers (Lee et al. 2015; Raval and Dourish 2016; Rosenblat 2018), most of the interviewees were attentive to their ratings. Yara, for instance, stated, "I watch my ratings very closely," even though "it usually doesn't change." Noah, too, paid close attention to his rating and expressed concerns over the punishment for not maintaining a good rating:

The average person, they think that three stars is average, four stars is good, five stars is excellent. So if you just give a basic ride to somebody, you don't do anything over and beyond, you pick them up, you take them somewhere, you drop them off, the ride is over, you're given three stars, it was an average ride. The problem is if you receive 100 four-star rides, you would be kicked off the platform because it's not 4.6. You would have to consistently deliver five-stars every single time.

Noah's comment reveals that surveillance is continuous as he must consistently get five-star ratings to "maintain a rating of 4.6 or higher," and thus employment opportunities. Several of the interviewees shared Noah's criticism and noted that Uber should inform passengers to rate properly. As Owen put it, passengers "don't realize that if you [drivers] get enough four-stars, that can eventually result in deactivation." Though Uber discursively delegates the duties of worker monitoring to passengers, the interviewees' normative judgment of ratings reorients the responsibility for oversight to the platform. From the interviewees' perspective, even though the passengers are tasked with rating themselves, they are not instructed on the evaluative criteria for doing so. Passengers' evaluation may be largely dependent upon their pre-existing understandings of what an appropriate rating is. Hence, the interviewees contemplated that the platform should take the responsibility of improving the design and actual implementation of the rating system.

As discussed earlier, Uber's rating system enables drivers to rate their passengers. As such, drivers are not simply the objects of surveillance but also the surveillors. Once drivers complete a ride, Uber's mobile application will immediately pop up with "Rate your rider" to remind drivers to engage in the act of evaluation. Drivers must rate their passengers before proceeding to accept a new ride. However, most of the interviewees reported that it is rare for them to rate their passengers lower than five stars unless their passengers were "rude," "obnoxious," or violated Uber's policy. Hamilton, for example, only gave a passenger a one-star rating once, as this passenger used "abusive language" during the service encounter. The inattention to the evaluation of passengers is in part because "customers are not penalized at all in their ratings," as Noah put it. He explained, "A customer can have a rating as low as possible, and still get a ride. I mean, I've picked up somebody who had a 3.2 rating... Any driver with a 3.2 rating would not be driving. You'll never see a driver with a 3.2 rating." His reflection is indicative of Uber's unequal treatment of drivers and passengers, though it is noteworthy that some drivers reportedly use passengers' ratings as proxies for social acceptability (Rose 2018). Giving someone a one-star rating, Nathaniel added, also means the loss of employment opportunities because "you won't match with them again."

_

³ Although drivers may develop their own evaluations of prospective passengers (e.g., whether they should accept passengers with a low rating), Uber does not formally punish passengers with a low rating. Yet, since September 2018, Uber has attempted to address this issue in Australia and New Zealand. Uber may temporarily deactivate the accounts of passengers whose average rating is lower than 4.0/5.0 in these two countries (BBC News 2018), but this policy has not been implemented in the United States, where I conducted interviews with drivers.

Among those in the sample, only Damon and Eaton admitted they would take the passenger's evaluation seriously. Consistent with Uber's discourse, they considered the rating system a way to hold passengers accountable. Eaton explained, "I personally use a rider's rating as a guideline. So if a rider has a really low rating, it kind of is [that] my antennas are up so I will pay attention to behaviors that may be of concern that another driver has marked." Inherent in Eaton's comment is an internalization of the standard embedded in ratings: a high rating is equated with a "good" passenger. It was common, however, for my interviewees to dismiss the accountability of passengers' ratings. Although Uber's rating system, by design, seems to enable drivers and passengers to monitor one another, drivers are more vulnerable to being monitored and deactivated in the rating game due to the platform policies.

While threats of being deactivated reveal the coercive dimension of surveillance, ratings are also engines of anxiety (Espeland and Sauder 2016), creating intense pressure toward compliance with the goals of the rating game. Noel, for example, offered, "It will ruin my whole day if I get a bad rating... This is my personal business, it's a reflection on my service, and when I get a bad rating, I take it personal[ly] and I try to figure out where I possibly have, was that customer not pleased." Eaton, meanwhile, reasoned, "Behaviors hurt your rating... If they just have bad behavior, they're always going to end up with a bad rating." Comments like this reaffirm the belief that ratings have become normalized as a standard for evaluating drivers' performance. Internalization is central to the disciplinary power of ratings; drivers will exercise self-discipline to maximize their ratings only when ratings have become naturalized as a standard for evaluating their performance.

The internalization of ratings motivates drivers to attend to the details of what counts and what factors seem to be most amenable to their control. Accordingly, the interviewees endeavored to allocate resources and perform emotional labor to satisfy the perceived expectations of their passengers. The allocation of resources concerns the additional amenities drivers offer, whereas the performance of emotional labor requires drivers to maintain a "professional" appearance and manage their interactions with passengers. These two reactive practices may be Uber's intended outcomes of implementing the rating system. Drivers learn to engage in such monitored performance (Anderson 2016) as they realize that passengers are responsible for rating their behavior. The allocation of resources entails drivers' considerations of the tangible things that are offered to their passengers. When asked about his reactions to ratings, Nathaniel noted, "What it does is, it creates an arms race. No longer are you just driving a person from A to B. Now, you have to have a charger. Now, you have to have internet. Now you have to have bottled water. You have to give out candy." Offering amenities, he added, distinguishes "elite service" from "just a transportation service." Similarly, Bale and Yara offered gum and candy to their passengers. It is noteworthy that offering amenities like bottled water is in line with Uber's suggestions for getting a good rating. Some interviewees, such as Nathaniel and Noah, did recognize the additional expenses when they offered amenities to their passengers. Noah explained, "What the customers don't realize with the charging the phone is that charge creates drain on the battery in my car. I have a \$300 battery inside my vehicle." Yet, they continued to offer amenities, as they considered that such behavior could help them maintain a high rating.

Another practice for maximizing their ratings is to perform emotional labor, which is analogous to the findings of extant studies (Raval and Dourish 2016; Rosenblat 2018). The interviewees recurrently used specific terms—being "professional" and offering "good services"—to describe how they interacted with passengers verbally and nonverbally to maintain a high rating. Owen, for example, reasoned, "What it takes to make a customer happy is good service." Good service, he continued, means "speaking to them professionally and providing them with everything that they need to get to their destination, and getting them there safely." Other interviewees attempted to spread "positivity" to their passengers. Yara claimed that she never made any complaints about Uber during service encounters. She reasoned,

Because they [passengers] do not want to get into an Uber car and hear me complain about my job. If you got in the car and they would hear, 'Ah, I hate my job, blah, blah, blah,' if they heard me complain, they would rate me lower ... I want to them to give me the best rating possible.

Her comment is indicative of the self-management of emotional displays during service encounters. Adele, meanwhile, noted that drivers should learn to "profile" their passengers to know their psychological state to better communicate with them.



Figure 1. A sign about ratings in an interviewee's vehicle.

While Owen, Yara, and Adele focused on the ways of interacting with their passengers, a small subset of the interviewees attempted to "game" the rating system. Gaming does not necessarily challenge the calculative logic of the rating system; instead, it may contribute to the internalization of ratings (Chan and Humphreys 2018; Sauder and Espeland 2009). The sign about ratings shown in Figure 1 is telling of drivers' gaming practice. Hart, who put this sign in his vehicle, explained that he expected every passenger to give him a five-star rating because he was aware that ratings would determine his employment opportunities. The sign, for him, was a way to convey friendliness and remind passengers to give him a five-star rating.

Overall, I have described how the interviewees made sense of the rating system and attempted to develop reactive practices to maximize their ratings. Uber drivers, however, have limited ability to manage their ratings as they lack full information about the rating system and cannot control how passengers evaluate their performance. Ratings produce uncertainty and anxiety, pressuring drivers to self-monitor their ratings and conform to the perceived expectations of service encounters. Not all interviewees responded to ratings in the same way. Indeed, some interviewees were critical of the unfairness of ratings but they still had to play the rating game unless they quit the platform.

Conclusion

In pointing out the discipline of ratings, my goal is not to make a technological determinist argument that drivers are necessarily subject to platform surveillance. Instead, I wish to unpack the socio-technical conditions through which drivers are constructed to be objects of (self-)surveillance. In Uber's discourse, the rating system is considered a boon to both drivers and passengers, with its assurances of accountability. In theory, drivers can monitor and evaluate passengers' performance but at the same time they remain subject to disciplinary power because of the platform policies. Drivers' reactive practices entail affective responses to the surveillance functions of the rating system, which in turn sustains and normalizes the disciplinary power of ratings.

The point is not to contend that the rating system is necessarily punitive. Instead, the analogy of the rating game acknowledges the interplay between Uber, its technical platform, and its drivers. Uber's discourse and platform work in tandem to construct the rules of the rating game, which inform drivers' expectations of service encounters. While the discipline of ratings incites drivers to play by the rules, drivers may decide what factors are amenable to their control so as to maximize their ratings. This observation resonates with

previous scholarship that has examined organizations' reactive responses to rankings (Espeland and Sauder 2016). Crucially, *reactive* ratings define what counts as relevant to drivers and wipe off alternatives by simplifying and standardizing information. As such, ratings can alter the socio-technical conditions under which drivers make work-related decisions.

It is noteworthy that drivers' reactive practices that are geared toward ratings evolve as Uber changes its policies and the technological design of the platform. Echoing Burawoy's (1979) observation of work games, changes in the rules may result in a "legitimation crisis," potentially threatening workers' willingness to participate in work games. Within the context of Uber, we should consider how drivers might modify their reactions to changes in the platform so as to better understand the ramifications of platform surveillance over time. As user-generated ratings and other kinds of performance metrics become taken for granted as part of digital platforms, we must be critical of surveillance functions and the ramifications of ratings. Understanding people's engagement with ratings on Uber and other on-demand labor platforms can help us explore socio-technical conditions that enable the surveillant relationship between platforms, workers, and consumers.

References

Anderson, Donald Nathan. 2016. Wheels in the Head: Ridesharing as Monitored Performance. Surveillance & Society 14 (2): 240-258.

Ajunwa, Ifeoma, Kate Crawford, and Jason Schultz. 2017. Limitless Worker Surveillance. *California Law Review* 105: 735-776. Ball, Kirstie. 2010. Workplace Surveillance: An Overview. *Labor History* 51 (1): 87-106.

BBC News. 2018. Uber to Block Low-Rating Riders in Australia and New Zealand. https://www.bbc.com/news/world-australia-45416404 [accessed December 29, 2018].

Bhuiyan, Johana. 2018. Uber Powered Four Billion Rides In 2017. It Wants to Do More—And Cheaper—In 2018. *Recode*, January 5. https://www.recode.net/2018/1/5/16854714/uber-four-billion-rides-coo-barney-harford-2018-cut-costs-customer-service [accessed December 29, 2018].

Burawoy, Michael. 1979. Manufacturing Consent: Changes in the Labor Process under Monopoly Capitalism. Chicago: University of Chicago Press.

Chan, Ngai Keung, and Lee Humphreys. 2018. Mediatization of Social Space and the Case of Uber Drivers. *Media and Communication* 6 (2): 29-38.

Cradeur, Jay. 2018. Uber Pro–New Rewards Program Helps Drivers Be More Successful. *The RideShare Guy*, November 1. https://therideshareguy.com/uber-pro/ [accessed November 26, 2018].

Cohen, Julie E. 2016. The Surveillance-Innovation Complex: The Irony of the Participatory Turn. In *The Participatory Condition* in the Digital Age, edited by Darin Barney, Gabriella Coleman, Christine Ross, Jonathan Sterne, and Tamar Tembeck, 207-226. Minneapolis, MN: University of Minnesota Press.

Espeland, Wendy Nelson, and Michael Sauder. 2016. Engines of Anxiety: Academic Rankings, Reputation, and Accountability. New York: Russell Sage Foundation.

Foucault, Michel. 1995. Discipline and Punish: The Birth of the Prison. New York: Vintage Books.

Gandini, Alessandro. 2018. Labour Process Theory and the Gig Economy. *Human Relations*. https://doi.org/10.1177/0018726718790002 [accessed December 29, 2018].

Hochschild, Arlie Russell. 1983. The Managed Heart: Commercialization of Human Feeling. Berkeley, CA: University of California Press.

Lee, Min Kyung, Daniel Kusbit, Evan Metsky, and Laura Dabbish. 2015. Working with Machines: The Impact of Algorithmic and Data-Driven Management on Human Workers. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 1603-1612. New York: ACM.

Manokha, Ivan. 2018. Surveillance, Panopticism, and Self-Discipline in the Digital Age. *Surveillance & Society* 16 (2): 219-237. Parekh, Sunil, and Ali Wiezbowski. 2018. Uber Pro (beta): Helping Drivers and Their Families Reach Their Goals, On and Off the

Road. *Uber Newsroom*, November 1. https://www.uber.com/newsroom/uberpro/ [accessed December 29, 2018].

Raval, Noopur, and Paul Dourish. 2016. Standing Out from The Crowd: Emotional Labor, Body Labor, and Temporal Labor in Ridesharing. In CSCW '16 Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing, 97-107. New York: ACM.

Rose, Rebecca. 2018. The shame of my very low Uber rating. *Financial Times*, July 4. https://www.ft.com/content/05df5746-7e56-11e8-8e67-1e1a0846c475 [accessed December 29, 2018].

Rosenblat, Alex. 2018. *Uberland: How Algorithms are Rewriting the Rules of Work*. California, CA: University of California Press. Saturday Night Live. 2017. *Five Stars*. https://view.yahoo.com/show/saturday-night-live/clip/60829414/five-stars [accessed December 29, 2018].

Sauder, Michael, and Wendy Nelson Espeland. 2009. The Discipline of Rankings: Tight Coupling and Organizational Change. American Sociological Review 74 (1): 63-82.

- Scheiber, Noam. 2017. How Uber Uses Psychological Tricks to Push its Drivers' Buttons. *The New York Times*, April 2. https://www.nytimes.com/interactive/2017/04/02/technology/uber-drivers-psychological-tricks.html [accessed December 29, 2018].
- Stark, Luke, and Karen Levy. 2018. The Surveillant Consumer. Media, Culture & Society 40 (8): 1202-1220.
- Ticona, Julia, and Alexandra Mateescu. 2018. Trusted Strangers: Carework Platforms' Cultural Entrepreneurship in the On-demand Economy. *New Media & Society* 20 (11): 4384-4404.
- Uber. 2015. In Response to the "Sharing" Economy: Issues Facing Platforms, Participants, and Regulators. https://www.ftc.gov/system/files/documents/public comments/2015/08/02049-96749.pdf [accessed December 29, 2018].
- Uber. 2018. *Uber Community Guidelines*. https://www.uber.com/legal/community-guidelines/us-en/ [accessed December 29, 2018].