# In Your Face: Digital Rider Ratings and the Facework Process in the Case of Uber

by

Grant Matthew Yeatts

A thesis submitted in partial fulfillment of the requirements
for graduation with Honors in Sociology.

Whitman College
2020

*Certificate of Approval*

This is to certify that the accompanying thesis by Grant Matthew Yeatts has been accepted in partial fulfillment of the requirements for graduation with Honors in Sociology.

______________________________

Matthew Gougherty, Thesis Chair

Whitman College
May 20, 2020

# Table of Contents

## Acknowledgements

First and foremost, I would like to acknowledge the land upon which I have written this thesis and studied for the past four years is the traditional homelands of the Cayuse, Umatilla, and Walla Walla peoples. It is important to acknowledge the history of this place and in remembering these communities we honor their lives, their legacies, and their descendants.

To the individuals I interviewed, I extend my deepest gratitude. Thank you for reaching out and your willingness to participate in this project. Thank you for sharing your stories and time with me. Without your generosity this project would not have been possible.

To my thesis committee members, Matthew Gougherty and Gilbert Mireles, thank you for your unwavering support throughout this process. Thank you for taking the time to read, edit, and make suggestions on the many iterations this project has gone through. This thesis would not be what it is today without your guidance, mentorship, and willingness to meet both in-person and virtually.

To Michelle Janning, thank you for helping me discover my love for sociology in Principles during my first year at Whitman. To Keith Farrington, thank you for your guidance as my major advisor. To the entirety of the sociology faculty at Whitman College including those I have yet to mention, Alissa Cordner, Helen Kim, and Alvaro Santana-Acuña, thank you for four wonderful years of inspiring study in the discipline. Whether I have interacted with you in class or across campus, you all have challenged me to think critically, consciously, and creatively. Your expertise and commitment to helping students does not go unnoticed. I will never forget you, nor my sociological imagination.

To my housemates and best friends, thank you for helping to keep me sane. This was not the senior year we had all hoped for, but the memories we have created and the laughter we have shared even amidst the dire COVID-19 pandemic has kept me going.

Finally to my family, Mom and Dad, I could not have done this without you. Lauren and Mark, I am so proud of you both. I love you all unconditionally and truly cannot thank you enough for loving me, supporting me, teaching me, inspiring me, and putting up with me for 22 years.

# Abstract

Uber is a contemporary technology company that matches consumers directly with service providers in the on-demand ride marketplace. In their management of drivers, Uber relies on consumer ratings of drivers to make employment decisions about which drivers are allowed to continue working and which drivers are deactivated, or removed from the app as a service provider. This project set out to investigate how and why Uber consumers rate the drivers who provide them with rides. Data from nine semi-structured interviews with Uber riders in a large, American city on the west coast reveal that Uber's rating prompt is the primary catalyst for the collection of consumer ratings. The criteria constructions riders use to make decisions about rating vary due to subjectivity, differential holdings of criteria, and criteria contradictions, despite agreement on the importance of safety. Raters were also found to fall into four unique categories given the nature of their experiences and rating deployments: full spectrum raters, partial spectrum raters, empathetic raters, and binary raters. The project concludes with a discussion of the implications of these findings for Uber and generalized contexts, and presents alternative visions of driver evaluation as well as avenues for continued research on quantification, evaluation, and digital life.

# List of Tables

## Introduction

Uber is a popular ride hailing service operating in over 700 countries across the globe that helps connect consumers (riders) directly with service providers (drivers) by way of their proprietary digital app interface for rides from one place to another. When requesting a ride, consumers select their destination and are matched with an available driver nearby, who in turn comes to the rider's current location and picks them up. After pickup, the app interface functions as a GPS navigation tool, showing drivers and passengers alike a map visual with the outlined route to the requested location and the car's current position until the ride is completed. After the ride, both the consumer and driver are prompted to evaluate each other on a scale of one to five stars. Although riders are rated by drivers at the end of each trip, they do not appear to experience negative impacts for receiving a low rating. On the other hand, drivers receiving poor ratings from their riders can be deactivated, or barred from using the Uber app to continue working as a driver (Chan 2019). Given this and the fact that Uber does not give riders prescribed criteria along which to rate drivers nor compensation for completing the rating, the questions follow: how and why do Uber riders translate their experience of a ride into a quantitative rating metric?

This question is important sociologically because Uber and other app-facilitated businesses are relatively new organizations of socioeconomic activity with distinctly disruptive business practices which warrant critical interrogation from scholars, policy-makers, and the public alike. Further, Uber relies on rider ratings to make concrete hiring decisions about drivers. While much of the literature has focused on driver responses to rider ratings and the general shape of the labor relation Uber constructs, no studies to my knowledge have approached this issue by collecting data from a sample of riders. Doing so will produce valuable data and insights as to how Uber riders may (or may not) engage with the rating system that requires their translation of a necessarily interactive experience with a driver into a rating that has salient ramifications for the driver's employment status.

In this work I first propose a synthesis of sociological literature relevant to this project. Next, I offer theoretical concepts with which to frame the data. I then turn to review the methodology employed, acknowledging the limitations to the study. In moving toward analysis of the data collected, I outline disciplinary forces that coerce consumers to participate in the rating act and develop an Uber rater typology to describe the ways in which different riders rate drivers. Finally, I conclude with a discussion of the implications of these findings, offer alternative visions of Uber driver evaluation, and present further avenues for future research relating to quantification and neoliberalism.

## Review of Literature

In order to best position this project in relation to other academic works, I conducted a review of predominantly sociological literature germane to the project's themes of employee evaluation and management, consumer rating and quantification, and decision making and bias drawn from English language sources. While this project employs theory and methodology common to the sociological discipline, I acknowledge that this project, along with the sociological discipline widely conceived, intersects with cross-disciplinary fields including legal studies, organizational management studies, and digital media studies that should not be discounted. Thus, in order to most holistically frame this project I include sources from these areas of scholarship as well, highlighting their relevance to the project as a whole.

### Uber Drivers and Algorithmic Management

Uber drivers constitute the labor supply that meets consumer's demand for rides. This service provider workforce consists of independent contractors that are spread widely geographically and whose numbers are constantly in flux. Yet, the company still maintains the management of this workforce in order to provide the best experience for its consumers (Rosenblat and Stark 2016). In order to do so, Uber employs what I will call the algorithmic rating apparatus that sources quantitative data from riders in order to make determinations about contractor wages and firing (Rosenblat, Levy, Barocas, and Hwang 2017). This strategy of assessing the quality of a product or service from aggregated consumer reviews emerged as ubiquitous with the development of digital technology (Stark and Levy 2018). The apparatus itself is considered algorithmic because it employs a machine in the form of a data processing computer in order to adjust metrics based on coded formulas as new data comes in and to define and notify which drivers are not meeting standards.

Content analyses of Uber's documented operations and digital interfaces have found that after each trip, the consumer is prompted on the app to rate the driver on a scale of one to five stars with the additional prompt to leave comments for that trip. The ratings from a driver's most recent 500 trips are then averaged to determine their overall rating. While prompted to, the rider is not required to rate the driver. In the case the rider opts out of rating, there is no impact on the driver's average score. The averaging of these ratings places all Uber drivers, despite their diversity, on a common scale of evaluation and as such they can be compared to one another by way of this metric. In order for a driver to continue working, Uber requires they have an average rating above a city-specific benchmark that usually hovers around 4.6 out of 5 stars. If a driver falls below this mark, they risk deactivation which entails temporary or permanent barring from the platform as a worker (Rosenblat and Stark 2016; Rosenblat et al. 2017). These grounds for deactivation coupled with their definition as independent contractors which does not require the company to guarantee minimum wages, renders this vision of work more precarious than traditional full-time employment. However, drivers are found to be motivated by the economic prospect Uber presents and therefore attempt to cope with this precarity by focusing on the advantages of the driver role while downplaying the risks they face (Peticca-Harris, deGama, and Ravishankar 2020).

## Quantification

Quantification is the process by which something or someone comes to be represented by numbers. In turn, the sociology of quantification, a relatively new field yet one that is intrinsically tied to the subject of this project, explores the ways in which our social worlds are increasingly represented by and determined by quantitative data. In this field, quantification is generally understood as a "social action that, akin to speech, can have multiple purposes and meanings. Only by analyzing particular instances of quantification in context can these purposes and meanings be revealed" (Espeland and Stevens 2008:405). This project will approach quantification as it plays out in the context of Uber consumers rating their ride experiences and how those ratings are then mapped onto drivers to represent their reputation. More generally, the successes of quantification are premised on the idea that people see numbers as objective due to their common use in mathematics and sciences, both disciplines in pursuit of objective truths. Therefore, individuals tend to trust quantitative data are accurate representations of what these data are measuring (Porter 1996).

One central concept in the sociology of quantification is commensuration, or "the valuation or measuring of different objects with a common metric" (Espeland and Stevens 2008:408). This process brings about the homogenization of all objects being quantified, no matter how disparate in their materialization, under a singular system whereby each object has a position relative to all others and differences or distances between objects can be quantified too. In the case of Uber, quantification is embedded in the algorithmic rating apparatus as the rider's rating of a driver serves as a numerical representation of that rider's experience with that driver. As such, all individual ride experiences become commensurate with one other as each is quantified along the star rating scale by the rider. Furthermore, drivers themselves become commensurate relative to each other as the company calculates a numerical average and dynamically assigns it to each driver. While drivers' average scores have no inherent meaning devoid of context, when they are framed against the maximum 5.0 average we are able to quantify each driver's apparent shortcomings using a continuous scale. We are also able to use these average scores as a common metric to ground the comparison of drivers to one another. In these instances, the differences in scores have meaningful magnitude due to the commensuration of quantification. Altogether, the commensuration of drivers by way of their average rating scores upholds the logic of deactivating drivers with lower scores than others.

Adjacent literature analyzes the quantification process in relation to the number of quantifiers, or the people or entities taking on the task of quantifying. In their empirical work on law school programs, Sauder and Espeland (2006) note that US News is the most popular ranking publisher and does not have notable competition. They go on to explore the rejection of a single quantifier observing that if there were multiple agencies ranking schools, this would "create a degree of ambiguity about the relative standing of schools and that this ambiguity allows schools to regain reputational control" (Sauder and Espeland 2006:206-207). Yet, in the same vein they note that "the proliferation of rankings would also tend to reinforce rankings as the legitimate mode of accountability in higher education" (Sauder and Espeland 2006:207). This is important to consider in the context of Uber, which sources rating data from a large population of consumers. Each rider becomes a quantifier when they decide to rate, so the inclusion of multiple quantifiers may reinforce the legitimacy of the ratings as was found in the literature. However, Uber's management of drivers still depends on a measurement of driver reputation that lacks ambiguity. The process of reducing a multitude of quantifiers' data to a single metric for each driver through averaging bridges this paradox for the company's benefit as the driver's average rating score is the only number published without opposition from any other source. As such, Uber prevents the introduction of any ambiguity that would empower drivers to challenge these reputational assessments made by Uber and thereby also the algorithmic rating apparatus and the outcomes it creates.

## The Work of Quantification

The sociology of quantification also traces the amount of work it takes to produce and process large amounts of quantitative data. For example, "when grading essays...teachers might use a 100-point scale to evaluate each one, relying on prior experience and comparison to quantify individual accomplishment" (Espeland and Stevens 2008:410). The same can be said about an Uber rider when they rate drivers, as they use a one to five star scale to evaluate and may quantify their experience based on what they have come to expect from Uber drivers as informed by their past experiences in this social context. Additionally, there is considerable labor taken on by Uber in aggregating all of their consumer data to publish average driver ratings. Wanting to minimize operating costs and maximize profits under basic capitalistic ideology, Uber automates this aggregation using their computerized algorithmic rating apparatus. This automation accomplishes its goals much faster and cheaper when quantitative data is provided as the input as computers can more easily manipulate, operate on, aggregate, and calculate numbers compared to qualitative linguistic sentiments that require much more complex processing techniques. Acknowledging the immense labor of generating and managing these data, we see that "quantification is often the work of large bureaucracies" as "we often forget how much infrastructure lies behind the numbers that are the end product of counting regimes. This is especially true when numbers circulate to places that are removed from the bureaucracies that manufactured them" (Espeland and Stevens 2008: 411). The digital world is paradoxically extremely local as the physical device that allows access to this world exists in the hand of the user (or wherever they put their device), while the components that make up the digital world such as softwares, apps, and interfaces can be located anywhere across the globe or in the immaterial digital cloud, obscuring the power and function of interested bureaucracies such as Uber.

## Responses to Quantification

When people are behind the object or experience being quantified, or a person is what is being quantified in the first place, quantification takes on an influential role in how people decide to live their lives and perform their organizational functions. One response individuals may have to the quantification of something they have a vested interest in is "reverse engineering" whereby they direct their energy and actions toward optimizing only those metrics the quantifier is measuring in order to receive a high score (Espeland 2016:280). Additionally, people become emotionally attached to the numbers that represent them, and thus express a multitude of emotions at different points including anxiety about their scores falling lower, happiness when their scores increase, and despair when their scores decrease. All of these responses have been observed in law school administrators as a response to certain aspects of the school's program being quantifiably measured and then ranked against other programs (Espeland 2016).

Responses to the quantification process have also been explored in the context of Uber through content analysis of Uber's discourse and an online forum of Uber drivers.

The company's collection of quantitative data and decision-making based on these data has led to the "mediatization," or redefinition based on numbers and digitality, of the social space of Uber drivers' work (Chan and Humphrys 2018:29). These researchers argue that in response to this mediatization, drivers develop an "algorithmic imaginary" whereby they frame, interpret, and understand social life in this mediatized space in relation to quantitative data. As a result, drivers accept and reproduce the value of quantification by internalizing the company's logics, leading them to shape their work practices and movement in ways that maximize benefits as defined by these algorithms similar to the reverse engineering responses observed among law school administrators (Chan and Humphrys 2018; Espeland 2016).

Additionally in the context of Uber, the algorithmic rating apparatus' management of drivers often results in drivers expressing anxiety about their average rating and feelings that additional, uncompensated, material and emotional labors are mandatory in order for them to optimize the ratings they receive (Rosenblat and Stark 2016; Gandini 2019). Materially, this pressure can manifest as drivers feeling the need to offer riders water, candy, or phone chargers. These strategies are suggested by the corporation to drivers who are at risk for deactivation, but all necessitate drivers invest additional money into providing this service for the rider (Chan 2019; Rosenblat et al. 2017). Emotionally, drivers must maintain a professional demeanor and gauge the willingness of the rider to converse so as to not upset them by talking too much or not enough (Rosenblat et al. 2017). Drivers generally may come to internalize these labors and view them as necessary in order to optimize their rating and keep their job. Drivers mainly discover these means of rating score optimization from weekly performance reviews from Uber and online communities of drivers sharing tips to maximize ratings (Rosenblat et al. 2017; Rosenblat and Stark 2016).

Finally, quantification has also been framed as leading to the "economization" of social life, or "the process through which individuals, activities, and organizations are constituted or framed as economic actors and entities" (Mennicken and Espeland 2019:233). This trend is in line with the premises of neoliberalism which will be discussed as part of the theoretical framing of this project. This economization also reinforces the rationality that we can assign numbers to experiences and people as a way to justifiably evaluate them. Just as quantification and Western logic are privileged in the contemporary economy, the application of economic frameworks to private social realms has led to these more private realms inheriting the same privileging of numbers and logic. Furthermore, as explained by Berman and Hirschman (2018) in their review of recent sociological works on quantification, the proliferation of data collecting instruments and the increasingly ubiquitous integration of these instruments into our daily lives has led to the conception of "the quantified self." This constitutes a new kind of subjectivity where rather than centering experience itself, social life and the self are reduced to be understood solely through quantifiable metrics.

## Rating Criteria Constructions

To best understand how Uber riders may or may not rate drivers, it is important to explore other contexts in which people evaluate people and services. In the literature, it has been observed in the specific context of Uber that a rider's experience of positive emotions during their ride is more likely to result in a positive consumer outcome (and rating for the driver) and the inverse is true as well with negative emotions leading to more negative outcomes (Chatterjee 2019). As drivers' employment statuses are determined by these ratings, they have a vested interest in providing the rider with an experience that aligns with the rider's meaning of a positive Uber ride. However, these are loose grounds for forming a consistent vision of the Uber experience that will result in positive emotions and outcomes across a multitude of rider subjectivities, showing the driver's interactive task of gauging how the rider they are currently with is feeling about the current experience.

Additionally, in the context of low-cost versus full-service airlines, "concrete" criteria such as value are relatively more important by consumers in low-cost contexts while "abstract" criteria such as beverage service are relatively more important in the full-service context (Chatterjee 2019:20). This exemplifies the way in which rating outcomes may be variable given the expectations the consumer has for the service and how important they see different criteria being in meeting these expectations. While the price of an Uber ride is greatly variable based on driver supply, rider demand, ride distance, the number of passengers, and many other variables, if we generalize these findings on airline services to ride services, we may hypothesize that consumers who consider Uber's services low-cost compared to other ride services may focus on considering concrete criteria when rating drivers.

Finally, as exemplified by Rivera's work on doormen working at exclusive nightclubs and their evaluation of patrons attempting to gain entrance, she finds that "the specific cues used to evaluate a given customer...were informed by shared status schemas about the relative material, moral, and symbolic worth of certain client groups that were derived from a combination of broader societal stereotypes" (Rivera 2010:247). In other words, doormen had to make meaning out of a potential client's external image, measure that against a stereotype of what could be considered a valuable client, and then make material decisions about whether or not the patron was let into the nightclub. Although in this context Rivera is exploring a worker's evaluation of consumers with one-sided consequences for the consumer, we can see that some evaluative decisions made about people are made on grounds of external presentation and characteristics. As will be further discussed in the next section, these mechanisms of evaluation could be biased on racist, sexist, or host of other discriminatory grounds if the rider holds prejudice against certain identity groups. If present and allowed to operate within the algorithmic rating apparatus, these prejudices would constitute a discriminatory threat against drivers, potentially resulting in the violation of laws protecting equal opportunity employment.

## Rating and Biases

Uber's reliance on rider ratings in making hiring decisions about drivers comes at the cost of requiring riders to translate their interactive experience of being the direct recipient of service into a quantitative measure (Stark and Levy 2018). When this process is undertaken by riders, there is room for multiple forms of bias and discrimination to find their way into concrete decisions, complicating the ethical considerations of using an algorithmic rating apparatus as a system to manage contractors. First off, multiple studies in the literature demonstrate that racial bias has material effects on consumer decisions in online marketplaces, which manifest as negative outcomes for racial minorities such as "lower offer prices and decreased response rates" (Rosenblat et al. 2017:263). Further, Luoh and Tsaur (2007) found that in service industry customer-waitperson interactions, employee gender has an impact on perceived service quality. Both of these show the risk that Uber riders could be incorporating discriminatory beliefs about a driver into their subjective rating criteria constructions and judgements of quality.

Beyond identity-based biases, bias in rating can occur as a result of how consumers perceive and engage with numerical rating systems. Consumers who rate products and services have been categorized in literature as optimistic (one that gives a high rating to every product regardless of perceived quality), pessimistic (one that gives a low rating to every product regardless of perceived quality), realistic (one whose ratings are directly related to perceived product quality), or unreliable (one whose ratings are inversely related to perceived product quality) (Lim and Tucker 2017). Additionally, while consumers are able to distinguish between generally high ratings and low ratings, they tend to collapse numerical categories into a simplified high and low binary that results in binary bias (Fisher, Newman, and Dhar 2018). Although these biases are not necessarily discriminatory, Uber takes into account ratings regardless of the way a rider would go about rating. With drivers needing to meet a quantitative threshold that hovers between four and five stars for maintaining contract employment (both 'high' ratings under binary bias), these minute disparities in the quantification of the quality of the service a rider receives could produce relevant outcomes for a driver's average rating.

In this section I have reviewed how recent academic works have employed interviews with drivers and content analyses of Uber's website and app interface to address questions about Uber and the condition of drivers. I have also shown how researchers have begun to develop a sociology of quantification in other rating and ranking contexts. Finally, I have investigated works on rating as a social act in other contexts, recounting how criteria and perception have been found to influence the way one rates. While Uber ratings have been explored from the driver's perspective and the general notion of consumer ratings have been analyzed in other contexts, this review calls attention to the absence of specific empirical studies of Uber ratings and their effects from the perspective of the rider. This highlights a gap in sociological literature this project helps to fill.

## Theoretical Frameworks

In addition to the literature previously discussed, I develop a theoretical framework which calls upon recent conceptions of neoliberalism and digital bureaucracy, Michel Foucault's notions of power and discipline, and Erving Goffman's symbolic interactionist concepts of faces and facework to construct a multilevel account of the organizational and interactional social phenomena subsumed by this project. First, neoliberalism and digital bureaucracy will be used to generally frame Uber's business model as well as the legitimation of the algorithmic rating apparatus as a management tool. Additionally, Foucault's work on power and discipline will be invoked to triangulate the relations between Uber, Uber consumers, and Uber drivers and the effects these relations produce as they play out in both physical and digital space. Finally, Goffman's definitions of lines, face, and facework will be called upon to frame the interactions between consumers and the drivers they ride with as well as the ways consumers go about rating drivers. Altogether, these theoretical concepts will serve as a composite lens through which the subsequent data will be subject to analysis.

### Neoliberalism

Neoliberalism is a politico-economic ideology that supposes the decentralization of production from one place to a multitude of places, the prioritization of the individual along with the protection of their freedom and property, the advocacy of free markets with minimal intervention by the state, and the minimization of organizational costs for corporations (Zwick 2018). Additionally, in the neoliberal marketplace "the sanctity of contracts and the individual right to freedom of action, expression, and choice must be protected...at all costs" (Harvey 2005:64). In practice, this ideal has materialized in the formulation of a technology-enabled gig economy where independent contractors are free to accept short-term contracts (or gigs) directly in the marketplace as facilitated by apps. Uber is one contemporary example of a corporation that utilizes these neoliberal business practices, using a proprietary app platform to facilitate the logistics of matching drivers with riders directly in the on-demand ride service marketplace.

Although Uber drivers are perceived to be working *for* Uber, they are not actually employees of the corporation. Rather, Uber drivers are best understood to be working *under* Uber as legally they are defined as independent contractors (Rosenblat and Stark 2016; Zwick 2018). This definition allows space for idyllic neoliberal rhetoric glamorizing the flexibility and freedom that Uber drivers can have in their job, while it also enables Uber to minimize organizational costs by excusing the company from having to provide drivers with the materials needed for service (i.e. automotive vehicles) as well as traditional employee rights and benefits like insurance and a guaranteed minimum wage (Zwick 2018). This strategy forms part of Uber's "neoliberal playbook" that ultimately functions to keep the corporation cost-competitive in the contemporary capitalistic marketplace all the while making the workers' conditions less stable (Zwick 2018:682). The situation of Uber drivers in relation to the company they are working under is salient because it plays a determining role in the shape of the forms of control that Uber maintains over its drivers despite their supposed independence. Carrying forward this conception of Uber’s neoliberal business practices, I will now turn to entangle these practices with a contemporary conception of Max Weber’s bureaucracy to show how Uber’s algorithmic rating apparatus is asserted as a legitimate way to manage the Uber driver workforce through digital quantification.

## Digital Bureaucracy

Max Weber first conceived of the bureaucracy as a system of social control that "rests upon expert training, a functional specialization of work, and an attitude set for habitual and virtuoso-like mastery of single yet methodically integrated functions" (Weber [1909-1920] in Lemert 2016:89). Further, these functions, or "the management of the modern office is based upon written documents ("the files")...There is, therefore, a staff of subaltern officials and scribes of all sorts" (Weber [1909-1920] in Lemert 2016:86). In this conception, the bureau privileges Western, techno-scientific rationality as a predominant cognitive orientation as supported by the notions that bureaucrats will carry out the functions of the bureaucracy efficiently and objectively through standardized files. As a result, the bureaucracy is able to employ this rationality to legitimately claim authority over those it governs on "rational grounds – resting on a belief in the legality of enacted rules and the right of those elevated to authority under such rules to issue commands" (Weber [1909-1920] in Lemert 2016:93).

Weber's original theory of bureaucracy has been picked up by Muellerleile and Robertson (2018) who advance it into the current moment, accounting for the advent of digital technology and its applications in bureaucratic realms. As these contemporary theorists propose,

> the digital bureaucracy would appear to achieve its legitimacy through apparent data and algorithmic neutrality. This is best exemplified through the discourse of "big data" where "big" implies not so much a large size, but that data is unfiltered, unadulterated, and semi-autonomous—as if it can speak for itself, or for "all of us." These ontological assumptions about

> data help justify the realist conception that people, institutions, and technical sensors simply collect data about a world that already exists. (Muellerleile and Robertson 2018:202-203)

In other words, in today's digital age, data has replaced "the files" that Weber notes are the material form of the bureaucracy's objectivity and facilitate its operation, so it is these data through which we can trace the legitimization process. When represented quantitatively, these data take on an especially objective character as numbers are assumed to be universal, unbiased by human cognition, and therefore can be rendered as trustworthy. As a whole, this conception of digital bureaucracy suggests that "quantification and numbers, with their distinct air of relative disinterest, produce legitimacy for the governance mechanisms of neoliberal market society" (Muellerleile and Robertson 2018:205). The perceived objectivity of numbers allows those who see these numbers to trust them on grounds that they are accurate representations of what they are purported to measure, producing legitimacy for their use in governance and management systems.

In the context of framing Uber's algorithmic rating apparatus as a neoliberal governance mechanism used by a digital bureaucracy, the aggregation of 500 individual consumer ratings produces the average rating score which obtains an objective character by way of the big data paradigm as well as the generally accepted neutrality of algorithmic mathematical averaging. Further, driver ratings are efficient as they are collected and circulated to the company in real-time, further bolstering the rationality of this tactic. This in turn gives Uber rational grounds to claim legitimacy for this strategy of managing drivers. However, this governance of drivers inherently depends on consumers' continual provision of driver ratings in the first place. In this sense, while Uber consumers are not necessarily bureaucrats as they are not working for nor compensated by Uber, they do take on some bureaucratic labor in rating the drivers they ride with. This labor serves Uber by helping to reify the algorithmic management system as a legitimate form of domination over drivers.

While this advancement of Weber's bureaucracy may be valuable in describing the organization of Uber's driver management system, one critique of Weber is that he does not adequately account for the functioning of power in social relations. In order to fill this gap in the theoretical framework and account for the relations between company, consumers, and drivers, I will now turn to Foucault and his conceptions of power and discipline.

## Power and Discipline

In order to account for Weber's lack of attention to power detail, we may invoke Foucault's conceptions of power and discipline as they are constituted and function within the relations of this contemporary digital bureaucracy. Foucault, in theorizing on the function of power and discipline, notes that "power is everywhere; not because it embraces everything, but because it comes from everywhere" (Foucault [1976] in Lemert 2016:361). In other words, power is pervasive and present in every social interaction, yet contextually situated. The form power takes in a specific context is determined by the relationships between the parties engaged and interacting in this sphere. As a result, power also produces material effects for each engaged party at each power site. Power in turn can also be used to discipline others toward "a certain policy of the body, a certain way of rendering the group...docile and useful" (Foucault [1975] in Lemert 2016:325).

Applied to the Uber context, power is visible in the triadic relationship between driver, rider, and corporation. Because the algorithmic rating apparatus mediates these relationships, we can posit it as a site of this power where these force relations coalesce. Upon inspection of the force relations present in the Uber context, the corporation, by way of aggregating riders' rating data, is the arbiter of employment with the ability to deactivate drivers due to a low average rating score with little to no recourse for deactivated drivers. This power relation best serves the interests of the corporation and riders as they are both concerned with maximizing positive user experiences and minimizing negative user experiences in order to maintain usership and enjoy using the service, respectively. This is ultimately accomplished through the corporation's deactivation of drivers deemed unworthy of employment.

The Uber rider may also be seen as privileged in the power relation between them and the driver on economic grounds as the rider is more or less directly paying the driver in return for a service. As such, in this neoliberal economic exchange there arises an expectation that the driver will meet the reasonable requests of the rider. Further, the corporation allows for the temporary transfer of this power over the driver to the rider after each ride by giving them the ability to rate. While drivers are also given the opportunity to rate riders after each trip, the apparent notion that riders do not experience concrete effects of poor ratings exemplifies the unidirectional functioning of this power relation between the rider and the driver. When a rider does rate the driver, the corporation's power vis-à-vis the driver is temporarily transferred to the rider, but riders then let this power cycle back toward the corporation when their rating data is collected. In this way, Uber riders can be considered the "middle managers" of the Uber driver workforce as they assume the function of evaluating drivers as requested by the company (Stark and Levy 2018:1212).

Additionally, as riders are not required to rate drivers yet Uber's neoliberal digital bureaucracy depends on quantifiable rider ratings to legitimize its governance, the company also has a vested interest in disciplining riders to rate drivers after each ride. In other words, Uber as a corporation desires its consumers to be "the ideal subject of late capitalism midway within hierarchies of everyday surveillance: just as the subject is herself surveilled, she is also positioned as responsible for managing others classed as subordinate" (Stark and Levy 2018:1203). While the company maintains power over drivers by way of the rider, they must also maintain power over the riders through some other disciplinary force. This disciplinary force will be rendered visible in the analysis through participants' responses about why they rate drivers.

Overall, the result of the algorithmic rating apparatus employed by Uber and the power relations it produces is a hierarchy in which the driver, the rider, and the company are situated relationally with the company atop the order, riders at an intermediary level, and drivers subject to discipline from both. As such, the power produced in these relations serves each party disparately (Chan 2019). While neoliberalism, digital bureaucracy, power, and discipline are useful concepts in fleshing out the organizational and power relations between Uber, riders, and drivers, the mandatory face-to-face interaction between drivers and riders during an Uber ride has thus far been neglected. In order to account for this microsociological level of analysis, I will incorporate Goffman's theory of faces and facework in interaction to frame the in-car encounter between riders and drivers as well as the individual rating given by the rider after the ride.

## Faces and Lines

Goffman's theory of facework and its constituent components is based in sociology's theoretical tradition of symbolic interactionism which was originally conceived by Blumer (1969) as resting on three premises: (1) that individuals act toward things based on the meaning one has for them, (2) these meanings are derived from the social interaction one has with others, and (3) that these meanings are negotiated and modified with others in interaction through an interpretive process. In other words, the ways we understand, orient ourselves in relation to, and behave toward everything and everyone in social life arise from our interactions with others. In this regard, symbolic interactionism forefronts the individual and the meaning-making process that occurs in our interactions with others, which fills in the microsociological gaps left by the other theoretical concepts that constitute more macrosociological and mesosociological frames, respectively.

Turning specific attention to Goffman, in his work on the presentation of self in interaction he develops the concept of a "face" or "the positive social value a person effectively claims for himself by the line others assume he has taken during a particular contact. Face is an image of self delineated in terms of approved social attributes" (Goffman 1967:222).[1] In order for any individual to present a face to others in an interaction, the individual must follow a "*line* – that is, a pattern of verbal and nonverbal acts by which he expresses his view of the situation and through this his evaluation of the

[1] Please excuse the reproduction of the patriarchal terminology used by Goffman in his original works. In this project, these theoretical notions are applicable to individuals regardless of their gender and other social identities.

participants, especially himself" (Goffman 1967:222). If one's face is the intended image of self to be presented by the individual in interaction, the line is the interactive steps the individual takes to embody that idealized image of self. Further, when an individual's line or behavior is consistent with the judgments of their face as made by other participants in the interaction, the individual is said to "*be in face*" while inconsistencies between one's line and face render them "*in wrong face*" (Goffman 1967:223). As these symbolic interactionist concepts are primarily concerned with face-to-face interactions and the meanings produced in each situation, they will be invoked to account for the interactional process that plays out between Uber drivers and riders when they come into contact during a ride.

To contextualize these theoretical concepts in this project, both the driver and the rider have a face they pursue by acting along the lines they believe will allow the other party to perceive them as being in said face. For the rider, they may attempt to pursue a face that they are pleasant to be with during the ride by way of a line that includes exchanging pleasantries with the driver. On the other hand, drivers may follow a line to show the rider they are a good service provider by following traffic laws. Intrinsic to these notions of faces and lines is the process by which the rider judges the consistency between the driver's line as carried out in the interaction and the face the rider expects the driver to take in order to determine if the driver is in face or in wrong face. If the rider expects the driver to pursue the good service provider face, it follows that meeting the rider's criteria for being a good service provider is the corresponding line the driver should take. Therefore, Uber consumers' criteria constructions will serve as the basis for analysis as to what lines are available for drivers in order to succeed in being judged in face.

This judgement is also quantified by the rider's rating of the driver after the ride. As such, a driver deemed to be a good service provider and therefore in face will be rated highly while a driver that does not meet the rider's expectations will be in wrong face and likely rated poorly. Due to the unidirectional power relation that privileges riders over drivers, riders have a relative disinterest in successfully being in face as there are limited threats when drivers find them to be in wrong face. Comparatively, drivers, who could be subject to deactivation if they are not rated highly enough by riders, will want to avoid being in wrong face as it may lead to poor ratings that are reflected in their average rating. As the driver's average rating score is constituted by a series of individual riders' judgements of the degree to which the driver is in face, the average rating score may constitute another sort of face for the driver; one that is digitally quantified. It is important to note that this digitally quantified face is of different essence than the faces outlined by Goffman. A Goffmanian face is an ideal image of self that is pursued in encounters with others while the digitally quantified face is a historic, accumulating, and dynamically updated measurement of one's achievement (or lack thereof) of being judged to be in the expected Goffmanian face by other interactants across common contexts. Furthermore, in the context of an Uber driver's digitally quantified face, riders can see the average score of the driver from the moment they are matched, even before the driver arrives at the pickup location. This shows how these digitally quantified faces circulate in the digital sphere and thus can precede the individual and any attempt they may make to embody the appropriate Goffmanian face. Finally, as a digitally quantified face is represented numerically it may obtain a more objective definition, again by way of the big data paradigm, than the Goffmanian faces that are conceptualized by each individual in the interaction. Carrying forward these notions, I will now turn to map the action of riders giving drivers poor ratings to the facework process Goffman outlines for when individuals are judged to be in wrong face.

## Facework

When individuals are found to be in wrong face by other participants, they may engage in "*face-work*" or attempt to "make whatever he is doing consistent with face" (Goffman 1967:226). One way an individual may engage with facework to regain face is through "*the corrective process*" which Goffman defines as

> When the participants in an undertaking or encounter fail to prevent the occurrence of an event that is expressively incompatible with the judgments of social worth that are being maintained, and when the event is of the kind that is difficult to overlook, then the participants are likely to give it accredited status as an incident–to ratify it as a threat that deserves direct official attention–and to proceed to try to correct for its effects. (Goffman 1967:230).

Goffman continues to outline this corrective process in its four steps: (1) the challenge whereby participants call attention to the offender being in wrong face, (2) the offering whereby the offender is given the opportunity to correct their behavior and regain face, (3) the acceptance whereby the other participants may accept the offering, and (4) the thanks, or gratitude shown by the offender to other participants for their acceptance of the offering. As previously mentioned, poor ratings from riders may result from a driver being judged to be in wrong face by the rider. Therefore, a rider giving a driver a poor rating may constitute their communication to the driver that they are in wrong face - or a challenge that begins a corrective process. In the subsequent analysis, I will cast the ways in which Uber riders rate against this step in the corrective process, showing the range of challenges made by different rater types when drivers do not meet the rider's criteria.

One limitation to the application of Goffman's facework as originally conceived to the context of Uber is the interest of an individual to maintain a particular face over time. As Goffman notes for an individual trying to convey a particular face, "an encounter with people whom he will not have dealings with again leaves him...free to suffer humiliations that would make future dealings with them an embarrassing thing to have to face" (Goffman 1967:223). In other words, if one's fellows in interaction will not be encountered again, the individual is free to be in wrong face without consequence outside of that particular interaction. In the context of Uber, due to the number of drivers and riders located in a densely populated urban area, a driver may only come into contact with a particular rider once. However, again harkening back to the function of Uber's digital bureaucracy, drivers may be deactivated as a result of the humiliations of being judged as being in wrong face by riders that Goffman assumes they would be free to endure. In order to advance this theory to the contemporary moment, we must account for the fact that these judgements of a driver being in wrong face are tracked through time and space by way of the driver's digitally quantified face. Due to our increased capacity to produce, publish, and continually modify these numbers, we can see how an incident of a driver being in wrong face in one interaction will be represented in the driver's digitally quantified face and carried forward to their subsequent interactions with other riders. In sum, while Goffman could not account for the sustained tracking of facial judgements as they take quantified form, this premise is made trivial by this theoretical advancement that accounts for digitally quantified faces.

When taken together, these theoretical concepts will serve as a framework to be used when analyzing what Uber riders reported about their experiences using Uber and rating drivers. The neoliberal digital bureaucracy will serve as a frame through which we can map Uber's digital infrastructure and goal of receiving ratings from their consumers which in turn legitimizes this type of driver management. Foucault's notions of power and discipline will be applied to a multitude of forces in explaining why it is that riders may engage in the act of rating without mandatory obligation to otherwise. Finally, the steps of Goffman's facework process and how they play out in the interaction between Uber drivers and riders will be traced in relation to how riders deploy ratings unto drivers.

## Methods and Data Collection

This project employs nine semi-structured phone and in-person interviews in order to collect qualitative data for this project. Interviews ranged in length from 29 minutes to 50 minutes with a median length of 39 minutes. Seven interviews were conducted over the phone and two were conducted in-person at a location of the respondent's choosing to ensure comfort and confidentiality. An audio recorder was used to capture what was said during all interviews. Immediately after each interview, I wrote a short memo summarizing the conversation. After the interviews were conducted and recorded, I transcribed each audio recording into a text document that was then uploaded to the Dedoose online software platform for the subsequent data coding process.

"The interview has been called the primary method used in qualitative research (Burnard, 1994; Doody & Noonan, 2013; Myers & Newman, 2007; Ryan, Coughlan & Cronin, 2009; Schultze & Avital, 2011)," matching it nicely to the qualitative nature of the research questions this project explores (Oltmann 2016:1). The interview method served as a means to gather qualitative data about how consumers go about engaging in rating drivers and why they do so. The interview questions spanned a range of relevant topics including, but not limited to, how consumers decide which star level to choose when rating a driver, what criteria do consumers consider when evaluating Uber drivers, and what consumers understand about how Uber utilizes the rating data they produce.[2] The interview method was appropriately suited for this project because it enabled the respondents to expound on the nuances and intricacies of their participation in and understanding of these acts. This method also enabled me as the researcher to probe

[2] See Appendix A for full interview protocol

respondents for more details and complexity when their initial responses to questions were succinct.

In the analysis, three rounds of coding were done. During the first round, a set of initial codes developed deductively from my review of the literature was applied broadly to transcript excerpts and interesting anecdotes were noted. During the second round of coding, I was concerned with refining the initial codes, which involved the application of child codes, as well as coding for inductive themes recognized during and crystallized after the first round of coding. In my inductive code application, I attempted to code along methodological lines of

> Grounded theory [that] requires the researcher to use a specified set of procedures to code data in a series of passes (open, axial, and selective). Data are examined for dimensions and properties, compared with similar phenomena, regrouped and reconceptualized until a provisional theory emerges inductively from the analysis. (Neff 1998:125)

However, I will note that these inductive codes were not developed devoid of external influence as my continual research of literature and consumption of other relevant media surely influenced what patterns were rendered most visible. Thus, I do not claim to have developed a grounded theory, but rather that I took this process as a basis to inform the steps I took during the inductive analysis. Finally, the third round of coding consisted of refining and reshaping inductive codes (as I did with deductive codes in round two) and confirming the relevance of earlier coded excerpts to the applied codes. Early in the third round I seemed to reach analytical saturation with the collected data, so I was comfortable moving to begin synthesizing for findings upon completion of round three.

## Sample

The sample was drawn beginning with my personal networks and employed snowball sampling to gather additional contacts outside of my network. The sample consisted of nine individuals who have used Uber and rated a driver at least once in the past six months. To limit confounding variables, I only drew my sample from Uber consumers in a large city on the west coast of America. In order to recruit participants, I leveraged my personal networks using Instagram, email, and the senior sociology cohort to make contact with five initial qualifying volunteers. In snowball sampling from this first round of contacts, which to my knowledge utilized a neighborhood Facebook group and word of mouth, I made contact with an additional four volunteers. Although I attempted to break past the second degree of separation by asking second round participants to help me further snowball sample, this did not result in any additional volunteers. Although I knew about half of the respondents before starting this project, I do not think my dual positionality as a researcher and acquaintance introduced any exceptional bias in participants' responses as I had not spoken with any of them about this topic prior to the interviews.

## Ethics and Positionality

Ethically, all respondents gave informed consent by signing a form detailing the procedural and topical details of the interview before beginning.[3] Additionally, before each interview the respondent was verbally reminded that they had the opportunity to pass on any questions they did not feel comfortable answering and could request to end

---

[3] See Appendix B for full Informed Consent Document

the interview and have their data removed from the study at any point. I checked in on how respondents were doing periodically throughout the interviews, but no underlying issues seemed to arise nor did I sense any serious distress in our conversations. The topics brought up in the interviews were not likely to not be sensitive for respondents as they pertained to a social context that the respondents willfully enter by choosing to use Uber. This was seemingly confirmed during the interviews as no respondents opted out of any interview questions nor did any end the interview early.

However, it was important for me to preserve interview respondents' confidentiality in order to minimize the risk they are identified by participating in this study. This was accomplished by my efforts removing personally identifying information, such as names and locations, from interviews as they were transcribed. After deidentifying, I assigned each transcript an arbitrary letter of the alphabet and gave each respondent a pseudonym that I came up with impulsively using the letter their transcript was identified with. Throughout the data analysis and writing process, all data was kept digitally in password protected files and accounts on a password protected computer. All audio and written copies of transcripts were deleted upon completion of this project.

As an Uber consumer myself, I was afforded general insider status as I was familiar with Uber's app interface and generalities common to my own experiences using the app. Further, my positionality as having loose social ties with some of the respondents enabled me to easily maintain rapport during these interviews. I attest that for these respondents with which I had loose ties, I had not shared my perspective on this subject matter with them before the interview. For respondents with which I did not have loose social ties before the project, I mainly established and maintained rapport by informing them that this is an academic project with no ulterior funding, motives, or interests. I attempted to mitigate the introduction of any researcher bias stemming from my positionality by framing open-ended questions neutrally and attempting to withhold my personal anecdotes and perceptions about Uber throughout the interviews.

## Limitations

These recruitment and data collection methods do introduce limitations to the sample and analysis. First, I acknowledge that the convenience of my sample has introduced disproportionate representation and biases. The sample was disproportionately white and college-educated, however participants were more evenly distributed along demographics such as gender and age. Socioeconomic status was quite ambiguous in the sample as I did not ask participants to disclose their income, however this project was only accessible to people with the resources to volunteer 30-50 minutes of their time to participate in an interview without any compensation. This may have resulted in disproportionate representation of individuals with the privilege of adequate leisure time as well as disproportionate representation of individuals who hold more extreme views of Uber such that they wanted to volunteer this time to speak with an academic researcher. I attempted to dampen this by obscuring the focus of my analysis, the rating, throughout the recruitment and interview process, but this limitation may still have led to the inclusion of consumers who rate drivers more extremely, or more or less often, than the general Uber consumer population in the area of interest. Furthermore, while some respondents noted similarities between using Uber in the city of interest and other cities around the United States, Uber does operate on a global scale. Due to the fact that an individual's participation in rating an Uber driver may differ due to other subsuming cultural values based on their race, class, or geography, these conclusions should not be generalized widely to individuals or groups outside of those represented.

Finally, sociologists have documented that what people say is not always consistent with what they do and that interview researchers are prone to committing “the attitudinal fallacy—the error of inferring situated behavior from verbal accounts” and suggest that studies concerned with action should be approached using an ethnographic method where the researcher observes behavior (Jerolmack and Khan 2014:179). This project admittedly commits this attitudinal fallacy as it is concerned with concluding how and why Uber consumers rate drivers based on what Uber consumers said during an interview. However, in writing these findings I attempt to frame all data as something that a respondent says, forefronting their acts of saying, recalling, noting, riffing, and opining as they did in the interview. Also, conducting interviews was the methodology that was logistically feasible and flexible enough to allow me access to the most detail as to how consumers participate in the driver rating process. It would be near impossible to validate the respondents’ verbal claims with their behavior because Uber maintains a data asymmetry whereby the corporation and the user are the only two parties that ever know what ratings consumers give each driver. Thus, I did not have access to an objective account of respondents’ individual rating histories. Furthermore, conducting an ethnographic project by riding along and observing Uber consumers rating drivers in real time was impossible due to geographical distance, temporal variation in when consumers actual rate drivers, and would have introduced a reactionary bias whereby the participants may have rated drivers differently when being observed by a researcher than they would have otherwise. I would like to believe that given the protection of confidentiality and trust in the researcher, respondents were not actively deceiving me with their responses and that what each said in their interview is relatively consistent with their behavior in action most of the time.

It is important for me to acknowledge that these limitations make my findings not widely generalizable across identity groups, but hopefully protects it from claims that it is tokenizing the individuals who participated. Rather, the generalizability of this study may pertain to the applicability of the findings to other rating contexts. In my exploration of how and why Uber consumers rate the drivers they interact with, I hope to offer an abductive analysis that is more so concerned with taking the limited observations obtained and generating the most realistic hypothesis that will then serve as a basis for future empirical tests on a larger scale (Timmermans and Tavory 2012).

## Findings and Analysis

The following section draws upon the review of literature, theoretical frameworks, and data collected to answer the central research questions: how and why do Uber riders engage with rating drivers? In the first section, I discuss Uber's discipline of riders to rate as it manifests in the app. Then, I transition to define the general notion of how Uber consumers' criteria construct Goffmanian faces for drivers to embody and trace the facework process in the context of these interactions. I will then show how consumers respond digitally when drivers are deemed to be in face. Finally, I conclude with an analysis of the range of responses different types of raters have when they find drivers to be in wrong face. These responses include quantified challenges, mediated challenges, facial protections, and mediated facial protections.

### Discipline to Rate

Rating data does not come from nothing; it must be generated by someone. Uber's algorithmic rating apparatus, similar to many other online product review platforms like Amazon and Yelp's, relies on consumers who have used the service to voluntarily rate their experience in order to produce rating data for whatever or whoever is being evaluated. This production of data requires time and labor from the consumer, as Bella summarizes rating products as "having to go back onto the website and like purposefully click through something." For consumers like Irma, this results in them rating products "really only if something is at one end of the spectrum or the other so if it's really exceptional or it's really terrible." Similarly, Bella will only go through the effort of rating a product if it did not live up to her expectations "more so that other people wouldn't have to spend money on the bad thing." Although only one respondent reported regularly going online to rate products they buy, all respondents reported rating their Uber drivers on a relatively regular basis. With their seemingly more frequent rating of Uber drivers than material products, the questions follow: why do Uber consumers rate Uber drivers with such frequency?

Within the sample, the most commonly reported reason consumers rate Uber drivers is because of Uber's in-app rating prompt. The rating prompt appears on the user's screen at the end of the ride, and does not allow users to engage in other in-app actions like requesting another ride until they have either gone through and completed the rating act or have chosen to opt out of rating by clicking outside of the rating prompt window. After each ride, Flora notes the rating prompt "pops up automatically afterward so it's like as long as I'm on my phone right there I'll be like okay cool 5 stars." In Flora's case, she often still has the Uber app open at the end of the ride and as such is almost immediately prompted to rate, so she does just that. Yet this form of discipline also has a persistent temporal quality insofar as the prompt remains present and prompting until user action is taken. For those who may exit out of the Uber app before their current ride ends, the prompt confronts these users the next time they open the app for whatever reason, regardless of how long it has been since that last completed ride. Adam describes his experience with rating his Uber drivers as such: "I'll always rate them but sometimes it'll be like a couple hours after...I just kind of forget and then the next time I open the app the rating thing will pop up or still be there and so then I'll just rate them." Likewise, Bella notes "I usually don't end up rating the person until a couple days later when I open the app again. Like there's probably one sitting in my app right now that's just unrated and probably has been for like 2 weeks" and Emily says "I don't rate until I order the next one 'cause it pops up your last ride and then I'll just you know rate and tip."

For users like Adam, Bella, and Emily, their experience rating drivers is reminded by the prompt when they reopen the app for whatever reason and it appears front and center on their device's screen. Compared to product ratings prompts that are usually distributed to consumers using digital messengers like email that exist external to the producer's interfaces (i.e. websites or apps), the Uber app's driver rating prompt is integrated into the platform that serves other functions beyond garnering ratings from consumers. This integration allows for the prompt to obscure and prevent the consumer's host of potential future actions, including requesting their next ride, until they have engaged in the rating act or clicked away from the prompt. This renders the Uber app's discipline of consumers to participate in the rating apparatus stronger than that exerted over consumers in material product rating prompts. The strength of Uber's rating prompt as a site of disciplining power is also directly correlated to the corporation's ability to garner the adequate amount of consumer data needed to facilitate the function of the algorithmic rating apparatus and legitimize their use of average ratings to manage drivers. Although the consumer's goal may be to continue using the app, the company's goal is to simply solicit a rating from the consumer for their most recent ride. The prompt as a disciplining force aligns these respective goals and gives the consumer a singular avenue through which the fulfillment of both parties' goals is achieved. Asserting the in-app prompt as one of the primary reasons consumers are disciplined to rate, I will now turn to analyze the landscape of the criteria constructions used by Uber consumers in their rating of drivers.

## Rider Criteria and Driver Faces

As riders rate drivers after each ride as a way to judge or evaluate them, riders must come up with a basis upon which they make this judgement. As drivers are judged in relation to these bases, or criteria construction, they can be conceived of as the ideal Goffmanian face the driver has an interest in pursuing through a corresponding line in order to be held in high esteem by the rider. Yet, as these criteria constructions are subjectively formed by each rider, we find agreement, differential holding, and criteria contradiction among respondents. Agreement occurs when multiple riders agree that a specific criterion is of a specific importance. Differential holding consists of agreement between riders on the criterion, but each rider may feel this criterion is of different importance. Finally, contradiction occurs when respondents hold oppositional stances on a criterion when evaluating.

Amongst the respondents interviewed, there was convincing agreement that safety was the most important criteria to consider. Daisy notes amongst her criteria used to evaluate drivers "obviously the big one of like safety...as you know they must be a safe driver." Gina also says in explaining her rating criteria "I expect them to be good drivers." She goes on to define good drivers as "safe drivers that take less risks than the traditional taxi driver." This criterion also extends to the physical car space, as Harry notes "I expect to have obviously like working seatbelts." Despite the sense that this criterion is commonplace and obviously important, safety is still somewhat ambiguously defined outside of following traffic laws. This shows that even within agreed upon criteria, rider subjectivity presents an ambiguity that makes the lines that are available to the driver in order to embody the appropriate face unclear.

Further confusing the patterning of rider criteria constructions is the differential holding of criteria like efficiency and contradiction as to whether rider's expect drivers to follow the GPS route outlined for them by the app. Regarding efficiency, Daisy holds this criterion of moderate importance as she explains good drivers are "generally on time or within a few minutes of when the app says." Additionally, regarding navigation practices Bella explains "I appreciate someone who is following the directions that the Uber app gives them 'cause sometimes you'll have a driver who you can tell is like going off of their own like mental map of the city and kind of like makes weird turns that aren't on the map which I always feel is like a little sketchy." Compared to Daisy and Bella, Carl is a rater that exalts efficiency almost to the level of safety and expects drivers not to follow the app's navigation suggestions entirely in order to achieve maximal efficiency. He recounts

> disappointing experiences are when people are literally relying on what the app is telling them what to do and clearly like visually if they just look where they're driving it doesn't make sense to go that way or the traffic pattern...you know...I want someone who has street feel because their full time job...or you know their current job is to drive you somewhere efficiently. So I like when folks are efficiently driving.

Despite the tension between Carl's perception that Uber drivers are working full-time leading him to expect that they will have and use knowledge of the area and its traffic patterns and the fact that not all Uber drivers are working for Uber full-time and may not have this knowledge, we can see that Carl likens the driver's entire service to achieving an efficient route. Comparatively, Daisy expects drivers to be on time, but she is willing to account for general minor inconsistencies, showing this criterion is less important to her than it is to Carl. This differential holding shows how even when consumers agree a criterion should be considered when evaluating the driver, there is variation in how important that criterion is to their overall criteria construction. Furthermore, Carl's expectation that drivers will not totally rely on the app's navigation suggestions contradicts Bella's sentiment that rides during which the driver diverges from the route are "sketchy," showing how different consumers can have opposing expectations for Uber drivers.

These examples of differential holding and contradiction show how different raters may make different meanings out of the similar interactive situations, leading to a range of possible criteria constructions. Coupled with the variation that may exist even within consumer agreement on a criterion, differential holding and contradiction further disprove a monolithic semblance of how consumers construct criteria. In sum, the rider's subjective criteria construction serves to define the situation and informs the Goffmanian face the driver will be interested in embodying in order to receive a high rating from the rider. However, to some degree these faces are ambiguously defined across many contextually similar situations due to the variation in how each individual consumer constructs their criteria. As a result of this ambiguity, it may be difficult for the driver to determine what line they should take at the outset of a ride.

## The Digital Facework Process: Retroactive Challenges and Proactive Interventions

Further making it difficult for drivers to determine what line they should take during a ride is the notion that no respondents reported ever communicating with the driver what they expect from them at the start of the ride. This tendency for riders to not communicate their expectations to the driver continues throughout the ride, even when drivers are judged by riders to be in wrong face during the encounter. Harry recalls during a particularly poor ride experience “there were multiple instances of [unsafe driving]. I would have been like hey we need to end this ride I don't feel comfortable with how you're driving, but at the time because it was such a short ride I was kind of like I don't really need to say anything.” This driver’s behavior did not align with Harry’s safety criteria thus leading Harry to judge this driver in wrong face. However, due to the expected short temporal nature of his interaction with this driver, Harry did not feel the need to say something. Had Harry said something to the driver about this inconsistency between his expectation and their behavior, this would have constituted a challenge to the driver’s Goffmanian face as it would have called attention to his judgement that the driver was in wrong face. Harry’s avoidance of making challenges during the interaction is consistent with Carl’s sentiment that “even if I had a negative experience I will always thank [the driver]. I usually just try to you know say thank you have a nice day have a good afternoon...Just some sort of pleasant goodbye no matter what I'm going to rate them at.” In other words, even if Carl judges the driver to be in wrong face as a result of his negative experience, he will avoid making the face-to-face challenge by thanking them at the end of the ride. Also, Irma explicitly notes “I'm not going to feel comfortable telling an Uber driver directly what I was unhappy with because I'm not confrontational. But if someone is real bad...I would be more inclined to give them a low rating afterwards.” In this case Irma recounts her personal discomfort with confrontation leads her to avoid challenging drivers’ Goffmanian faces during their face-to-face interaction as well. All of these examples of riders not making face-to-face challenges to drivers they deem in wrong face may also be framed as the rider's own facework process by which they do not want to be perceived as the driver's disciplinarian by calling out their failures and as such avoid the in-person challenge that would cast them so in the eyes of the driver.

However, as Carl and Irma note, they would be inclined to rate this driver poorly after they have exited the car, which would have a negative impact on the driver's average rating and digitally quantified face. As such, we can see how poor consumer ratings constitute challenges to the driver's Goffmanian face that are permanently logged in the driver's digitally quantified face. Further, given the company's disciplining of riders by way of the rating prompt, consumers may be more inclined to pursue this digital route of challenging drivers over the alternative of a face-to-face verbal challenge. These digital challenges are made retroactively as riders do not rate until after they have left the car and the face-to-face encounter with the driver is over. Therefore, the corrective process of facework occurs in a digital space where there is no agency for the driver to be reactive. As the rider is no longer with the driver when they rate, the driver cannot make an offering in an attempt to regain their Goffmanian face after the rider's digital rating challenge has occurred. Thus, the corrective process cannot be carried out in full and the driver is subject to endure this humiliation to their digitally quantified face without recourse.

Further preventing the corrective process from executing in full is the notion that most raters' challenges do not contain any information that identifies why the rider judged the driver to be in wrong face. Daisy and Carl were the only respondents who mentioned giving qualitative comments to drivers frequently as Carl riffs “I'll share explicitly what gave my rating because to me if I was them I would be frustrated to get a negative rating and not know why. So I will try at least...you know whether or not they agree with my star rating I'll try to just write like why I did that so they can improve” and Daisy adds “I do like that they've added lately like buttons that you can click on to suggest why you rated the ride that way.” While they still prevent drivers from making an offering and completing the corrective process without decreasing the driver’s digitally quantified face, these raters at least inform drivers about what their shortcomings were in that interaction, better informing the line that the driver may want to take in future interactions. While rider criteria constructions may range on a subjective basis as previously discussed, this qualitative feedback would be useful regarding commonly agreed upon criteria as well as criteria that are differentially held. On the other hand, the fact that seven of the nine respondents do not usually make these suggestions to drivers shows how drivers are left to wonder how to avoid being judged in similar lights by future riders with no basis for correction at all.

In response to rider tendencies to make retroactive digital challenges that bar drivers from making corrective offerings and affect their digitally quantified faces, drivers may proactively intervene in the rider’s rating process. These interventions attempt to incite an in-person challenge before the rider’s digital challenge is made. Daisy notes this by saying “20% of the time there might be like a sign or some reminder saying like please you know...give me a five out of five and if you don't want to do that let me know.” These signs are proactive interventions as they are visible to riders from the moment they get in the car and thus before the rider may be able to judge the driver as being in wrong face. In this case, a five-star rating would not constitute a challenge as the driver's average score would increase, so the driver's sign indicates their acceptance of this outcome without feeling the need to intervene in the rider's rating process. However, in contexts where the driver would be judged in wrong face, these signs intervene in the rider's generally normal process of digitally challenging the driver's face by communicating the driver's willingness to discuss any challenges the rider may want to make in-person before they rate. In this regard, the driver can be seen as trying to discipline riders to make challenges in-person rather than digitally, which would enable the driver to maintain their agency to make an offering to the driver in response to the original challenge and complete the corrective facework process. While these interventions do empower the driver by allowing them the ability to challenge, Daisy reports only about one fifth of drivers actually do this. In turn, we must still keep close attention to the ways in which riders do end up deploying their ratings in the majority of these Uber ride contexts. Carrying forward the idea that a rider's judgement of drivers being in Goffmanian face during the face-to-face interaction is made digitally and therefore tracked in the drivers' digitally quantified faces, the following sections will explore the nature of these judgements as they are quantitatively made under the five star rating system by different types of raters in different contexts.

## Drivers In Face

Recalling the general notion that a rider's rating criteria inform the character of the Goffmanian face a driver is interested in pursuing, we explore the effects of interactions during which the driver is judged by the rider to be in face. In Uber rides where the consumer's most important criteria, including safety, are met, most all consumers will give the driver a five-star rating. Amongst the respondents, seven report rating their Uber drivers five stars the majority of the time. Adam notes “for me personally if they do a good job, five stars. Like, I don't get really picky in how I rate them.” and Carl says “I feel like my expectations are reasonable or somewhat low so I'm willing to give someone five stars if they got me where I needed to go, there was nothing negative about the experience.” Likewise Daisy feels “unless it's a pretty poor experience or I feel like they did something that I really took issue with I probably will default to a five” and Flora explains “I feel like at this point five stars is like the default and like you have to do something pretty wrong for me to go down from there.” Furthermore, Harry notes “I think...for the most part I want to say I've always given a driver five stars” and Irma says “most Uber drivers I give five stars...as long as there's nothing exceptionally wrong with the drive I will give them a five star.” Noting the consistent claim that in order for a rider to rate the driver five stars the rider’s expectations must be met and nothing else the rider would find unsavory occurred, we can see that the driver’s behavior in the interaction, or their line taken, is consistent with the ideal Goffmanian face as defined by that rider’s rating criteria construction. As such, these rides constitute interactions where the driver is perceived by the rider to be in face. Continuing, we see that most riders reward drivers they deem to be in face during their encounter by rating in a way that will increase the driver’s average rating score (assuming their average is not at a perfect 5.0 stars before the ride). Thus, this reward is reflected in the driver’s digitally quantified face.

However, as there were inconsistencies with how riders constructed rating criteria, there are also inconsistencies in how riders deploy their rating when they deem the driver to be in face. Gina, whose criteria have been met in most all of the rides she has taken, does not align with other raters for these driver in face contexts as she notes “I pretty much give everybody a four 'cause I'm a...I don't know if that's a tough grader or not but whenever I'm grading anything I never give anybody the top spot because there's always possibility to do better.” Although she still perceives the drivers she gives four stars to be in face during their interaction relative to her constructed criteria, she does not quantitatively reward these drivers’ digitally quantified faces as other riders would. Instead, as most active drivers have average scores above four stars, Gina’s actions may be seen by drivers as challenges stemming from their being in wrong face during their face-to-face encounter with Gina or other similar raters. The notion of this inconsistency is also captured by Carl who, although he rates drivers who meet his criteria with five stars, says each individual

> rating will, in it's very inconsistently small way, influence [the driver’s] average. So if they're a 4.9 rated driver and I have a good experience but I rate them 4 out of 5 I'm basically punishing them. So it's funny how you know a 4 out of 5 is a good rating in many other contexts but if you're giving them a 4 out of 5 and their average is 4.9...you're lowering their average. So to them it's like anything below 4.9 would be bad.

This sentiment purports that riders’ judgements of drivers being in or in wrong face can be collapsed to a binary whereby the judgement maps to the simple effect the individual rating has, be it an increase or a decrease, on the driver’s digitally quantified face. While this may be salient to contextualize raters like Gina and explain the inconsistency between her judgement of drivers being in face and the effects she has on drivers’ digitally quantified faces, this logic overlooks the potential for differential outcomes based on quantifiable differences between the driver's average rating and the individual rating given by a rider after an encounter. The following sections will explore the range of quantifiable challenges riders can make to drivers' faces in their rating of interactions where the rider judges the driver to be in wrong face.

## Full Spectrum Raters: Quantifiable Challenges

Full spectrum raters are riders who are willing to engage with the entire range of the rating system to evaluate drivers in a way they deem matched to the outcome. Within the sample, two Uber consumers reported having ever rated a driver at either the one-star or two-star level, which is indicative that they are full spectrum raters. The first of these is Daisy, who rated a driver two stars recounting " I was probably more concerned for my safety than usual...he blew through an intersection where I think he thought he had the right of way to turn and someone else just drove through the intersection and it was a close call to say the least." Also in this category of full spectrum raters is Carl, who mainly values efficiency in an Uber driver. When recalling his worst Uber experience, Carl reports it being one when he requested

> kind of a standard route that I would take...going to my doctor's office and then back to my job...the person was extremely late like you almost wonder if the app is broken 'cause it looks like they're sitting a block or two away so it was unclear why it took them so long to get me...it was sort of like this culmination of like it's not the experience I ordered, I ordered an UberX to get somewhere quickly, the app said I would get there on time, the driver didn't meet that standard...so like even before I even got in

> the car I was frustrated with this driver and then they were just not efficient with driving so I ended up late to a meeting at work because of all that.

Carl ended up rating this driver one star citing “the most negative experience is when they just totally aren’t driving logically.” Between Daisy and Carl, they both have strong criteria of safety and efficiency, respectively, which in their worst Uber experiences were violated by the driver. In this regard, these drivers were judged as being in wrong face. As such, Daisy and Carl initiate a corrective process in challenging the driver’s face by rating them poorly, or calling attention to the fact that the driver did not meet their expectations. The driver receives this challenge by way of their digitally quantified face, as after a driver receives a rating lower than their current average score, this average score decreases. When a driver sees that their average rating score has dropped, it calls attention to the fact that a rider has not rated them at the generally normal five-star level, indicating to the driver their inability to meet the rider’s criteria and thus an inability to be in appropriate face. Furthermore, these challenges are quantifiable insofar as individual Uber rides are commensurate under the five-star rating system. Therefore, the lower the individual rating given by the rider, the larger the drop in the driver’s average score, indicating greater discrepancies between the driver’s interactional face and the line they followed in the interaction.

The quantifiability of these challenges is further apparent when we cast full spectrum raters against partial spectrum raters. In the sample, four consumers were identified as partial spectrum raters, or raters who have had experiences where they have rated as low as three or four stars. This rater type is constituted by two rater subtypes: the holding type and the forgiving disciplinarians. Holding type partial spectrum raters are those who have not experienced extreme rides and thus they did not feel the need to rate below three stars. However, while no holding type partial spectrum raters had ever rated a driver below three stars, most all did conceive of situations under which they may give drivers a one-star or two-star rating, indicating they may be cast as full spectrum raters if they were to experience an Uber ride where their primary criteria were not met. These raters are held here insofar as they may be more specifically defined by another rater type under more extreme circumstances.

Holding type partial spectrum raters help us discern the quantifiability of the challenges made to drivers' faces as a first step in the corrective facework process. Among them is Adam, who rated a driver three-stars when, as he recalls "I had a friend once close the door and he didn't slam it but it was a little bit more forceful than just slowly closing it and the dude got all upset. Like "Don't slam my doors!" like got all mad and it was the start of the ride so it was kind of like you need to chill man." Adam cites his conception of drivers like this as "mediocre, kind of sassy, three." In this instance, Adam's primary criterion of getting from point A to point B was fulfilled as he got to his destination safely, but one of his secondary criteria that drivers will be courteous and accommodating toward riders was not. Gina is another partial spectrum rater of holding subtype as she also rated a driver three stars when her primary criteria of safety was met but her secondary criterion of timeliness was not. She recalls "there was a lot of difficulty in...telling the driver where we were and then him being able to get there because...roads that are usually two way are only one way so the guy had a hard time getting to our location where he was coming from" and notes "three would be if there was only one thing wrong like the car was dirty or they showed up late or one minor thing." Although this experience takes on a similar criteria dimension to Carl's as they both pertain to timeliness and efficiency, Gina does not exalt this criterion as primarily as Carl does, therefore explaining the disparity in outcomes between these two raters. By casting Gina and Adam's partial spectrum rating against Carl and Daisy's full spectrum rating, we see how rider challenges to drivers' faces are also quantified by the rating system proportionally to how the rider constructs their criteria in the first place. When a criterion that a full spectrum rater holds primary to their criteria construction, as Carl does with efficiency and Daisy does with safety, is not met by the driver, the challenge made by the rider will be of greater magnitude than that coming from a holding type partial spectrum rater if one of their secondary criteria, such as efficiency for Gina or courteousness for Adam, is not met. This is derived again from the commensurate differences between the driver's average rating score and the score that constitutes the challenge. In other words, the amount that a driver's average rating will go down after receiving a one-star rating from a full spectrum rater is meaningfully larger than the amount that their rating would drop after receiving a three-star rating from a partial or full spectrum rater.

## Forgiving Disciplinarians: Mediated Challenges

The other subtype of partial spectrum rater is the forgiving disciplinarian. These raters are characterized by their experience of a ride where some of their primary criteria were not met, yet they were not willing to rate the driver below three stars. Harry is one of these partial spectrum raters who recounts

> I think the only time I've never given a driver five stars, I gave them maybe three or four, was when I felt really unsafe because...the driver

> made no attempt to actually come to my location...And then when we got off the freeway she had no idea how to merge and like we almost got in an accident. So that was where I was like oh gosh this is another red flag and then when she was pulling over...she pulled over in a really unsafe spot where it was like if someone's gonna come behind us they'd be blocking the intersection so like this is just not a good place to pull over because you could cause someone else to have an accident. So that was where I was like I didn't feel safe in that ride.

Harry's near-collision Uber experience is similar to Daisy's as safety is held highly in both of their criteria constructions as primary factors in determining how they rate and in both situations they judged the driver they were with to not be the safest. Furthermore, Harry holds efficiency of the pickup, which was also violated by this driver, as a secondary criterion in his construction which is in agreement with Gina's construction. Arguably, the driver Harry was with may be considered further in wrong face than those drivers Daisy and Gina were with relative to each rider's criteria construction on grounds that Harry's driver violated multiple of his criteria of both primary and secondary importance. Yet, Harry still went on to rate this driver three or four stars while Daisy rated her driver who only violated the safety criteria with two stars and Gina gave a similar rating for a less serious offense. By recalling that he did not rate this driver lower because he felt bad, Harry can be seen as a sort of forgiving rater who although their criteria were not met, may not rate a driver as low as a full spectrum rater would under like circumstances.

Additionally, Flora is another forgiving partial spectrum rater. In describing her rating deployments, she says

> it also takes a lot for me to give somebody a low rating. Just because I feel bad like it's their job or at least their side job but I know like the rating aspect is really impactful on their side. So I would say like if it's a situation where I felt unsafe, then maybe like three, four stars which like saying that out loud seems ridiculous but I feel like at this point five stars is like the default and like you have to do something pretty wrong for me to go down from there.

Although Flora has not experienced a ride where her primary criteria were not met, she also notes the difficulty she would have rating a driver below three stars, even in an encounter where the driver did not meet her primary safety criterion. As forgiving disciplinarians, Harry and Flora carry out mediated challenges to drivers who are in wrong face as they do not lower their rating to the same magnitude as full spectrum raters. These challenges are mediated by their consideration of the negative outcomes associated with a driver's average rating dropping too low and thus their consideration for protecting drivers' digitally quantified faces. In sum, forgiving disciplinarian raters will penalize a driver's digitally quantified face to a quantifiably less degree than full spectrum raters will in similar ride outcome contexts. Yet, forgiving disciplinarian raters still engage in the rating act, produce data for the corporation to use in their management of drivers, and in turn contribute to the quantified evaluation of drivers by lowering the driver's average score. Thus, these protections afforded to the driver's digitally quantified face do not replace the challenge that is inevitably still made by forgiving disciplinarians.

The following section explores another rater type, the empathetic rater, that extends the consideration of a driver's digitally quantified face wholeheartedly, offering full protection of face in place of a challenge.

## Empathetic Raters: Facial Protection

While full spectrum raters and partial spectrum raters engage with challenging the faces of drivers they deem to be in wrong face during their encounter by lowering the driver's average rating score, raters of another type, empathetic raters, actively forgo rating in order to protect the driver's digitally quantified face. In action, empathetic raters are those who have only rated drivers five stars and express considerable emotional response to the prospect of rating a driver less than five stars. As such, empathetic raters do not rate drivers they deem to be in wrong face below five stars. Rather, they opt to forgo the rating of the driver in these contexts.

Emily is an empathetic rater, as can be gleaned from her description of her rating deployment:

> I just either rate five stars or no stars....even in saying that I think oh this person I should have rated in between but I just didn't rate at all because he wasn't the most skilled driver...there's a couple times where in the one short ride he had taken a left and got into the wrong lane and quickly swerved into the right lane...or the correct lane...And I just didn't rate him so that would be one and another one was just recently where...he seemed very agitated...but he got us there safely and everything he just I think was agitated at the traffic, so then I just didn't rate him.

Irma is another empathetic rater who notes “the only time I might forgo [the rating] is if I didn't like it. And then I feel like I should rate them low but I feel bad about it so then I don't. So then I just don't rate them. I'm sure it's not like very helpful for other people but there it is.” She goes on to say this occurs

> usually when they haven't really been five stars but I like them as a person, you know, like they seem like a decent human being and they're doing their job and it just wasn't perfect...I feel bad if I lower their rating, that's what it is. I feel bad if I gave a lower rating, that's going to lower their star rating and maybe negatively affect their life. So that I'm conflicted.

Moreover, these empathetic raters are similar to the forgiving disciplinarian partial spectrum raters as their consideration of the implications of actually deploying a rating less than five stars is the basis upon which they offer these facial protections as Emily shares “the reason why I don't rate is because the lower the score I know it's harder for them to get more rides.” However, an empathetic rater’s actions do not carry the punitive weight that a forgiving disciplinarian’s rating of drivers at the three-star or four-star level does, as again not rating a driver has no effect on their average rating. In this case, empathetic raters fully protect the driver’s digitally quantified face by bypassing the rating act altogether. While empathetic raters are willing to protect the digitally quantified faces of drivers they judge in wrong face during the interaction, they still do not award these drivers as they reserve five-star ratings for drivers they judge to be in face.

By ignoring the rating prompt after a ride with a driver in wrong face in order to offer this protection, empathetic raters also resist Uber’s attempt to discipline its consumers to rate. This presents a tension to Uber's successful exercise of power over its consumers and shows how empathetic raters do not contribute to the legitimation of Uber's digital bureaucracy when they opt out of producing rating data that would penalize a driver's digitally quantified face.

## Binary Raters: Mediated Facial Protection

The final rater type present in the sample is the binary rater, exemplified by Bella who is only willing to rate drivers one star or five stars and requires that the ride is memorable to even rate the driver at all. Presumably there could also exist other subtypes of binary raters such as ones that never opt out of rating drivers or others who may not construct as extreme a binary, but these sentiments were not expressed by respondents in the sample so for now they should remain hypothetical.

Bella, in describing how she rates Uber drivers, notes "if I don't remember specifically...a good thing or a terrible thing, then I will usually just not rate at all." She goes on to say "if I'm going to rate it's going to be for a reason." Drivers who Bella has rated five stars include one driver who "had little primping bags behind both chairs so there was deodorant on the back of both seats and there was other stuff too like a water bottle or like Ibuprofen" and another driver who Bella remembers "stopped the car when we got back to my house and helped carry the [grocery] bags up to my front porch which was extremely kind." These were instances where Bella did not expect the driver to make these provisions, so when they did Bella remembered them as especially good experiences and rated the driver five stars. Under Bella's rating deployment, a memorably good driver is able to effectively carry out a line such that Bella judges them to be in a face of even higher esteem than the one she expects. As Bella mentions, these drivers not only meet all of her expectations, but do the additional work of providing amenities or assistance that Bella does not generally expect from Uber drivers. Bella in turn rewards these memorably good gestures with a five-star rating which increases the driver's overall average. Bella's five-star rating, similar to other riders' default five-star ratings, confirms the driver's line is consistent not only with her definition of the situation but with one of higher driver expectations and thus the driver is rewarded by way of their digitally quantified face.

On the contrary, a driver who violates Bella's primary criteria of safety to a high degree will be memorably bad and judged to be completely in wrong face. As a result, Bella challenges this great inconsistency between the driver's line and her expectations with the largest quantifiable challenge she can by rating the driver one star. In these extreme cases, Bella's response to drivers being in wrong face are similar challenges to those a full spectrum rater would make under the same conditions as they are both willing to give out the worst possible rating score.

Yet, Bella's most common action after an Uber ride is to opt out of the rating, having no impact on the driver's average score. For those drivers who span this middle range of Bella's binary, Bella may find them to be in or in wrong face yet she will not rate in either case. For some drivers, they may meet Bella's expectations and she will judge them to be in face. However, as they have not stood out as memorably good or bad, Bella will not confirm their being in face by way of the rating. Although their digitally quantified faces will not be rewarded as they may under the same conditions by other rater types, Bella will still protect these drivers' digitally quantified faces as she does not make a challenge by not rating lower than five stars. On the other hand, drivers who may not meet one of Bella's secondary criteria may be judged by her to be slightly in wrong face. But assuming the degree of their being in wrong face is small enough such that Bella does not remember taking huge issue with their behavior, they too are protected by Bella not rating them. Bella frames these protections by calling on her knowledge of how Uber employs rider ratings, explaining

> I know that ratings can be really important to Uber drivers and...I know that if you slide below a certain rating you can get deactivated and not be able to access that right to drive anymore through that app. So I don't...I don't just like hand out threes, fours, and twos just for fun because I don't want to be the reason that someone is having their score...their like overall score lowered.

In this case, as was the case with the forgiving disciplinarians and the empathetic raters, Bella's decision to protect a driver's face is mediated by her knowledge that drivers who fall below a certain average rating threshold are deactivated. While she is willing to extend this protection to drivers in wrong face similar to empathetic raters, Bella will still challenge drivers who do not meet her primary criteria as full spectrum raters do. In this way, the rating deployments of binary raters like Bella can be considered a sort of hybrid of full spectrum rater and empathetic rater deployments. However, binary raters have unique enough character to constitute their own type.

The following table serves as a final summary of the Uber rater typology developed throughout these sections. The table recaps each type of raters and its frequency in the sample, the scores given to drivers judged in face, the range of rating scores each type has given to drivers they judge in wrong face, and an application of how each type rates drivers in wrong face to the facework process. I do not claim this to be an exhaustive classification of raters, as further investigations into Uber rider rating practices may discover new types of raters that were not represented in this project's sample.

**Table 1: Uber Rater Typology**

| Rater Type | Driver In Face: Scores Given | Driver In Wrong Face: Scores Given | Facework Process |
|---|---|---|---|
| Full Spectrum Raters (n = 2) | 5 | 1, 2, 3, 4 | Quantified Challenge |
| Partial Spectrum Raters: Holding Subtype (n = 2) | 4*, 5 | 3, 4 | Quantified Challenge |
| Partial Spectrum Raters: Forgiving Disciplinarians (n = 2) | 5 | 3, 4 | Mediated Quantified Challenge |
| Empathetic Raters (n = 2) | 5 | None | Facial Protection |
| Binary Raters (n = 1) | None**, 5 | None, 1 | Mediated Facial Protection and Quantified Challenge |

* Gina rates drivers she judges to be in face four stars.
** Bella does not rate drivers in face who were not memorable

## Discussion and Conclusion

This project set out to answer the questions as to how and why Uber riders quantifiably rate the drivers they ride with. The data collected demonstrates how the subjective meaning made for Uber rides by each individual rider leads to a loose patterning of criteria constructions across multiple riders. Furthermore, we show how riders' rating constitutes contemporary extensions of the facework process outlined by Goffman, developing digitally quantified faces as a way to track confirmations of driver's being in face as well as challenges. In this application we also see how drivers are unable to make offerings in response to challenges, and thus may attempt to proactively intervene in riders' rating process. In regard to how riders do rate, we find most riders rate drivers five stars when nothing goes wrong, confirming the consistency between the driver's line and the rider's criteria and increasing the driver's digitally quantified face. Further, we see a typology of raters develop which helps demonstrate the quantifiability of challenges to a driver's face, the range of challenges each rater type will make under different circumstances, and the facial protections afforded to drivers by some rater types on the basis of these raters' considerations of how low ratings may affect drivers ability to work.

We also show Uber's in-app rating prompt is the preliminary mechanism for getting riders to rate drivers. As such, the rating prompt should be considered a disciplining force that aligns rider behavior with the company's ends of garnering enough rating data to legitimize their management of drivers by way of these numbers. While these rater types may be contrasted in the quantifiable effects they produce for drivers (i.e. forgiving disciplinarians will not rate as low as full spectrum raters for the same offense), it is their simple engagement and participation with the rating apparatus that allows this digital bureaucracy to obtain legitimate character and continue to exist. In light of this analysis, we can see how any rating act works to reify the digitally bureaucratic system in place.

On the contrary, we are also able to observe sites of tension between consumers and the corporation whereby riders protect drivers' faces rather than challenging them. These consumer rationales usually form on the basis of empathizing with drivers and not wanting to lower their average rating score. For binary raters and empathetic raters, this results in no rating being given to drivers who do not meet all of their expectations. This orientation toward inaction works to oppose the digital bureaucratic apparatus in place that reifies itself by way of soliciting consumer data. By not providing a rating, these rater types prevent decisions that rest on these data from being made legitimately. After all, if every consumer opted out of rating Uber drivers, Uber would not be able to justifiably deactivate drivers on the grounds that their average rating score was not high enough as they would not even be able to calculate drivers' averages.

Although disproportionate representation may have been present in this sample as a result of the recruitment methodology employed, these rating processes are still generalizable to other rating contexts. As mentioned earlier, rating prompts are sent to consumers by a number of businesses after the customer has purchased a product. Although they generally take weaker form as product rating prompts are not central to the company's digital interface, we can still apply the disciplinary nature of these prompts to see the underlying value of the legitimacy produced by these quantitative ratings for these businesses as digital bureaucracies. In addition, a host of other neoliberal gig economy businesses, including Lyft and Airbnb, aggregate consumer ratings to quantitatively represent the supposed quality of each service provider working their platforms. Finally, although they do not rely on consumer ratings, quantified evaluations are used to determine who is hired and who is fired by employers across the technology sector and beyond. As Carl briefly describes during the interview, “I'm hiring manager in my job so I'm used to giving ratings on making hiring decisions.” In this capacity, any company that makes human resource decisions on the basis of numerical performance metrics or ratings engages in a similar quantified judgement of how well an individual’s line matches the expectations of another. As such, power is produced in these relations of quantification and differentially serves each party based on who or what is quantifying and who or what is being quantified.

Turning to investigate alternatives to quantified evaluation in the Uber context, if we believe the use of subjective consumer ratings in managing workers is unfair and we are to render the hiring decisions made on this quantified basis unjust, we must educate users about the role they are playing in this bureaucratic machine, the true shape and nature of this machine and its produced effects, and foster a large-scale movement of users rejecting the quantified rating system and opting out of rating. Yet, driver management may not want to be entirely done away with on grounds of the importance of consumer protections. In this instance, requiring a more qualitative form of evaluation and feedback for drivers may be an alternative. This surely would take more labor, both on behalf of the consumer who would be required to spend more time generating qualitative feedback and on behalf of the corporation in assuming the task of processing such data. However, qualitative feedback would allow the driver and corporation to better identify the bases upon which rider judgements were made. For drivers, they would be able to discern why riders rated them in such a way, better informing their corrective process and taking of lines during future Uber rides. Furthermore, under this mode of evaluation, poor rider ratings that were made on discriminatory bases would be visible as such, allowing the company to invalidate these ratings such that they do not disproportionately punish the digitally quantified faces of drivers from marginalized identity groups. Yet, qualitative types of management would not dismantle Uber's digital bureaucracy, but rather would offer a new vision of bureaucracy that privileges qualitative feedback; albeit a vision that would take more work to operate and may ultimately fail in the hypercompetitive neoliberal economic space.

An additional alternative to ratings and reviews may be to discard the subjective constructions of consumer rating criteria in favor of an incident-based metric. As all consumers reported safely getting from their pickup location to their destination as a primary criterion they have for drivers, logging incident reports as they would be generated by police departments when Uber drivers were involved in collisions may be a means of consistently operationalizing this safety criterion. While seemingly more detailed or objective, this strategy would only account for safety as it pertains to crash avoidance, which may overshadow other important factors that contribute to a rider feeling safe. Moreover, I would offer caution that this management tactic may inherit bias present in the incident reports as they are still constructed by potentially biased people and institutions.

A final proposed alternative would be to discard driver evaluation altogether, making bureaucratic tendencies, dependencies, and legitimation impossible. Consumer protections would not be in place, however this would render the condition of consuming such a service precarious just as the condition of providing such a service is, equalizing the power relation between drivers and riders. After all, the conceptions of "good" and "bad" drivers are socially and culturally constructed in the first place and therefore may not be the basis upon which we allow material decisions about which people have access to this kind of work to be made. All in all, this avenue, while more dangerous for all in practice, would allow us to transparently view the underlying foundation upon which Uber and other businesses of this type rest. This in turn could lead to mass disillusionment with the well-disguised, auto-rationalizing, and pervasive digital bureaucracies these neoliberal businesses are founded upon.

I would like to conclude this thesis by offering suggestions and avenues for future research on similar subject matters. Reflecting on the attitudinal fallacy committed in this project, I propose any future qualitative research about individuals' rating tendencies confirms the sentiments expressed by respondents are consistent with a log of each individual's rating history. As this may be difficult to access from the researcher's positionality, cooperation from respondents in finding and sharing this documentation if it exists will be crucial. Additionally, due to the homogeneity of the sample interviewed for this project, I propose similar questions may be asked to larger samples drawn from wider and more diverse networks that better reflect the demographics of the people who use Uber in a particular place or across a range of places. This approach would enable researchers to analyze criteria constructions, rating deployments, and rater types across multiple social identity groups and would allow for the amplification of the voices of those who were not heard in this project. I will also offer that this project approached the

Uber ride as an interaction between one driver and one rider, as was the case in most of the interactions respondents recalled. Yet, the Uber platform allows groups of riders up to four people as well as ride-sharing between unacquainted parties. This presents the potential for a plethora of studies on interaction ritual and decision-making processes in groups and among unfamiliar faces. Finally, this project began with findings that Uber riders rate drivers more commonly than they rate products they buy online. I believe a comparative analysis of how rater tendencies are impacted by the degree to which the person or thing they are rating can be perceived as a human actor would shed valuable light on the interrelation of perception, agency, and reactivity in rating practices. This too, in line with the continued exploration and advancement of this conception of the digitally quantified face, would allow for study of the degree to which people may be objectified or dehumanized through their representation in the digital sphere by way of numbers. This in turn may have value to posthumanist conceptions of technology and the self which were not addressed in this project. All in all, I encourage future sociological research to continue to explore the social construction and use of numbers, the quantified evaluation of people and products, and the overall digitization of life. After all, in our technophilic world that seemingly only accelerates in its pace of advancement, we may find ourselves confronted by dystopian visions of a digital and quantified life sooner than we expect. That is, if we are not already living one.

# Appendix A: Interview Protocol

**Interview Protocol**

**Demographics**
- Gender, Age, Race, Occupation, Highest Education Level
- Do you live in [west coast city]? Why were you taking Uber in [west coast city]?

**Frequency/Type/Use**
- When was the last time you took Uber in [west coast city]?
- How often do you take Uber? What type of Uber do you use (pool, uberx, uberxl, black)?
- Why do you take Uber? What function(s) does Uber have?
- Do you take Uber in other cities? Any noticeable differences? Do you behave differently?

**Driver/Expectations**
- Before getting in the car, what do you expect of an Uber driver?
    - What other expectations do you have of an Uber driver during the ride?
- What qualities/characteristics do you like in an Uber driver?
- What qualities/characteristics do you dislike in an Uber driver?
- What are the primary characteristics you judge an Uber driver based on?
    - Are there any other/secondary characteristics you judge an Uber driver on?
- What characterizes a 5/4/3/2/1 star Uber ride?
- Tell me about the best Uber you had. Do you remember any aftermath (rating/tip)?
- Tell me about he worst Uber you had. Do you remember any aftermath (rating/complain)?
- Before getting in the car, what do you expect of an Uber driver?

**Rating**
- What do you do when an Uber ride comes to an end?
- Do you always rate your Uber driver?
- How long does it take you to rate your Uber driver?
- If you've ever felt conflicted, how did you decide what star level to rate the driver?
- Do you have a default rating?
    - What is it? Why? When do you give the driver a default rating?
- Do you ever rate products you buy? Why/Why not?

**Business/Awareness**
- What do you know/think about how Uber drivers are managed?
- What do you know/think about how Uber drivers are hired?
- What is your impression of Uber as a business?
- Do you think Uber's rating system is an appropriate way to manage workers?
- Is there anything else you want to comment on regarding what we've talked about/Uber more generally?

# Appendix B: Informed Consent Document

**Uber Rating Interview Informed Consent (18+ years old)**

Thank you for agreeing to participate in this interview research about how Uber consumers rate the Uber drivers they interact with. The interview should take about 30-45 minutes, but there is not a strict time limit so I am happy to stay and talk as long as you would like. The purpose of this research is to explore how people who have taken at least one Uber in the Seattle area in the past six months decide how to rate Uber drivers and what they may know about how Uber uses these ratings.

There are no foreseeable risks with this research. However, in talking about experiences using Uber some negative memories may be brought up. Remember that your participation in this research is voluntary. At any point throughout the interview, if you are feeling uncomfortable and would like to change the topic of conversation, take a break, or end the interview, just let me know. Benefits to participating in this research may include pleasure in talking about personal experiences using Uber as well as learning more about Uber.

All interviews will be recorded using a voice recorder. Recordings will be destroyed upon completion of this project. All information will be kept confidential. Any individual identifying information will be removed from the written transcripts of all interviews. Similarly, any individual identifying information will not be included in any analysis or report that is produced from this project.

If you have any questions about this project at any point, please do not hesitate to contact Grant Yeatts, student of Sociology at Whitman College, at:

Phone: (425) 922-8708
Email: yeattsgm@whitman.edu
Mail: 280 Boyer Ave Walla Walla, WA 99362

If you have further questions regarding the academic sponsorship and/or ethical approval of this project, please direct them to any of the institutional contacts listed below:

Matthew Gougherty, Thesis Supervisor (goughemt@whitman.edu)
Whitman College Institutional Review Board (irb@whitman.edu)

By signing below, you are consenting to participate in the interview research described above.

____________________________ ___________________________ __________

Signature Full Printed Name Da

## References


Berman, Elizabeth Popp and Daniel Hirschman. 2018. “The Sociology of Quantification: Where Are We Now?” *Contemporary Sociology* 4 7(3):257-66.

Blumer, Herbert. 1969. *Symbolic Interactionism: Perspective and Method*. Berkeley, CA: University of California Press.

Chan, Ngai Keung. 2019. “The Rating Game: The Discipline of Uber’s User-Generated Ratings.” *Surveillance & Society* 17(1/2):183-90.

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

Chatterjee, Swagato. 2019. “Explaining Customer Ratings and Recommendations by Combining Qualitative and Quantitative User Generated Comments.” *Decision Support Systems* 119:14-22.

Espeland, Wendy. 2016. “Reverse Engineering and Emotional Attachments as Mechanisms Mediating the Effects of Quantification.” *Historical Social Research* 41(2):280-304.

Espeland, Wendy Nelson and Mitchell L. Stevens. 2008. “A Sociology of Quantification.” *European Journal of Sociology* 49(3):401-36.

Fisher, Matthew, George E. Newman, and Ravi Dhar. 2018. “Seeing Stars: How the Binary Bias Distorts the Interpretation of Customer Ratings.” *Journal of Consumer Research* 45(3):471-89.

Foucault, Michel. 1975. "Biopolitics and the Carceral Society." Pp. 322-25 in *Social Theory: The Multicultural, Global, and Classic Readings Sixth Edition,* edited by C. Lemert. Boulder, CO: Westview Press.

Foucault, Michel. 1976. "Power as Knowledge." Pp. 361-65 in *Social Theory: The Multicultural, Global, and Classic Readings Sixth Edition,* edited by C. Lemert. Boulder, CO: Westview Press.

Gandini, Alessandro. 2019. “Labour Process Theory and the Gig Economy.” *Human Relations* 72(6):1039-56.

Goffman, Erving. 1967. *Interaction Ritual: Essays in Face-to-Face Behavior.* New Brunswick, NJ: Transaction Publishers.

Harvey, David. 2005. *A Brief History of Neoliberalism*. Oxford, England: Oxford University Press.

Jerolmack, Colin and Shamus Khan. 2014. "Talk is Cheap: Ethnography and the Attitudinal Fallacy." *Sociological Methods & Research* 4 3(2):178-209.

Lim, Sunghoon and Conrad S. Tucker. 2017 "Mitigating Online Product Rating Biases Through the Discovery of Optimistic, Pessimistic, and Realistic Reviewers." *Journal of Mechanical Design* 139(11):1-27.

Louh, Hsiang-Fei and Sheng-Hshiung Tsaur. 2007. "Gender Stereotypes and Service Quality in Customer-Waitperson Encounters." *Total Quality Management* 18(9):1035-1054.

Mennicken, Andrea and Wendy Nelson Espeland. 2019. "What's New with Numbers? Sociological Approaches to the Study of Quantification." *Annual Review of Sociology* 45:223-45.

Muellerleile, Chris and Susan L. Robertson. 2018. "Digital Weberianism: Bureaucracy, Information, and the Techno-rationality of Neoliberal Capitalism." *Indiana Journal of Global Legal Studies* 25(1):187-216.

Neff, Joyce Magnotto. 1998. "Grounded Theory: A Critical Research Methodology." Pp. 124-35 in *Under Construction*, edited by Christine Farris and Chris M. Anson. L ogan, UT: Utah State University Press.

Oltmann, Shannon M. 2016. "Qualitative Interviews: A Methodological Discussion of the Interviewer and Respondent Contexts." *Forum: Qualitative Social Research* 17(2):1-16.

Pettica-Harris, Amanda, Nadia deGamma, and M. N. Ravishankar. 2020. "Postcapitalist Precarious Work and those in the 'Drivers' Seat: Exploring the Motivations and Lived Experiences of Uber Drivers in Canada." *Organization* 27(1):36-59.

Porter, Theodore M. 1996. *Trust in Numbers*. Princeton, NJ: Princeton University Press.
Rivera, Lauren A. 2010. "Status Distinctions in Interaction: Social Selection and Exclusion at an Elite Nightclub." *Qualitative Sociology* 33(3):229-55.

Rosenblat, Alex, Karen E.C. Levy, Solon Barocas, and Tim Hwang. 2017. "Discriminating Tastes: Uber's Customer Rating as Vehicles for Workplace Discrimination." *Policy & Internet* 9(3):256-79.

Rosenblat, Alex and Luke Stark. 2016. "Algorithmic Labor and Information Asymmetries: A Case Study of Uber's Drivers." *International Journal of Communication* 10:3758-84.

Sauder, Michael and Wendy Nelson Espeland. 2006. "Strength in Numbers? The Advantages of Multiple Rankings." *Indiana Law Journal* 81(1):205-27.

Stark, Luke and Karen Levy. 2018. "The Surveillant Consumer." *Media, Culture & Society* 40(8):1202-20.

Timmermans, Stefan and Iddo Tavory. 2012. "Theory Construction in Qualitative Research: From Grounded Theory to Abductive Analysis." *Sociological Theory* 30(3):167-86.

Weber, Max. 1909-1920. "The Bureaucratic Machine." Pp. 86-90 in *Social Theory: The Multicultural, Global, and Classic Readings Sixth Edition,* edited by C. Lemert. Boulder, CO: Westview Press.

Weber, Max. 1909-1920. "The Types of Legitimate Domination." Pp. 92-4 in *Social Theory: The Multicultural, Global, and Classic Readings Sixth Edition,* edited by C. Lemert. Boulder, CO: Westview Press.

Zwick, Austin. 2018. "Welcome to the Gig Economy: Neoliberal Industrial Relations and the Case of Uber." *GeoJournal* 83(4):679-91.