Robotic and Incompetent: The Role of Quantification in Producing Customer Mistreatment, Including Racism and Sexism in Low-Wage Retail Work

Adam Storer
emlyon business school
storer@em-lyon.com

Thank you to Kira Choi, Bernard Forgues and Gordon Sayre for their valuable feedback on earlier versions of this paper. This paper was funded in part by the ASA DDRIG Award.

Robotic and Incompetent: The Role of Quantification in Producing Customer Mistreatment, Including Racism and Sexism in Low-Wage Retail Work

ABSTRACT

Customer-service organizations increasingly rely on newly available data sources and algorithms to inform managerial practices, potentially altering the frontline service interactions. Drawing on qualitative and quantitative data from two original surveys of employees from 17 foodservice and retail companies, as well as computational text analysis of 2 million Yelp reviews, I provide evidence linking quantification with increased customer-originating mistreatment, including racism, and sexism. Qualitative text analysis shows that quantified work outputs, such as item-scanning speed, can lead employees to feel as though they appear *robotic* in the eyes of customers, whereas the quantification of work inputs, such as adjustments to the supply of employees based on up to the minute sales data can leave employees feeling they appear *incompetent* to customers. In a quantitative analysis, I show that both processes are associated with higher levels of customer mistreatment. However, only organizational processes that lead employees to appear incompetent are associated with higher levels of sexist and racist remarks. In a computational text analysis of Yelp reviews I explain this finding, demonstrating that the appearance of employee incompetence plays into consumer's deep-rooted stereotypes that women and workers of color are less capable than their white and male counterparts. Together these findings demonstrate how frontline workers help algorithmic management succeed by absorbing, in likely unmeasured ways, the negative impacts of operational friction produced by algorithmic decision-making in the form of mistreatment, including racism and sexism.

INTRODUCTION

Organizations increasingly rely on newly available data sources and algorithms to inform managerial practices (Kellogg et al., 2020). Although these new decision making processes can alter the workplace in unexpected ways, for instance by leading employees to reduce effort (Ranganathan & Benson, 2020), employers often ritualize the use of numbers (Mazmanian & Beckman, 2018) and rely on them without fully understanding what they mean (Anthony, 2021) or their potential unintended consequences. In the context of customer-oriented platform work, algorithmic reliance on service ratings can give customers outsized power, leading platform workers to search for new ways to relate to and placate customers and algorithms (Bellesia et al., 2023; Cameron, 2022; Rahman, 2021). This outsized power can lead to abuse from customers using platforms (Maffie, 2020), a process that occurs in traditional customer service jobs where employees are taught that the customer is always right (Korczynski & Evans, 2013; Korczynski & Ott, 2004).

In traditional service work, quantification and the algorithmically-informed managerial practices that new sources of data make possible may have additional unintended consequences. In this setting, employees are traditionally given standardized scripts and ways of doing their work (Ikeler, 2016; Leidner, 1993) while at the same time being asked to provide cheerful and individualized customer experiences (Hochschild, 1983; Korczynski & Evans, 2013). As a result, even before the shift to a focus on algorithms and quantification, employees were often torn between meeting the needs of the employer and those of the customer (Korczynski & Ott, 2004). This dynamic is likely amplified by quantification, what processes are and are not quantified, and the extent to which managers use algorithms based on these numbers to drive decisions about how work is done. This paper considers the context of foodservice and retail work to understand how the quantification of work outputs, such as the number of items scanned by a cashier, and the quantification of work inputs, such as the ratio of staffing levels to demand at half-hour intervals based on cash register data, can produce an environment that puts numbers above service quality, leading to higher rates customer mistreatment, including racism and sexism.

For service sector workers, mistreatment by customers, which can range from low-grade incivility with ambiguous intent to harm, to outright abuse, is known to present serious organizational problems, including customer sabotage, lower quality customer service and worse task performance. While a large literature has inspected the dynamics of mistreatment (Gabriel & Diefendorff, 2015) as well as mistreatment's consequences (Groth et al., 2019), surprisingly little research has sought to identify mistreatment's antecedents. To the extent that the literature does examine this question, it focuses either on power differentials between customers and service workers (Korczynski & Evans, 2013; Maffie, 2020) or individual-level explanations, such as employee's service orientation and agreeableness or the customer's attitude (Sliter & Jones, 2016). This paper instead considers the extent to which organizational decision-making based on up-to-the-minute data creates a context where employees are more likely to experience customer mistreatment, including racism and sexism.

To measure this relationship, I draw upon two original surveys of frontline employees from 17 foodservice and retail stores conducted in July 2020 (N=546) and July through August of 2022 (N=1,271) and a sample of customer reviews from Yelp (N=2 million) over three studies. In study one, I conduct a qualitative analysis of free-text descriptions of how employees interpret the relationship between the metrics they are measured on and customer service. In study two, I measure the association between quantified managerial practices and employee's experiences of customer-originating mistreatment, racism and sexism. The third and final study draws upon a sample of 2 million Yelp reviews to consider the customer perspective, using computational text analysis to determine the extent to which frictions in customer service may play into consumers' racial/ethnic and gender bias.

Results from the first study indicate that workers are acutely aware of management's focus on these metrics, and frequently discuss how quantification of work inputs and outputs are misaligned with the high-quality customer service they are asked to deliver. In terms of work output, respondents describe a process in which an emphasis on speed or data collection over service quality leads employees themselves to act *robotically* in their interactions with customers. In terms of work inputs, respondents note that processes such as adjustments to staffing levels at half-hour increments based on algorithmic predictions of customer demand often leave them under-resourced. Respondents feel customers then blame operational issues, such as long lines, on *incompetent* employees.

In study two, I consider the association between the operational decisions that leave workers feeling robotic and incompetent, and employee's experiences of customer mistreatment, including racism and sexism. Findings indicate that while operational decisions contributing to appearances of roboticism and incompetence are both significantly associated with exposure to customer mistreatment, only an appearance of incompetence is associated with customer-originating sexism and racism. I suggest that the appearance of sexist and racist remarks surrounding incompetency is due in part to negative racial/ethnic and gender stereotypes surrounding competency. In study three I inspect this negative competency stereotype among North American consumers using a computational text analysis of Yelp reviews. Findings indicate that concepts such as competency and efficiency are coded in text as more white and masculine than their opposing concept pairs of incompetency and inefficiency. These results provide additional evidence that when organizational policies lead employees to appear incompetent, their actions will be perceived by consumers in a gendered and racialized way.

These findings contribute to broader discussions surrounding quantification and discrimination while at the same time highlighting a likely undermeasured outcome in the research on the effect of the quantification – workplace mistreatment. In effect, frontline workers in this sector absorb the negative impacts of algorithmic management in unmeasured ways, making managerial practices seem as though they are working in the short run, but producing potentially long-lasting effects on the firm as well as its employees. Employers in this sector should be especially aware that decisions regarding which workplace activities to measure, how to produce those measurements, and the decision-making processes based on those metrics have direct effects on frontline workers' experiences of mistreatment, including racist and sexist interactions.

THEORY

Customer Mistreatment and Its Antecedents

Customer mistreatment has received increasing attention over the last few decades, and the negative effects of customer mistreatment have been well documented (Groth et al., 2019; Schilpzand et al., 2016). For example, when workers are mistreated, they may engage in counter-productive activities at work, such as sabotaging customers (Wang et al., 2011). In addition, workers who experience incivility are more likely to engage in emotional labor (Gabriel & Diefendorff, 2015), meaning that they must actively alter their true or projected emotional state in order to maintain a positive customer service interaction. As workers become disengaged, they are more likely to engage in a more transparent form of emotional labor, leading to worse customer satisfaction (Groth et al., 2009). The effort put into dealing with a negative interaction may also cause workers to lose focus on the task at hand, impacting productivity (Porath & Erez, 2007). Moreover, when workers experience incivility it not only impacts the current interaction, but may impact worker well-being and quality of service in the days after a notable incident (Groth & Grandey, 2012).

While customer mistreatment is typically thought of as separate from racism and sexism in the workplace, mistreatment can present a key source of racism and sexism in the workplace (Cortina, 2008). Although it has received less attention in the management literature, related sociological literature has identified customer service interactions as a key source of racism and sexism. For instance, women must cope with sexual harassment from customers (Good &

Cooper, 2016) and transgender workers confront transphobia and abuse among customers (Hadjisolomou, 2021). Race similarly plays an important role in customer service and emotional labor (Mirchandani, 2003).

Together, the literature shows that the reduction of customer mistreatment can lead to fewer experiences of racism and sexism in the workplace, less draining emotional labor for employees, and higher productivity for the organization. Yet relatively few studies have identified organizational pathways for reducing customer mistreatment. One common macro-level explanation for customer mistreatment stems from Korczynski & Ott's (2004) study of customer sovereignty. The authors argue that the notion in many service establishments that the customer is always right puts frontline service workers in a subservient position. This leads customers to enact their power over frontline workers in various ways, including through customer mistreatment (Korczynski & Evans, 2013). Maffie (2020) shows that the same dynamic occurs in platform work, where customer's ability to rate platform workers leaves them with more power than workers.

A second line of research has focused on a set of explanations stemming from how customers perceive the service climate. Combining store busyness, cleanliness, quality of layout, and workplace attractiveness into a single measure of service climate, Sliter & Jones (2016) find that a worse service climate is associated with more incivility. They also find limited qualitative evidence, but not quantitative support, for the impact of training on incivility. While this study provides an important first step in drawing a connection between specific operational decisions and incivility, the construct itself is particularly diffuse, drawing on a snowball sample of a broad range of jobs, while combining a multitude of complex properties into a single metric.

This study builds on this research by offering two additional explanations for how operational decision-making informed by processes of quantification can lead to service workers experiencing mistreatment as a result of being perceived by customers as either a) robotic, or b) incompetent.

Quantification of Work Outputs and Roboticism

Quantification of work processes themselves have been of interest to both managers and management scholars since Taylor (2004) introduced the theory of scientific management. As technology develops, and publicly traded companies are heavily scrutinized by investors, companies have increasingly relied on quantification to promote efficiency, transparency, and accountability (Kellogg et al., 2020). This produces a tension between the numbers organizations rely on to make decisions and the complex situations the numbers attempt to represent. Over time and through organizational practices, the numbers can take on a meaning of their own (Mazmanian & Beckman, 2018). As the numbers become more complex, analysts themselves may have trouble making sense of what the numbers truly mean (Anthony, 2021) yet may continue to rely on them in order to justify their role in the organization (Stice-Lusvardi et al., 2023). As these outputs are measured, they may become increasingly performative (MacKenzie & Millo, 2003), with activities increasingly driven by, rather than measured by, the numbers. This may become counterproductive, since when workers feel the outputs they are measured on do not accurately reflect the complexity of their work, they may reduce effort (Ranganathan & Benson, 2020). Customer ratings of the quality of one's work have received attention in platform work, where autonomous platform workers alter how they work in order to improve their numbers (Cameron, 2022). There is some evidence that, when customers are given more power through a rating system, they are more likely to take advantage of this power and abuse platform workers (Maffie, 2020).

This is an important tension for front-line foodservice and retail, which as opposed to the autonomous nature of platform work, presents itself as a key site of routinization. Employers often ask employees to stick to specific scripts, and to produce experiences that are essentially indistinguishable from one another regardless of the specific establishment (Leidner, 1993). Over time, low-level frontline work has become increasingly routinized and deskilled (Ikeler, 2016). From this perspective, the institutionalization of new measurements, coupled with deskilled and routinized interactions, may create operational friction due for an instance to a focus on speed of service rather than fully addressing customer concerns, and reduce the quality of the interpersonal interactions between employees and customers.

This problem is compounded by the fact that in many retail settings, employees are typically exposed to a form of bureaucratic management with multiple rungs of low-level managers who have little decision-making latitude and often are distinguished by making only a few extra dollars per hour (Bolton & Houlihan, 2010; Hadjisolomou & Simone, 2021). These middle managers, who play a vital role in providing feedback on how new programs, such as the collection and implementation of new quantitative measurements, are implemented (Chown, 2021), may be helpless in the context of retail and foodservice.

This paper hypothesizes that when quantification and routinization are misaligned with quality customer service, it may create a scenario in which employees are forced to act in a way that appears *robotic* to customers, for instance due to their seeming unwillingness to help with a customer's specific issue in favor of meeting their numbers. Since customers are conditioned to expect personalized service (Korczynski & Ott, 2004), these robotic interactions may create an operational friction in which customers are unhappy and lash out.

Hypothesis 1: When work output metrics are not aligned with customer service, employees will experience greater levels of mistreatment due to their apparent roboticism.

Quantification of Work Inputs and Incompetence

In addition to measurements of work output, decisions regarding how to effectively measure the need for work inputs, in terms of staff, turnover, and reliance on new employees, may also impact the customer service interaction, but likely in a different manner.

Staffing Levels. A key operational decision, often informed by algorithmic approaches, is staffing levels. Many customer service industries operating on relatively small margins rely on complex algorithms in order to dynamically match the labor supply to demand (e.g. Atlason et al., 2008; Cezik & L'Ecuyer, 2008). In order to produce these models, businesses must quantify not only their labor costs, but also changes in the quality of the service that is being provided. In the case of call-centers, these models quantify the acceptability of the service being provided by including, for instance, the number of calls that can be marked as resolved (Ren & Zhou, 2008). Flexible staffing models are particularly common in food-service and retail. Lambert (2008) describes how employers engage in flexible staffing practice, such as leaving workers on-call or making last-minute scheduling changes based on demand within a store. For instance, employers may decide to send a worker home early if sales over the last hour do not meet with complex models predicting the sales to staff ratio.

Yet some studies have identified that these algorithmic decisions may miss clear factors. In a study of clothing retail, for example, Misra & Walters (2022) describe employees worrying about children of shoppers running in and out of the store, since each store entrance is recorded automatically and included in higher level quantified decision making processes concerning how well the store is doing. Butler & Hammer (2019) similarly identify fast-food managers frustrated over sales metrics that they feel are not indicative of the quality of the work they are doing, panicking over how snow or hot weather will impact measures of their store's performance. Moreover, these managers also discuss how corporate-mandated staffing decisions lead to no-shows and turnover. At the same time, there is not a staffing buffer against these issues, leading to more stress and lower job quality.

The implementation of these models can have unintended consequences for staffing and effort. In the nursing sector, research has shown that staffing at what may seem to be at optimal levels according to some metrics can produce absenteeism (Green et al., 2013) and worse patient care (Berry Jaeker & Tucker, 2017). In foodservice, lower staffing levels can lead workers to shift their activities, for instance by reducing their sales efforts in order to cope with a larger workload (Tan & Netessine, 2014). In call centers, employment systems that lead to higher quits and dismissals have been shown to reduce the quality of customer service (Batt & Colvin, 2011). On the other hand, research shows that when employers engage in more responsible scheduling practices, they may benefit through improving sales and boosting productivity (Kamalahmadi et al., 2021; Kesavan et al., 2022).

The effects of these approaches to minimizing staffing may also have unintended and often unmeasured impacts on employees themselves. Involuntary schedule instability can impact employee well-being in the form of worker health and job satisfaction as well as their material hardship (Schneider & Harknett, 2019a, 2021). In addition, Storer (2022) argues that frontline workers have higher turnover intentions when they attribute negative interactions with customers to organizational decisions such as understaffing.

I argue that these staffing decisions informed by a quantitative focus on closely matching customer demand and staffing levels can lead to operational friction, creating a scenario in which too few employees are asked to serve too many customers. These decisions may lead customers to see long-lines and unfinished tasks, not as the fault of organizational decisions but as the fault employees who are perceived as *incompetent*.

Hypothesis 2: As staffing levels decrease, employees are more likely to experience customer mistreatment due to the employee's apparent incompetence.

Training. A related operational decision impacted by algorithmic approaches to staffing that may impact customer service is the level of training and experience among employees. Among national retailers, organizations often expect high levels of turnover, even searching for optimal levels of turnover (Siebert & Zubanov, 2009). One aspect of the negative impact on turnover is not only understaffing, but the loss of human capital as new workers enter. De Stefano et al. (2019) highlight that turnover effects are heightened when workers are replaced with less-experience replacements. In a study of workers at Burger King, Kacmar et al. (2006) find that high levels of turnover create higher wait times for customers, impacting sales and profits for the corporation.

In companies that rely on low levels of staffing and expect high levels of turnover, when workers leave they are likely to be replaced by less-experienced counterparts. As Butler &

Hammer (2019) describe, food-service and retail organizations that operate at a minimum level of acceptable staffing are not likely to pay for additional workers to be on staff to provide additional training. As a result, the extent to which adequate training is provided to workers likely impacts the presence of operational errors. Algorithmic decision making that leads to turnover, then may also create a workforce of poorly-trained workers who are more likely to move slowly and make mistakes, leading to customer perceptions of incompetence.

Hypothesis 3: When new hires are not adequately trained, they are more likely to experience mistreatment due to their apparent incompetence.

Incompetence and Racial/Ethnic and Gender-Based Discrimination

While the hypotheses above focus on generalized mistreatment, there is ample reason to believe that frontline workers experience multiple racist and sexist remarks from customers (Billingsley, 2016; Hadjisolomou, 2021; Kern & Grandey, 2009). While it might be reasonable to suggest that racist and sexist remarks simply originate from serving racist and sexist customers, organizational decision-making may also contribute to the production of an environment in which these remarks are more common. For this reason, I consider a body of research on negative competence stereotypes, suggesting that incompetence is closely tied to negative racial/ethnic and gender stereotypes.

Research shows that men and women are often held to different standards, and men are often perceived as more competent than women even when they perform equally as well as their male counterparts (Foschi, 2000). In academia women and women of color specifically often find themselves to be presumed incompetent by students and colleagues (Muhs et al., 2012). In addition, Carton & Rosette (2011) show that this negative competence stereotype impedes advancement for Black leaders. As a result, in a recent *Academy of Management Annals* article Phillips et al. (2022) argue that organizational scholars need to be aware of the fact that women and racial/ethnic minorities face negative competence stereotypes in the workplace.

While research has focused on the effects of the negative incompetence bias from those within the organization, it is just as likely that these negative competence stereotypes could impact service work. Moreover, organizational decision making may play a role in activating these stereotypes among customers. In the context of inadequate training and understaffing, customers are likely to mistake outcomes such as longer lines or more common mistakes from employees not as the result of an organization's staffing policies, for instance, but as the result of individual worker's competence. Service sector employees who are not given enough training often struggle with feelings of incompetence (Sallaz, 2015) and frustration that customers do not understand that organizational failures are not the fault of the employees (Storer, 2022). By producing an environment in which workers may appear to be incompetent to outsiders, organizations may be triggering negative competence stereotypes that are especially harmful to women and employees of color. These employees, as a result, may end up receiving more comments that are not only uncivil, but that involve racial/ethnic slurs or sexist language.

Hypothesis 4a: Employees who experience organizational decisions that lead to an appearance of incompetence will experience greater levels of racial/ethnic and gender-based discrimination.

It could also be the case that roboticism may subvert stereotypes, specifically for women, in ways that are harmful. Given that women are often found in care work (Dwyer, 2013) and emotional labor is a gendered activity (Hoschild, 1983), customers may expect women, more than men, to be especially kind, caring, and attentive to their needs. If quantification of work outputs forces employees to put speed and efficiency over customer care, female employees may fail to meet the expectations put on them by customers. As a result, the following hypothesis may hold:

Hypothesis 4b: Employees who experience organizational decisions that lead to an appearance of roboticism will experience greater levels of gender-based discrimination.

DATA & METHODS

To test these hypotheses, I rely on data from two original surveys of frontline retail workers at 17 national chains that were fielded in July of 2020 and July of August of 2022, as well as a random sample of 2 million Yelp Reviews. The surveys targeted large-scale grocery chains (Albertson's, Kroeger, Safeway) retailers (The Home Depot, Lowe's, JOAANN, Michael's, Hobby Lobby), super centers (Target, Walmart) warehouse clubs (Costco, Sam's Club), Dollar Stores (Dollar Tree, Dollar General), pharmacy retail (Walgreens, CVS) and one Coffee Shop (Starbucks). Targeting was done through Facebook Advertisements, a system that allows for targeting of individuals who publicly post on their Facebook profile that they work at one of the companies listed above. This method has been validated (Schneider & Harknett, 2019b) and used to study this specific population (Schneider & Harknett, 2019a; Storer et al., 2020).

The Facebook Advertisement sampling technique is particularly attractive because it allows for an employer-identified sample without partnering with organizations. Workplace surveys which focus only on a single company can produce bias at the individual level due to worker's fears of the surveys being monitored by their employer. They may also produce bias at the organizational level if the companies that are willing to work with researchers are significantly different from those that are not. In addition, many multi-company studies rely on data from publicly available convenience samples, such as written reviews provided on websites like Glassdoor.com. This strategy allows for a sampling of workers at specific companies which provides the opportunity to control for unobserved differences in management strategies between employers, with a known data-generating mechanism and an ability to collect additional individual level controls. Appendix A includes a detailed description of the surveys, including response rates, comparisons to other survey techniques, techniques to address item non-response, and a consideration of the costs and benefits of using individual-level data to measure employment practices, as well as strategies to address halo effects.

The paper also draws on the Yelp Academic Dataset to uncover underlying customer biases in reviews of organizations. The full dataset includes nearly 7 million reviews that were algorithmically selected by Yelp as "Recommended Reviews," a process that takes into account the quality and reliability of reviews on the platform. The full sample includes over 150,000 organizations and covers 11 metropolitan areas in North America. From this sample, I randomly sample 2 million reviews in order to produce word embeddings to understand how opposing word pairs such as "competent" and "incompetent" or "efficient" and "inefficient" relate to one another in semantic space.

STUDY ONE: QUALITATIVE ANALYSIS

In Study One I provide qualitative evidence from free-text response survey questions to understand the way that frontline workers understand managerial practices as driven by metrics. Employees provide a key linkage point between metrics and key decisions regarding who is working when, and how that work is completed. At the same time, employees have a unique perspective due to their ability to not only understand extent to which decision making is quantified, but the impact of the numbers on how work is performed and interpreted by customers.

I draw two free-text response questions. The first, from the 2022 survey asks employees to describe "What kinds of metrics or numbers does management focus on?" I also draw on a 2020 survey asking "Share some of your most/least favorite things about the customer service experience at [employer.]" The qualitative evidence is meant to provide evidence of validity of the theoretical model and deepen the reader's understanding of how employee's see these linkages occurring within the context of foodservice and retail jobs. This analysis is not meant to provide an exhaustive list of metrics mentioned or aspects of customer service that appear in the dataset.

Results

Functional Metrics. In response to the survey question regarding metrics, many workers discuss a variety of customer service metrics that seemed to be at odds with high quality customer service. For instance, some workers report being measured based on cash register scanning speeds, at the expense of customer service. One respondent writes, "They want people in and out, had customers ask me a question they didn't want me to take time to answer said I didn't have time.” Cashiers are asked to focus heavily on scan times, with many noting the number of items per minute scanned as a key metric. One cashier mentions that “now and again they remind us that we're slow, and post our employee number next to our rating.” Beyond the register, sales associates also noted a conflict between measurements surrounding the speed of their work and customer service. One respondent identifies multiple metrics:

> *“time it takes to get product unloaded from boxes BY SALES FLOOR people and how fa[s]t the fulfillment team can pick merch off the floor. Many of the workers do not help guests in fear of not get[ting] things done on time.”*

Another common complaint was the impact of corporate-directed membership and rewards plans altering the activities of cashiers. One respondent notes that, for them, the main imperative is “signing up customers for the reward program above all else. Stores are measured weekly to make a certain number o[r] cashiers can be fired.” Workers asked to follow scripts to at the register describe how customers express frustration over “being badgered” about upgrading their membership. Beyond memberships, other stores focus on the collection of emails, and the sale of company credit cards, a process that can create an ethical dilemma for employees. One respondent notes that, although they were measured on the number of credit cards sold, they found:

> *“this meant having to consciously counteract my own reflexive empathy and desire to help the customer with [w]hat they actually wanted and/or needed -- which corporate -claims- should be our priority -- and push something at them that virtually no customer wanted or needed.”*

Although this topic is well covered in the question regarding metrics, this topic also arises in ways that match the key themes in the customer service question. For example, one respondent makes the connection explicitly that the “[c]ompany…prioritizes upselling and being fast more than connecting with customers at times.”

Staffing and Precarious Schedules. In response to the question about metrics, many workers claimed that the most important metric their employers seemed to care about was the ratio of sales to the number of staff on the floor. Many respondents noted how the company kept track of, for example, "how many customers I serve in 30-minute time periods and how many people are on shift at a time," noting that this causes them difficulty. For instance, some describe a process of "under scheduling to meet weekly budget, profit over last year" and "high numbers regardless of situations it causes employees." A clear theme in these responses notes how understaffing makes the job seemingly impossible, with customer service suffering. One respondent notes:

> *"They want all ready-to-go foods, inline warmers, grab & go bins filled at all times, but don't provide enough staff hours to complete the task. Customer service also suffers due to constant understaffing."*

Another respondent makes a link between understaffing, metrics, and customer service:

> *"We have surveys we're meant to have filled out by customers. If I get anything but the highest scores, it's a negative survey. Most customers have a problem with the line length and how the prices have changed but not been updated on the stickers, as well as shelves that are out of stock. One person cannot ring, stock, and change stickers at the same time due to how busy our store is."*

In this example, long lines and worse customer service may lead workers to seem incompetent, unable to keep the store functioning properly. This leads customers to give lower ratings, a metric that leads not to systemic change, but to punishment.

These themes also appear in the survey regarding the Pros and Cons of customer service, where employees make an explicit connection between understaffing and negative interactions with customers. One employee notes "[s]ome yell or cuss us out because of our insane workload [we] are expected to do and they don't understand." Many note this indirectly, putting the blame on impatient customers without realizing that this context is produced through managerial strategies:

> *"when customers are angry or i[m]patient it makes things a bit more stressful and makes me feel like I'm not doing enough to help them and get them what they need. [I] don't want them to think I'm a bad worker."*

Together these pieces of evidence show that employees see customers as coming in with a set of expectations that, due to understaffing, are not possible to meet. As a result, employees are left feeling unfairly evaluated and harshly treated by customers who don't understand their workload.

Training. The metrics question did not surface responses based on training, likely because the link between staffing, turnover, and training does not neatly fit into a single metric. However, the study of the Pros and Cons of customer service show that employees see this as a key issue. One employee notes:

> *"Customers expect me to have product knowledge, which is something I've never been trained on, and I was also never trained on the location of merchandise which is the main reason I speak with customers."*

Others echo this feeling of not having enough training to help customers, noting that it's difficult "[n]ot knowing what [customers] need or location of the item needed." This is especially apparent in stores that serve specific needs, such as home improvement stores. One respondent notes:

> *"[customers] come expecting expert help on how to plumb or wire [their] whole house. they cannot pay those professionals to work in a store. [T]he wage doesnt compair[sic]. [W]e get a few retired folks from those fields that need something to do and they kind of teach the rest of u*s."

One worker notes that the combination of understaffing and undertraining can be particularly embarrassing, saying customers “wait to[o] long when I need an answer for something for the customer. It is so embarrassing to stand there and wait so long for help.” At times, this lack of training is perceived by employees as a lack of experience. For instance, one respondent notes: “I’m not as experienced as everyone else quite yet so I’m not always able to help to the fullest ability.” In frontline customer service jobs, training and experience seem to play an important role in shaping the extent to which customers see employees as incompetent, and also how employees see themselves.

Together, the qualitative evidence illustrates clear conceptual linkages between metrics, staffing and training on the one hand, and how work is conducted and the quality of customer interactions on the other. Specifically, descriptions of misaligned metrics, understaffing, and undertraining often reflect a friction in the customer service experience that is sometimes linked explicitly to customer mistreatment. Moreover, this evidence provides support from employees that measurements of work output, such as the number of individuals who sign up for a rewards program, lead employees to seem robotic in eyes of customers. On the other hand, understaffing and undertraining are often discussed in the context of a customer’s failure to understand the heavy workload, or feeling embarrassed about not having enough information to help. Instead, employees believe that customers perceive them as incompetent, a feeling that can employees also internalize.

STUDY TWO: QUANTITATIVE ANALYSIS OF EMPLOYEE SURVEYS

Results from the study one indicate that there is a linkage between algorithmically-informed decision making, both in terms of how metrics are measured and implemented, as well as their impact on staffing and training, and the quality of customer service from the vantage point of frontline employees. In study two, I use regression analysis of the 2022 survey (N=1,271) to measure the association between employee's experiences of managerial practices informed by quantification and experiences of mistreatment, racism, and sexism from customers.

Independent Variables

I include three independent variables in the analysis.

Misaligned Metrics. This question asks, "At my [Store], metrics and numbers are at odds with **good customer service.**"

Understaffing. This question asks, "In general, there are **not enough** workers on the floor at my [Store]."

Inadequate Training. This question asks, "At my [Store], new hires are given **adequate training.**" This question is reversed in order to match the direction of the other dependent variables.

Dependent Variables

I include three measures of mistreatment. Each mistreatment measure asks individuals to specify how frequently different types of customer interactions occur. Individuals can respond on a five-point scale which includes the response categories "Never," "Once or Twice," "Monthly" "Weekly" or "Daily." *Customers Yell* asks how often customers raise their voice at the respondent. *Customers Sexually Harass* asks how often customers sexually harass the respondent or others. *Customers Racist* asks how often customers use racial slurs towards the respondent or others.

Controls

I also include a series of controls in the analysis.

Individual-Level Controls. At the individual-level, I control for whether the respondent is *White,* a *Cis-Gender Male*, their *Age*, and whether they speak *English as a Second Language. I* also control for the amount of *Education* the respondent has, whether they are currently *Enrolled* in school, whether they have *Children,* and whether they are *Cohabitating* with a partner.

Job-Level Controls. At the job-level, I control for *Frequency of Customer Interactions*. I also control for *Tenure* at the company as well as *Managerial Level*. In addition, I control for *Hourly Wage*, their *Usual Hours* on the job, the extent to which their *Hours Vary* from week to week, whether they are a *Full-Time* worker, and whether they are a *Current Worker.*

Halo-Effect Controls. One key concern is that responses may include a halo-effect, where satisfied workers are likely to under-report negative events and over-report positive events. To account for this possibility, I take a conservative approach, controlling for multiple factors that may also be impacted by the independent variables. I control for *Job Satisfaction Plans to Look for a New Job, Self-Reported Effort* on the job*,* and the *Self-Report Meaning* that frontline workers take from their work beyond their wages. I also control for negative views of the organization that may impact reports of company strategies such as training or understaffing.

Establishment-Level Controls. In addition to individual-level controls, I control for a series of establishment-level controls. Since the rates of mistreatment may differ by the race and gender of both coworkers and customers, I include self-reports of the following controls: *Coworker White,* *Coworker Cisgender-Male, Customer White, Customer Cisgender-Male. I* also control for the likelihood that understaffing and a focus on non-functional metrics may be the result of working in a poorer location, where the company may be struggling. To control for this, I include controls for *Customer Class Background* as well as whether the *Company is Doing Well* financially.

Company-Level Controls. To control for differences in managerial strategies between companies, I also include *Company-Level Fixed Effects*. This is particularly important since theories of customer sovereignty suggest that the extent to which a company, as policy, puts customer power over coworkers plays an important role in producing customer mistreatment.

Descriptive Statistics

The following table demonstrates the means for each of the variables included in the analysis.

Table 1 About Here

Table 1 shows that the sample is split nearly evenly between part-time and full-time employees, and includes roughly two thirds frontline employees, and one third low level managers or department heads. Respondents also interact with customers extremely frequently. From a demographic perspective, the sample leans White and non-cis-gendered male, which is typical of Facebook samples of these groups. In addition, about one fifth of respondents are enrolled in school, and about half are cohabitating and have children.

In addition, Figure 1 below demonstrates the frequency of responses in each of the five categories for the dependent variables.

Figure 1 About Here

Figure 1 shows that only roughly 10% of respondents never experiencing mistreatment, while roughly 30% experience yelling on a regular basis. However, the remaining 60% are distributed evenly among each of the next three categories, with 40% experiencing mistreatment on a weekly or daily basis. Racist and sexist remarks, on the other hand, are much less common, with over 50% of respondents reporting that they never experience them, and another 30% noting that these comments have occurred once or twice. However, a significant number of respondents report customer racism or sexism occurring on a somewhat regular basis, with slightly higher reports of customer sexism.

Figure 2 About Here

Figure 2 shows the distribution of responses for the independent variables, specifying perceptions of operational issues in terms of understaffing, inadequate training, and misaligned metrics. A large majority (over 75%) of employees agree or strongly agree that their store is understaffed. Employees are more split over the quality of training, with around 35% expressing favorable opinions about training quality, and 35% expressing unfavorable opinions.

Analytic Strategy

Following advice from Angrist & Pischke (2009) I present results from standard Ordinary Least Squares regressions, rather than Ordinal regressions, since the results are typically not more accurate but are more easily interpretable. However, in Appendix B I also include ordinal regressions, which present the same pattern of results. In addition, since the sampling strategy of the survey was based on targeting specific companies, I cluster standard errors based on company. In each regression I present I include all controls listed above.

Results

Table 2 below demonstrates a step-wise model with increasing sets of controls for the dependent variable of customer mistreatment, measured as how often customers raise their voice at employees. For reference, each independent and dependent variable is measured on a 5-point scale.

Table 2 About Here

Results show that, without controls, there is a significant association between each of the three independent variables and customer mistreatment. A one step increase in each of misaligned metrics, inadequate training, and understaffing is associated with between a 13% and 15% increase in the likelihood of experiencing customer mistreatment on a more regular basis. Compared to the most favorable condition, where individuals believe their store is well staffed, well trained, and has functioning metrics, a combined 4-step movement to the most negative reports for each of these strategies combined is associated with about a full two-step increase in experiences of mistreatment. This would move employees from the baseline of a response of 1.65 on a 5-point scale to 3.65, associated with a movement from experiences of mistreatment occurring less than once a month, to these experiences occurring on a monthly or weekly basis.

As controls are included, this effect size remains roughly the same in the first three models, which iteratively add individual-level controls, company fixed-effects, and self-reported information about customer and coworker demographic composition, as well as perception of the establishment's financial well-being. The effects, though significantly attenuated, continue to be significantly when including the halo-effect controls of *Job Satisfaction, Plans to Search for a New Job, Self-Reported Effort,* and *Meaning From Work.* It is important to note that these controls present a strong test of the hypotheses, since the independent variables likely also influence factors such as job satisfaction. There is particularly significant attenuation when adding halo-effect controls for the inadequate training variable. This should not be particularly surprising since, in the qualitative evidence, there seemed to be a strong link between the ability to effectively demonstrate and use training and job satisfaction or feelings of self-worth. In the final model, a one-step increase in any of the three independent variables is associated with a .10 to a .12 point increase in mistreatment. For each independent variable, the average marginal effect at means indicates a one step increase is associated with between a 3% and 4% increase in experiences of mistreatment compared to the control). Together, this provides significant evidence in favor of *Hypotheses 1, 2 and 3,* which hypothesized that misaligned metrics, understaffing, and inadequate training would be associated with customer mistreatment even after controlling for halo effects.

Figure 3 presents the average marginal effects of the joint movement from the best to the worst conditions along each of these three independent variables. The joint stepwise condition is provided since each of the independent variables have roughly the same effect size.

Figure 3 About Here

As Figure 3 shows, in the best-case conditions, employees would expect to experience mistreatment rarely while on the job, option two of the 5-point scale. In a completely neutral condition, employees on average would experience mistreatment only a monthly basis. In the most extreme scenario, employees fall at the half-way mark between monthly and weekly experiences of customer mistreatment.

Results from Table 2 have shown that mistreatment is associated with each of the three independent variables. The next set of analyses, presented in Table 3 below, presents results for sexist and racist remarks, in addition to customers simply raising their voices at employees. Given that the halo effect controls present a strong test of the association between the independent variables and these different forms of mistreatment, for each dependent variable results are shown with and without halo effects. For reference, the first two columns in Table 3 reproduce Models 4 and 5 of Table 2.

Table 3 About Here

Results from the analysis show significant effects for the association between inadequate training and understaffing and customer's making sexist and racist comments. A one-step increase in reports of inadequate training are associated with a .15 point increase in customer sexism in the models where halo effects are not controlled for, and .09 points in the halo-effects specification on a 5-point scale. Results also indicate significant, but smaller results for understaffing. A one step increase in understaffing is associated with a .08 point increase in customer's sexist remarks without controlling for halo effects, and a .07 point increase when controlling for the halo effects. A similar pattern holds for customers making racist remarks. A one-step increase in perceptions of inadequate training is associated with a .13 point increase without halo effect controls, and a .08 point increase in the control condition. The effects of understaffing are significant yet smaller, with a one point increase in understaffing associated with a .07 point increase in racist remarks without halo effects, and a .05 point increase when halo effects are controlled for.

While inadequate training and understaffing continue to predict sexist and racist remarks, misaligned metrics do not have the same effect. Misaligned metrics are not associated with sexist or racist remarks in either specification. Together, these results provide support for *Hypothesis 4a*, that understaffing and inadequate training lead to higher levels of sexist and racist remarks due to the connection between an appearance of incompetence and the negative competence stereotype. However, I do not find support for *Hyothesis 4b*, that the appearance of roboticism subverting expectations of high-quality emotional labor will lead to higher levels of sexist remarks.

STUDY THREE: COMPUTATIONAL TEXT ANALYSIS OF YELP REVIEWS

Results from study two suggest that frontline workers experience additional instances of customer-originating racism and sexism when algorithmically-informed managerial practices make them seem incompetent. The suggested reason provided by *Hypothesis 4a* relies on the negative competency stereotype, which states that customers expect women and workers of color to be less competent than their white and male counterparts. The final study in this paper considers *Hypothesis 4a* from the perspective of Yelp reviews, testing for evidence of racial/ethnic and gender bias in consumer's evaluations of perceived competency. This analysis is meant to show that, if employees are indeed put in a situation in which they are set up to seem incompetent, this incompetence will be interpreted in a gendered and racialized manner.

To conduct this analysis, I draw on recent advances in computational text analysis, using word embedding models to map how concepts relate to one another. This method, popularized by Kozlowski et al. (2019), analyzes cooccurrences of words within large text corpora using machine learning models. These models produce a high-dimensional vector space in which each word in a corpus is associated with a vector (otherwise described as a word embedding), and words that are used in similar contexts are located more closely to one another within this vector space. For a more detailed description of word embedding methods and their usage in this paper, see Appendix C.

These word embeddings have been shown to encode culture in meaningful ways (Garg et al., 2018; Kozlowski et al., 2019). The most classic example demonstrating the meanings encoded in word embedding by considering the case of analogies. For instance, with word embeddings produced using a large corpus of English text when performing the following operations for the embeddings tied to the following words: "King"-"Man"+"Woman," the most similar vector found is for "Queen." Kozlowski et al. (2019) take this method a step forward by producing semantic axes based on word pairs, for instance a gender axis with "Man" on one end and "Woman" on the other, and identifies where other words fall on this axis. For instance, when considering sports, "Softball" appears closer to the "Woman" side of the axis, and "Baseball" is much closer to the "Man" side.

Using *Word2vec,* a common tool to produce word embeddings, I construct word embeddings of Yelp reviews, producing a gender axis and a race/ethnicity axis. To reduce sensitivity to word choice, I began with the terms "man," "woman," "white" and "black," searched word embeddings for nearest neighbors of each term, and averaged the search term with relevant, closely related words (i.e. "man," "guy," "dude," "gentleman" or "white," "caucasian"). I then chose a series of four opposing concept pairs to include in the analysis: competent-incompetent; efficient-inefficient; helpful-unhelpful; slow-fast. For each concept in each concept pair I averaged the word embedding of the word itself with the word embeddings for the 20 most similar words in the space. For example, the top four most similar words in embedding space to incompetent were "inept," "uncaring," "unprofessional," and "ignorant." A discussion of similar words, word choice, and sensitivity analyses are provided in Appendix C.

By analyzing the distance between an opposing word pair such as "competent"-"incompetent" on the axes of "Man-Woman" and "White-Black" in the word embedding space derived by Yelp Reviews, it is possible to determine whether the positive word "competent" is more closely related to "Man" and "White" compared to the negative word "incompetent." If the positive words are more closely associated with "White" and "Man" than their negative counterparts in this word embedding space, this would suggest an underlying negative competency bias in the Yelp review corpus.

Results

Figure 4 below demonstrates the results for the 4 concept pairs projected onto an X-axis of gender and a Y-axis of race/ethnicity. For ease of interpretation, distances are presented in Z-scores calculated based on the average score of each word in the corpus, as well as the standard deviation. A one point negative shift values along the X-axis can be interpreted then as a word being one standard deviation more masculine than the average word in the corpus, and a one point negative shift the Y-axis with one standard deviation more White. As a result, the lower-left quadrant can be associated with whiteness and masculinity, and the upper right quadrant of the figure can bae associated with black-ness and femininity.

Figure 4 About Here

The figure demonstrates that, for each of the four concept pairs, the negative concept is more closely associated with femininity and blackness than its positive counterpart. For three out of the four, the positive concept is explicitly associated with whiteness and masculinity, and the corresponding negative concept is explicitly associated with black-ness and femininity. For the concept of efficiency, for example, the positive word-pair is coded in Yelp embedding space as one standard deviation more masculine, and 1.5 standard deviations more White than the negative word "inefficient." Similar results hold for the other three pairs. These results highlight that, in the context of Yelp reviews, perceptions of competency are indeed gendered and racialized in the ways predicted by the negative competency stereotype. These results provide additional support for *Hypothesis 4a*, predicting that women and racial/ethnic minorities are more likely than their white and male counterparts to experience bias and discrimination when they are put in positions where they are made to seem incompetent.

CONCLUSION

This article has demonstrated that the implementation of quantified management strategies in customer service domains is associated with customer-originating mistreatment directed at frontline service workers. From the perspective of frontline employees, qualitative survey findings highlight that that quantified decision making has effects on both the quality of customer service and how customer's perceive employees. Quantification of work outputs that emphasize speed lead employees to appear cold and *robotic*. Quantification of work inputs, for instance optimizing a sales to staffing ratio in thirty minute increments, lead employees to feel understaffed, undertrained and *incompetent* in the eyes of customers. Quantitative survey evidence provides additional confirmation for these processes, showing that managerial practices associated with roboticism and incompetence are associated with customer mistreatment. Moreover, when employees experience managerial practices that make them appear incompetent to customers, they are more likely to be exposed to racism and sexism. In a final computational analysis of Yelp reviews I demonstrate how this association between incompetency and racist and sexist remarks is likely due to the negative competency stereotypes that customers hold regarding women and racial/ethnic minorities. Theoretical implications for quantification, customer mistreatment and workplace discrimination are discussed below.

Theoretical Implications

Quantification. This research makes multiple contribution to the literature on quantification. First, I consider the impact of quantification in an overlooked area: low-wage service work. This type of work is ubiquitous in today's service-focused economy, is highly bureaucratized (Bolton & Houlihan, 2010), and is the subject of ongoing, algorithmically-informed efforts to improve efficiency through processes of quantification, automation, and roboticization. With respect to service-work, a focus on platforms has led to important discussions regarding the power dynamics between customers, employees, and managers. For instance, customer reviews fundamentally alter the nature of work and control in platform work (Bellesia et al., 2023; Cameron & Rahman, 2022; Curchod et al., 2020; Maffie, 2020). In bureaucratized frontline foodservice and retail work, on the other hand, this research suggests that employees act as an important buffer in quantification efforts, absorbing the negative effects of inefficiencies produced through quantification. Employee's experiences of increased stress and overwork, as well as customer mistreatment, including sexism, and racism, become unmeasured externalities as upper management fine-tune's algorithms designed to improve organizational efficiency.

As opposed to Chown's (2021) description of professionals engaging in feedback loops with managers to make new programs work better, frontline workers must instead find ways to make the new programs work. In doing so, they must make difficult decisions between doing what the organization claims it does (i.e. provide friendly and high quality service to customers) and what the organization wants (i.e. improve efficiency and make the numbers look good) (Korczynski & Ott, 2004; Misra & Walters, 2022; Storer, 2022). Low-level employees suffer the consequences of this dual-pull, personally absorbing the negative effects of an increased focus on algorithmically-informed managerial practices by making themselves the target of mistreatment, racism and sexism. At the same time that some numbers become increasingly important in decision-making, these negative externalities likely go unmeasured, contributing to the black box surrounding the meanings of numbers and prediction for managers (Anthony, 2021).

The research also provides unique insight into the effect of quantification on how service work is performed. While research shows that quantification can reduce effort in factory work (Ranganathan & Benson, 2020), and alter strategies in relatively autonomous platform work (Cameron, 2022), customer-facing bureaucratized work provides less room to maneuver. This article argues that in this context, quantification impacts not only the quality of work outputs, or how employees feel and how employees respond to the system, but also how quantification impacts how third-parties view service providers. In particular, I show that quantification of different aspects of work can have different results on how employee's feel they are perceived – as robotic or incompetent. This is particularly important since research shows that managers have gone through great effort to develop display rules and motivate workers to conduct emotional labor in order to alter customer's perceptions of the service interaction (Hochschild, 1983; Leidner, 1993). As a result, quantification and the coupled processes of quantification rituals (Mazmanian & Beckman, 2018) and performativity (MacKenzie & Millo, 2003) may unintentionally alter service quality, undoing some of the work that other parts of the organization, such as branding and marketing, have emphasized. This is especially clear from the qualitative results, where worker's must decide between helping customers and meeting their restocking quotas, scan times, or program enrollment numbers. From the customer's perspective this creates a situation in which the organization sets expectations that its operations are not designed to meet, and employee's bare the brunt of this friction.

Customer Mistreatment and Workplace Discrimination. This study also addresses a clear gap in the literature regarding customer mistreatment's organizational antecedents, and shows how organizational decision-making can produce stereotype threat for employees. Uncovering a middle ground between "bad apple" customers and employees (Sliter & Jones, 2016) and the cultural setting in which customers are told they are always right (Korczynski & Ott, 2004), this article shows how a series of operational decisions produce customer mistreatment. These studies makes clear the fact that organizations can control who experiences mistreatment, sexism and racism. This study has focused specifically on the relationship between quantified decision making and customer-based mistreatment, but there are likely more operational decisions at work. A promising body of work has focused on how managerial abuse, for instance, trickles down to third-party observers in the form of coworkers (Xu et al., 2020). This dynamic is also likely in play among customers, and likely varies with respect to race and gender. In the context of a complicated service triangle (Leidner, 1993), managers likely have the power to buffer workers from mistreatment, or to compound or activate perceptions of incompetence for customers. The interplay of organizational decision making and managerial latitude in the context of service work presents a key source of future research on mistreatment.

In addition, scholars in management agree that curbing workplace racism and sexism remains one of the grand challenges of this century, and research in this area has risen precipitously over recent decades. Yet while customer-facing service work represents a sizable portion of the economy, the management literature typically focuses on discrimination from others within the organization, whether it be at point of hire, in pay negotiations and evaluations, or in promotion. In the context of service work, customer-originating discrimination likely has different antecedents than other sources of workplace discrimination. This article makes the important contribution of expanding the scope of research on both workplace discrimination and on customer mistreatment to include the extent to which employees are exposed to racist and sexist remarks from customers. While research has been particularly concerned with discrimination in the workplace, there is a clear gap in the extent to which this literature focuses on mistreatment from organizational outsiders. This gap extends to research on customer-based mistreatment, where only a few qualitative studies have emphasized the importance of customer-originating discrimination. While this research takes a first step in quantifying experiences of racism and sexism from customers, more needs to be done to unpack these dynamics.

Implications For Managers

In the context of quantification and algorithmic work, managers in retail and foodservice implementing new programs surrounding, for example, coupons, memberships, credit cards, or the use of automated ordering and checkout counters, are likely to consider uptake, sales, and customer satisfaction surveys in order to quantify their success. Research on quantification suggests that these metrics are at risk of being seen as concrete indicators of a program's efficacy, rather than a series of suggestive numbers. This research shows that such algorithms would likely miss a key characteristic of employee experience in the rollout of such programs – the experience of customer-originating mistreatment. While customer-originating mistreatment has been shown to be a key factor in predicting productivity as well as turnover, this article draws a clear connection between mistreatment and practices such as quantification, staffing decisions, and the amount of training new employees are given. Without measuring mistreatment, there may be long-term unmeasured impacts, such as turnovers, that could be difficult to explain.

Managers devoted to reducing racism and sexism in the workplace should be especially concerned about the extent to which such decisions may make frontline workers seem incompetent in the eyes of customers. While some decisions force employees to favor efficiency over quality customer service, making them seem *robotic*, this article suggests that decisions such as cutting staff and training time can lead workers to seem *incompetent*. When women and workers of color are put in positions that make them seem incompetent, this may trigger customer's gendered and racialized negative competency stereotypes, leading these workers to experience higher levels of racism and sexism in the workplace. Managers concerned about the extent to which customers receive such comments should consider systematically collecting this data and including it in measurements of operational efficacy. Managers should be especially aware that decisions to cut labor costs to increase profit may come at a higher cost to female workers and workers of color who, as a result, may experience more discrimination.

More broadly, as new programs and operational changes are implemented, managers should be aware that frontline workers have a unique perspective on service quality that may be lost by higher level managers concerned specifically with customer well-being. This research suggests that this channel for insight is underutilized, with managers focusing on prioritizing customer experience surveys, which while important, likely do not capture the whole picture. As such, as organizations grow and become more heavily bureaucratized, managers concerned with the quality of the customer experience would benefit from developing and maintaining feedback channels between frontline employees and high-level managers. This is particularly important in the context of complex managerial hierarchy, where there are potentially multiple levels of managers between a frontline employee and a store manager, let alone between store manager and corporate managers.

REFERENCES

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics. [electronic resource]: An empiricist's companion*. Princeton University Press.

Anthony, C. (2021). When Knowledge Work and Analytical Technologies Collide: The Practices and Consequences of Black Boxing Algorithmic Technologies. *Administrative Science Quarterly*, *66*(4), 1173–1212. https://doi.org/10.1177/00018392211016755

Atlason, J., Epelman, M. A., & Henderson, S. G. (2008). Optimizing Call Center Staffing Using Simulation and Analytic Center Cutting-Plane Methods. *Management Science*, *54*(2), 295–309. https://doi.org/10.1287/mnsc.1070.0774

Batt, R., & Colvin, A. J. S. (2011). An Employment Systems Approach to Turnover: Human Resources Practices, Quits, Dismissals, and Performance. *Academy of Management Journal*, *54*(4), 695–717. https://doi.org/10.5465/amj.2011.64869448

Bellesia, F., Mattarelli, E., & Bertolotti, F. (2023). Algorithms and their Affordances: How Crowdworkers Manage Algorithmic Scores in Online Labour Markets. *Journal of Management Studies*, *60*(1), 1–37. https://doi.org/10.1111/joms.12870

Berry Jaeker, J. A., & Tucker, A. L. (2017). Past the Point of Speeding Up: The Negative Effects of Workload Saturation on Efficiency and Patient Severity. *Management Science*, *63*(4), 1042–1062. https://doi.org/10.1287/mnsc.2015.2387

Billingsley, B. L. (2016). Racialized and Classed Contexts: Shifting Audiences and Changes in Emotional Labor Among Restaurant Servers. *Sociological Inquiry*, *86*(4), 641–657. https://doi.org/10.1111/soin.12135

Bolton, S. C., & Houlihan, M. (2010). Bermuda Revisited?: Management Power and Powerlessness in the Worker–Manager–Customer Triangle. *Work and Occupations*, *37*(3), 378–403. https://doi.org/10.1177/0730888410375678

Butler, P., & Hammer, A. (2019). 'A Minute's a Life-Time in Fast-Food!': Managerial Job Quality in the Quick Service Restaurant Sector. *Work, Employment and Society*, *33*(1), 96–111. https://doi.org/10.1177/0950017018777710

Cameron, L. D. (2022). "Making Out" While Driving: Relational and Efficiency Games in the Gig Economy. *Organization Science*, *33*(1), 231–252. https://doi.org/10.1287/orsc.2021.1547

Cameron, L. D., & Rahman, H. (2022). Expanding the Locus of Resistance: Understanding the Co-constitution of Control and Resistance in the Gig Economy. *Organization Science*, *33*(1), 38–58. https://doi.org/10.1287/orsc.2021.1557

Carton, A. M., & Rosette, A. S. (2011). Explaining Bias against Black Leaders: Integrating Theory on Information Processing and Goal-Based Stereotyping. *Academy of Management Journal*, *54*(6), 1141–1158. https://doi.org/10.5465/amj.2009.0745

Cezik, M. T., & L'Ecuyer, P. (2008). Staffing Multiskill Call Centers via Linear Programming and Simulation. *Management Science*, *54*(2), 310–323. https://doi.org/10.1287/mnsc.1070.0824

Chown, J. (2021). The Unfolding of Control Mechanisms inside Organizations: Pathways of Customization and Transmutation. *Administrative Science Quarterly*, *66*(3), 711–752. https://doi.org/10.1177/0001839220980015

Cortina, L. M. (2008). Unseen Injustice: Incivility as Modern Discrimination in Organizations. *Academy of Management Review*, *33*(1), 55–75. https://doi.org/10.5465/amr.2008.27745097

Curchod, C., Patriotta, G., Cohen, L., & Neysen, N. (2020). Working for an Algorithm: Power Asymmetries and Agency in Online Work Settings. *Administrative Science Quarterly*, *65*(3), 644–676. https://doi.org/10.1177/0001839219867024

De Stefano, F., Bonet, R., & Camuffo, A. (2019). Does Losing Temporary Workers Matter? The Effects of Planned Turnover on Replacements and Unit Performance. *Academy of Management Journal*, *62*(4), 979–1002. https://doi.org/10.5465/amj.2017.0291

Dwyer, R. E. (2013). The Care Economy? Gender, Economic Restructuring, and Job Polarization in the U.S. Labor Market. *American Sociological Review*, *78*(3), 390–416. https://doi.org/10.1177/0003122413487197

Foschi, M. (2000). Double Standards for Competence: Theory and Research. *Annual Review of Sociology*, *26*, 21–42.

Gabriel, A. S., & Diefendorff, J. M. (2015). Emotional Labor Dynamics: A Momentary Approach. *Academy of Management Journal*, *58*(6), 1804–1825. https://doi.org/10.5465/amj.2013.1135

Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, *115*(16), E3635–E3644. https://doi.org/10.1073/pnas.1720347115

Good, L., & Cooper, R. (2016). 'But It's Your Job To Be Friendly': Employees Coping With and Contesting Sexual Harassment from Customers in the Service Sector. *Gender, Work & Organization*, *23*(5), 447–469. https://doi.org/10.1111/gwao.12117

Green, L. V., Savin, S., & Savva, N. (2013). "Nursevendor Problem": Personnel Staffing in the Presence of Endogenous Absenteeism. *Management Science*, *59*(10), 2237–2256. https://doi.org/10.1287/mnsc.2013.1713

Groth, M., & Grandey, A. (2012). From bad to worse: Negative exchange spirals in employee–customer service interactions. *Organizational Psychology Review*, *2*(3), 208–233. https://doi.org/10.1177/2041386612441735

Groth, M., Hennig-Thurau, T., & Walsh, G. (2009). Customer Reactions to Emotional Labor: The Roles of Employee Acting Strategies and Customer Detection Accuracy. *Academy of Management Journal*, *52*(5), 958–974. https://doi.org/10.5465/amj.2009.44634116

Groth, M., Wu, Y., Nguyen, H., & Johnson, A. (2019). The moment of truth: A review, synthesis, and research agenda for the customer service experience. *Annual Review of Organizational Psychology and Organizational Behavior*, *6*, 89–113. https://doi.org/10.1146/annurev-orgpsych-012218-015056

Hadjisolomou, A. (2021). Doing and Negotiating Transgender on the Front Line: Customer Abuse, Transphobia and Stigma in the Food Retail Sector. *Work, Employment and Society*, 0950017020977331. https://doi.org/10.1177/0950017020977331

Hadjisolomou, A., & Simone, S. (2021). Profit over People? Evaluating Morality on the Front Line during the COVID-19 Crisis: A Front-Line Service Manager's Confession and Regrets. *Work, Employment and Society*, *35*(2), 396–405. https://doi.org/10.1177/0950017020971561

Hochschild, A. R. (2012). *The managed heart. [electronic resource]: Commercialization of human feeling*. Berkeley : University of California Press. (Original work published 1983)

Ikeler, P. (2016). Deskilling emotional labour: Evidence from department store retail. *Work, Employment and Society*, *30*(6), 966–983. https://doi.org/10.1177/0950017015609031

Kacmar, K., Andrews, M. C., Van Rooy, D. L., Chris Steilberg, R., & Cerrone, S. (2006). Sure Everyone Can Be Replaced … But At What Cost? Turnover As A Predictor Of Unit-

Level Performance. *Academy of Management Journal*, *49*(1), 133–144. https://doi.org/10.5465/amj.2006.20785670
Kamalahmadi, M., Yu, Q., & Zhou, Y.-P. (2021). Call to Duty: Just-in-Time Scheduling in a Restaurant Chain. *Management Science*, *67*(11), 6751–6781. https://doi.org/10.1287/mnsc.2020.3877
Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at Work: The New Contested Terrain of Control. *Academy of Management Annals*, *14*(1), 366–410. https://doi.org/10.5465/annals.2018.0174
Kern, J. H., & Grandey, A. A. (2009). Customer incivility as a social stressor: The role of race and racial identity for service employees. *Journal of Occupational Health Psychology*, *14*(1), 46–57. https://doi.org/10.1037/a0012684
Kesavan, S., Lambert, S. J., Williams, J. C., & Pendem, P. K. (2022). Doing Well by Doing Good: Improving Retail Store Performance with Responsible Scheduling Practices at the Gap, Inc. *Management Science*, *68*(11), 7818–7836. https://doi.org/10.1287/mnsc.2021.4291
Korczynski, M., & Evans, C. (2013). Customer abuse to service workers: An analysis of its social creation within the service economy. *Work, Employment and Society*, *27*(5), 768–784. https://doi.org/10.1177/0950017012468501
Korczynski, M., & Ott, U. (2004). When Production and Consumption Meet: Cultural Contradictions and the Enchanting Myth of Customer Sovereignty. *Journal of Management Studies*, *41*(4), 575–599. https://doi.org/10.1111/j.1467-6486.2004.00445.x
Kozlowski, A. C., Taddy, M., & Evans, J. A. (2019). The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings. *American Sociological Review*, *84*(5), 905–949. https://doi.org/10.1177/0003122419877135
Lambert, S. J. (2008). Passing the buck: Labor flexibility practices that transfer risk onto hourly workers. *Human Relations*, *61*(9), 1203–1227. https://doi.org/10.1177/0018726708094910
Leidner, R. (1993). *Fast food, fast talk. [electronic resource]: Service work and the routinization of everyday life*. Berkeley : University of California Press, c1993.
MacKenzie, D., & Millo, Y. (2003). Constructing a Market, Performing Theory: The Historical Sociology of a Financial Derivatives Exchange. *American Journal of Sociology*, *109*(1), 107–145. https://doi.org/10.1086/374404
Maffie, M. D. (2020). The Perils of Laundering Control through Customers: A Study of Control and Resistance in the Ride-hail Industry. *ILR Review*, 0019793920972679. https://doi.org/10.1177/0019793920972679
Mazmanian, M., & Beckman, C. M. (2018). “Making” Your Numbers: Engendering Organizational Control Through a Ritual of Quantification. *Organization Science*, *29*(3), 357–379. https://doi.org/10.1287/orsc.2017.1185
Mirchandani, K. (2003). Challenging Racial Silences in Studies of Emotion Work: Contributions from Anti-Racist Feminist Theory. *Organization Studies*, *24*(5), 721–742. https://doi.org/10.1177/0170840603024005003
Misra, J., & Walters, K. (2022). *Walking Mannequins: How Race and Gender Inequalities Shape Retail Clothing Work*. University of California Press.
Muhs, G. G. y, Niemann, Y. F., González, C. G., & Harris, A. P. (2012). *Presumed Incompetent: The Intersections of Race and Class for Women in Academia*. University Press of Colorado.

Phillips, L. T., Jun, S., & Shakeri, A. (2022). Barriers and Boosts: Using Inequity Frames Theory to Expand Understanding of Mechanisms of Race and Gender Inequity. *Academy of Management Annals*, *16*(2), 547–587. https://doi.org/10.5465/annals.2020.0314

Porath, C. L., & Erez, A. (2007). Does Rudeness Really Matter? The Effects of Rudeness on Task Performance and Helpfulness. *Academy of Management Journal*, *50*(5), 1181–1197. https://doi.org/10.5465/amj.2007.20159919

Rahman, H. A. (2021). The Invisible Cage: Workers' Reactivity to Opaque Algorithmic Evaluations. *Administrative Science Quarterly*, 00018392211010118. https://doi.org/10.1177/00018392211010118

Ranganathan, A., & Benson, A. (2020). A Numbers Game: Quantification of Work, Auto-Gamification, and Worker Productivity. *American Sociological Review*, *85*(4), 573–609. https://doi.org/10.1177/0003122420936665

Ren, Z. J., & Zhou, Y.-P. (2008). Call Center Outsourcing: Coordinating Staffing Level and Service Quality. *Management Science*, *54*(2), 369–383. https://doi.org/10.1287/mnsc.1070.0820

Sallaz, J. J. (2015). Permanent Pedagogy: How Post-Fordist Firms Generate Effort but Not Consent. *Work and Occupations*, *42*(1), 3–34. https://doi.org/10.1177/0730888414551207

Schilpzand, P., De Pater, I. E., & Erez, A. (2016). Workplace incivility: A review of the literature and agenda for future research. *Journal of Organizational Behavior*, *37*(S1), S57–S88. https://doi.org/10.1002/job.1976

Schneider, D., & Harknett, K. (2019a). Consequences of Routine Work-Schedule Instability for Worker Health and Well-Being. *American Sociological Review*, *84*(1), 82–114.

Schneider, D., & Harknett, K. (2019b). What's to Like? Facebook as a Tool for Survey Data Collection. *Sociological Methods & Research*, *51*(1), 108–140. https://doi.org/10.1177/0049124119882477

Schneider, D., & Harknett, K. (2021). Hard Times: Routine Schedule Unpredictability and Material Hardship among Service Sector Workers. *Social Forces*, *99*(4), 1682–1709. https://doi.org/10.1093/sf/soaa079

Siebert, W. S., & Zubanov, N. (2009). Searching for The Optimal Level of Employee Turnover: A Study of a Large U.K. Retail Organization. *Academy of Management Journal*, *52*(2), 294–313. https://doi.org/10.5465/amj.2009.37308149

Sliter, M., & Jones, M. (2016). A qualitative and quantitative examination of the antecedents of customer incivility. *Journal of Occupational Health Psychology*, *21*, 208–219. https://doi.org/10.1037/a0039897

Stice-Lusvardi, R., Hinds, P. J., & Valentine, M. (2023). Legitimating Illegitimate Practices: How Data Analysts Compromised Their Standards to Promote Quantification. *Organization Science*. https://doi.org/10.1287/orsc.2022.1655

Storer, A. (2022). Serving Like an Organization: How Foodservice and Retail Workers Interpret Their Interactions With Customers. *Work and Occupations*, *Advanced online publication*, 07308884221134582. https://doi.org/10.1177/07308884221134582

Storer, A., Schneider, D., & Harknett, K. (2020). What Explains Racial/Ethnic Inequality in Job Quality in the Service Sector? *American Sociological Review*, *85*(4), 537–572. https://doi.org/10.1177/0003122420930018

Tan, T. F., & Netessine, S. (2014). When Does the Devil Make Work? An Empirical Study of the Impact of Workload on Worker Productivity. *Management Science*, *60*(6), 1574–1593. https://doi.org/10.1287/mnsc.2014.1950

Taylor, F. W., 1856-1915. (2004). *The Principles of Scientific Management. [Electronic resource]*. Project Gutenberg, 2004.

Wang, M., Liao, H., Zhan, Y., & Shi, J. (2011). Daily Customer Mistreatment and Employee Sabotage Against Customers:Examining Emotion and Resource Perspectives. *Academy of Management Journal*, *54*(2), 312–334. https://doi.org/10.5465/amj.2011.60263093

Xu, E., Huang, X., Jia, R., Xu, J., Liu, W., Graham, L., & Snape, E. (2020). The “Evil Pleasure”: Abusive Supervision and Third-Party Observers’ Malicious Reactions Toward Victims. *Organization Science*, *31*(5), 1115–1137. https://doi.org/10.1287/orsc.2019.1349

TABLES AND FIGURES

Table 1: Descriptive Statistics

Variable	Mean	Min/Max	Variable	Mean	Min/Max
Dependent Variables			**Job Quality Controls**		
Customers Raise Voice	3.02	1/5	Job Tenure	3.38	1/5
Customers Sexist	1.84	1/5	Full-Time	55.39%	
Customers Racist	1.81	1/5	Current Worker	89.38%	
			Managerial Level		
Independent Variables			*Store Manager*	4.80%	
Understaffing	3.99	1/5	*Middle Manager*	30.45%	
Inadequate Training	3.07	1/5	*Frontline*	64.75%	
Misaligned Metrics	3.51	1/5	Hourly Wage	16.30	
			Usual Hours	32.91	
Demographics and Human Capital Controls			Hourly Variability	0.40	0/1
Has Kids	53.25%		Freq. Customer Interactions	4.71	1/5
Cohabitating	52.79%				
Cis-Gender Male	20.43%		**Self-Reported Establishment Controls**		
White	84.83%		Prop. Customers Male	3.87	1/7
Age	39.24		Prop. Customers White	4.34	1/7
Education	1.91	1/3	Prop. Coworkers Male	3.26	1/7
Enrolled in School	18.41%		Prop. Coworkers White	4.55	1/7
English as a Second Language (ESL)	12.90%		Customer Class Background	3.61	1/5
			Company Doing Well	5.50	1/7
N	1271		**Halo Effect Controls**		
			Job Satisfaction	4.49	1/7
			Turnover Intentions	3.46	1/7
			Job Meaningfulness	4.47	1/7
			Self-Reported Effort	6.00	1/7

Table 2: The Effects of Perceived Problems with Understaffing, Inadequate Training and Misaligned Metrics on Frequency of Customers Rasing Their Voices at Employees

	Model 1	Model 2	Model 3	Model 4	Model 5
	b/se	b/se	b/se	b/se	b/se
Understaffing	0.162***	0.129***	0.134***	0.125***	0.103***
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Inadequate Training	0.172***	0.189***	0.191***	0.196***	0.124**
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
Misaligned Metrics	0.179***	0.142***	0.138***	0.125***	0.107**
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
ESL		0.235	0.210	0.212	0.143
		(0.13)	(0.13)	(0.12)	(0.10)
White		-0.128	-0.125	-0.106	-0.138
		(0.13)	(0.12)	(0.12)	(0.12)
Cis-male		-0.206*	-0.184*	-0.115	-0.111
		(0.09)	(0.08)	(0.08)	(0.08)
Kids		-0.157	-0.154	-0.162	-0.159
		(0.08)	(0.09)	(0.09)	(0.09)
Cohabitating		0.123	0.136	0.141	0.138
		(0.09)	(0.08)	(0.09)	(0.08)
Age		-0.013**	-0.013**	-0.013**	-0.011**
		(0.00)	(0.00)	(0.00)	(0.00)
Hourly Wage		0.004	-0.004	-0.003	-0.002
		(0.01)	(0.01)	(0.01)	(0.01)
Usual Hours		0.006	0.004	0.004	0.002
		(0.00)	(0.00)	(0.00)	(0.00)
Prop. Customers Male				0.069	0.075
				(0.05)	(0.05)
Prop. Customers White				-0.081	-0.077
				(0.04)	(0.04)
Prop. Coworkers Male				0.016	0.007
				(0.05)	(0.05)
Prop. Coworkers White				0.021	0.027
				(0.03)	(0.03)
Customer Class Background				0.059	0.031
				(0.05)	(0.05)
Freq. Customer Interactions				0.162*	0.168*
				(0.08)	(0.08)
Company Doing Well				-0.009	0.007
				(0.02)	(0.02)
Job Satisfaction					-0.117**
					(0.03)
Turnover Intentions					0.013
					(0.02)
Job Meaningfulness					-0.023
					(0.03)
Self-Reported Effort					0.008
					(0.03)
Constant	1.228***	1.467**	1.830***	0.876	1.625**
	(0.10)	(0.36)	(0.35)	(0.47)	(0.48)
N	1271	1271	1271	1271	1271
Ind. Controls	No	Yes	Yes	Yes	Yes
Fixed Effects	No	No	No	Yes	Yes
Company Controls	No	No	Yes	Yes	Yes
Halo Controls	No	No	No	No	Yes

Standard Errors Clustered by Company; Multiple Imputation for Item Non-Respose; Only selected variables from each control group displayed in each model

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: The Effects of Perceived Problems with Understaffing, Inadequate Training and Misaligned Metrics on Frequency of Customer Mistreatment, Sexism and Racism

	Customers Yell		Customers Sexist		Customers Racist	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
	b/se	b/se	b/se	b/se	b/se	b/se
Understaffing	0.125***	0.103***	0.084**	0.066*	0.066***	0.047*
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)
Inadequate Training	0.196***	0.124**	0.150***	0.085**	0.130***	0.077*
	(0.03)	(0.04)	(0.02)	(0.02)	(0.03)	(0.03)
Misaligned Metrics	0.125***	0.107**	0.044	0.026	0.011	-0.005
	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
ESL	0.212	0.143	0.375*	0.309*	0.349**	0.284*
	(0.12)	(0.10)	(0.13)	(0.13)	(0.11)	(0.10)
White	-0.106	-0.138	0.107	0.082	-0.039	-0.056
	(0.12)	(0.12)	(0.09)	(0.08)	(0.11)	(0.10)
Cis-male	-0.115	-0.111	-0.045	-0.051	0.035	0.033
	(0.08)	(0.08)	(0.11)	(0.11)	(0.07)	(0.07)
Kids	-0.162	-0.159	-0.135	-0.126	0.022	0.024
	(0.09)	(0.09)	(0.08)	(0.08)	(0.08)	(0.07)
Cohabitating	0.141	0.138	0.138	0.133	0.068	0.069
	(0.09)	(0.08)	(0.07)	(0.07)	(0.05)	(0.05)
Age	-0.013**	-0.011**	-0.010**	-0.007*	-0.005	-0.002
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Hourly Wage	-0.003	-0.002	-0.008	-0.006	-0.007	-0.005
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Usual Hours	0.004	0.002	0.006	0.004	0.010**	0.009*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Prop. Customers Male	0.069	0.075	0.053	0.058	0.115*	0.116*
	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)
Prop. Customers White	-0.081	-0.077	-0.060	-0.056	-0.077	-0.071
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Prop. Coworkers Male	0.016	0.007	0.051	0.042	0.097*	0.091
	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)
Prop. Coworkers White	0.021	0.027	-0.004	-0.003	-0.078	-0.077
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)
Customer Class Background	0.059	0.031	0.058	0.032	0.011	-0.014
	(0.05)	(0.05)	(0.06)	(0.06)	(0.04)	(0.04)
Freq. Customer Interactions	0.162*	0.168*	-0.027	-0.016	-0.025	-0.023
	(0.08)	(0.08)	(0.06)	(0.06)	(0.06)	(0.05)
Company Doing Well	-0.009	0.007	-0.036	-0.015	-0.035	-0.018
	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)
Job Satisfaction		-0.117**		-0.075**		-0.050
		(0.03)		(0.02)		(0.03)
Turnover Intentions		0.013		0.032		0.055*
		(0.02)		(0.02)		(0.02)
Job Meaningfulness		-0.023		-0.031		-0.022
		(0.03)		(0.03)		(0.03)
Self-Reported Effort		0.008		-0.033		-0.001
		(0.03)		(0.05)		(0.03)
Constant	0.876	1.625**	1.818**	2.534***	1.614**	1.896**
	(0.47)	(0.48)	(0.56)	(0.54)	(0.48)	(0.50)
N	1271	1271	1271	1271	1271	1271
Ind. Controls	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Company Controls	Yes	Yes	Yes	Yes	Yes	Yes
Halo Controls	No	Yes	No	Yes	No	Yes

Standard Errors Clustered by Company; Multiple Imputation for Item Non-Respose; All Controls, Not all controls are displayed in the regression results reported.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure 1: Histograms of Dependent Variables

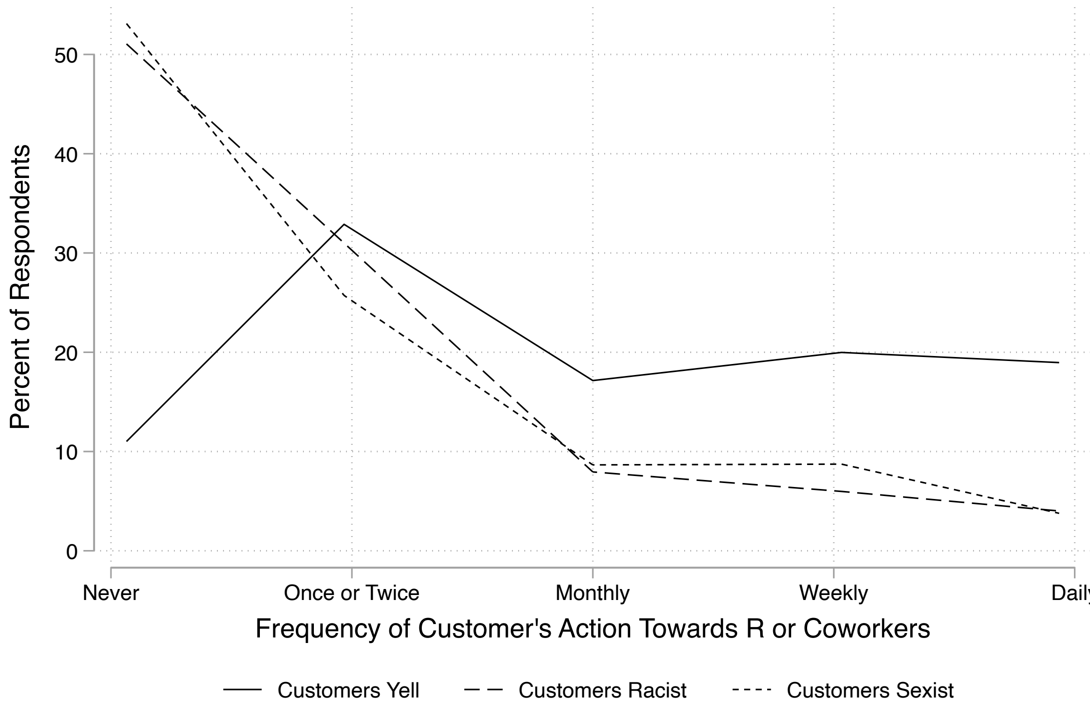


Figure 2: Histograms of Independent Variables

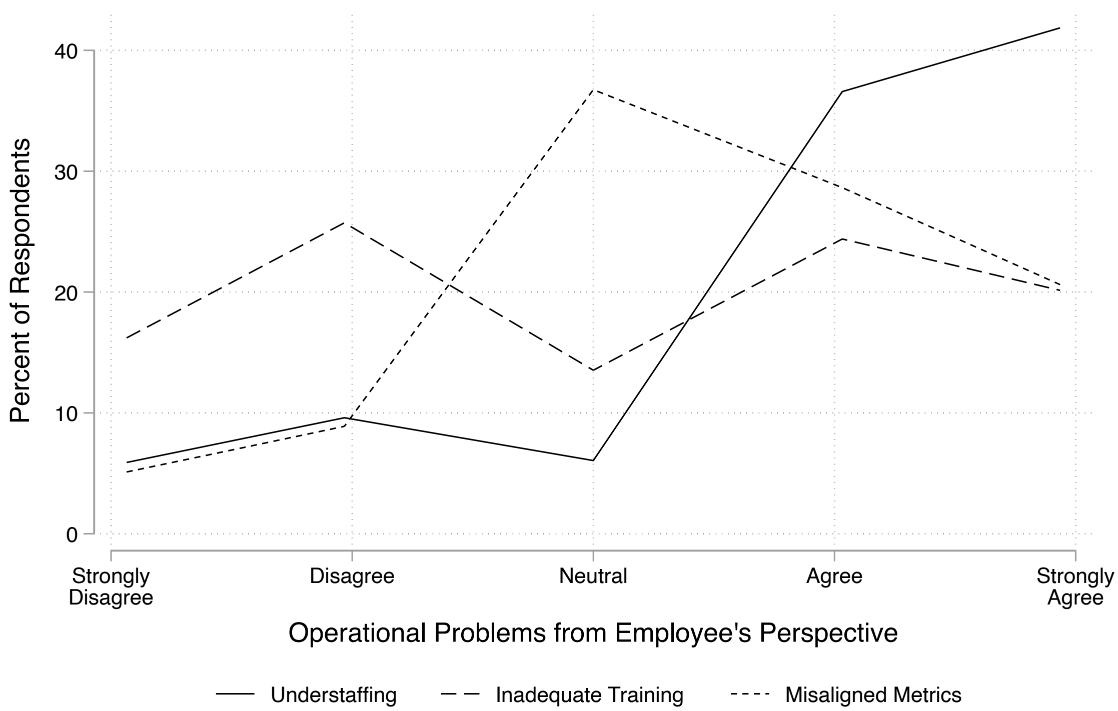


Figure 3: Average Marginal Effects of Joint Effects of Operational Decision Making on Experiences of Customer Mistreatment

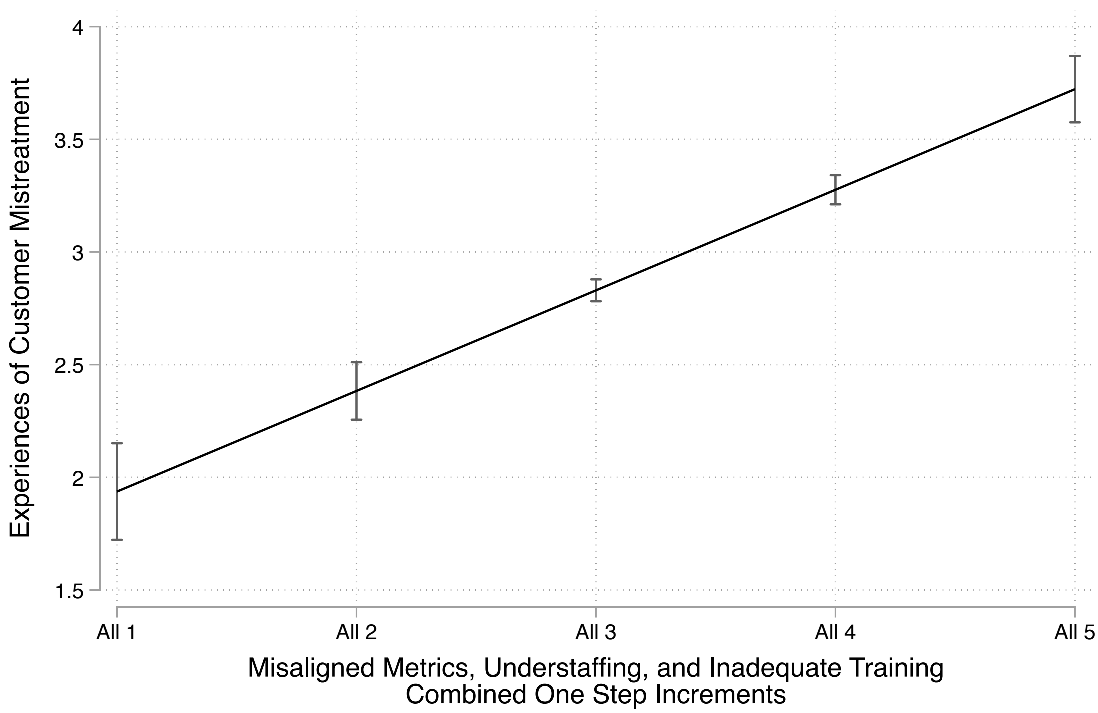
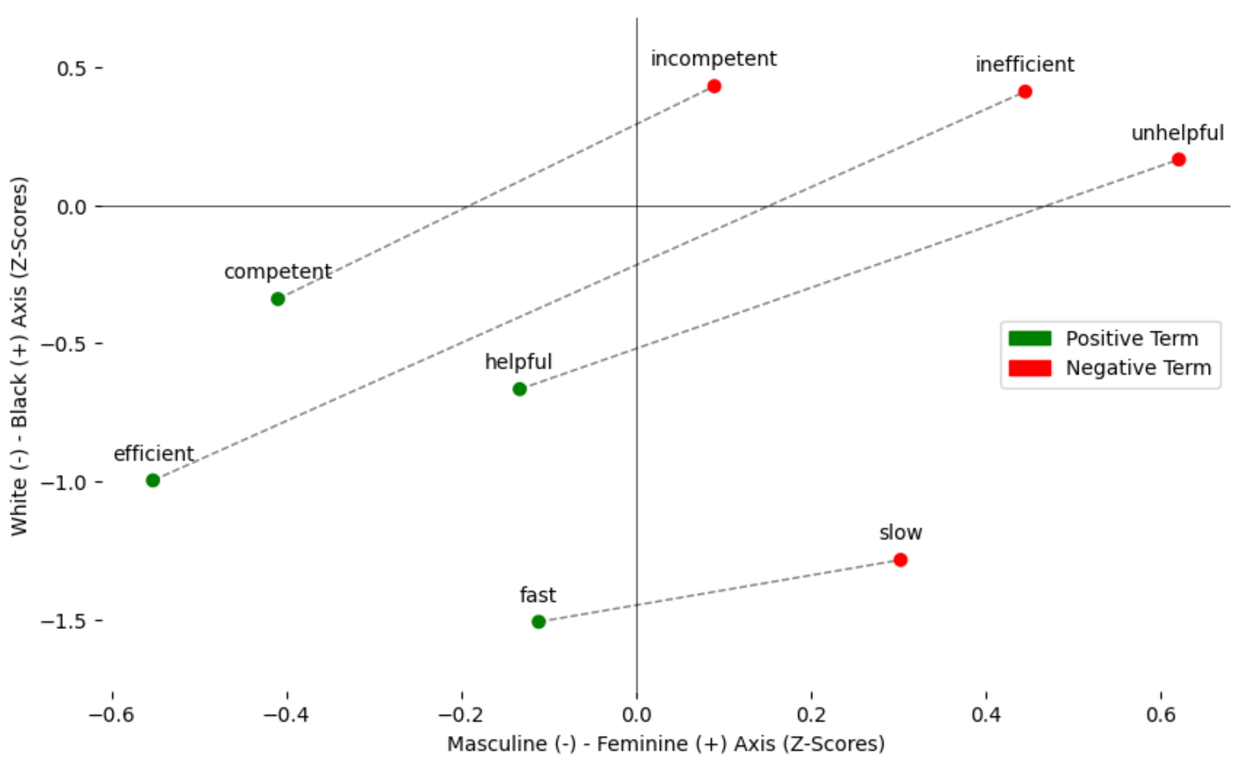


Figure 4: Gender and Racial/Ethnic Bias in Yelp Reviews of Service Competency in Word-Embedding Space



APPENDIX A

The quantitative study relies on the survey conducted between June and August of 2022. The click-through-rate of that survey is roughly 3%, which is roughly in line with Schneider & Harknett (2019b) and falls in line with the performance of other advertisements on the Facebook platform. While this seems low, non-employer driven surveys often have relatively low response rates (Keeter et al. 2017) and at the same time, online surveys have been found to be relatively accurate even among non-representative samples (Goel, Obeng, and Rothschild 2015). According to statistics provided by Qualtrics, roughly 91% of respondents completed at least 15% of the survey, and roughly 20% completed the entire survey. Compared to platforms such as Prolific or Amazon Mechanical Turk, with average response rates between 50% and 60% the response rate may seem small in comparison. However, the Facebook survey method presents significant upsides, including an opportunity to survey individuals at specific companies, and the ability to draw on a much larger pool of Americans on Facebook or Instagram, compared to the pool of individuals who opt-in to responding to surveys on these two platforms.

To contend with item non-response, I use multiple imputation for survey respondents who completed at least 85% of the survey. This strategy allows me to recover 522 responses that included at least one missing data point, for a total sample size of 1,271. Missing data was not imputed for dependent variables. Although demographically the sample leans white and female compared to estimates of the service sector provided by national surveys, I do not weigh the survey to match these statistics. This is because it is not currently possible to weigh the demographics of the sample in a way that definitively matches the demographics of the specific company.

In the qualitative study I also rely on a supplementary survey fielded in July of 2020 of 546 respondents at a smaller set of the same companies (The Home Depot, Lowe's, Walmart, Target, Walgreens, CVS, Starbucks). I use this survey to supplement the qualitative analysis by considering an open ended question asking employees to discuss the Pros and Cons of customer service at their employer.

Individual-Level Metrics for Establishment and Organizational Outcomes

Rather than drawing on objective or organizationally driven measures, this study uses individual-level data to measure both dependent variables, experiences of mistreatment, and organizational-level independent variables. With respect to the dependent variables, while measures such as customer experience surveys may be collected systematically by organizations, employee's own experiences of mistreatment are left unmeasured. As a result, these data must be collected from individuals. Organization-level dependent variables, on the other hand, such as the effectiveness of metrics, quality of training, and understaffing may be available, or have proxies, at the organization level. While it may be considered preferable to use organization-level measures of independent variables such as staffing due to its status as an objective measure, rather than a subjective measure collected by individuals, I identify two sets of arguments for why individual-level metrics with halo-effect controls provide a more suitable solution.

In the case of staffing, organization-level metrics may be problematic if staffing increases overall but changes in staffing even at the establishment level leave some functions well-staffed and other functions understaffed. This is particularly a concern in the case in establishments with a broad range of functions and departments, and could be particularly misleading when considering annual reports of large national companies. In this scenario, some objective measures of organizational operations would sacrifice conceptual depth for conceptual breadth. Second, collecting objective data on staffing levels in a specific function at a specific establishment would require working with a specific employer, which would sacrifice breadth in the organizational sample. In addition, this method could potentially introduce non-response bias at the organization level by only including employers that would be willing to work on such a research project. Quality of training may be even more difficult to measure, since individual managers may vary in the extent to which they value training within their own establishment.

This study, as a result, chooses to use employee-level information about these organizational measures, capturing their experiences of training, staffing, and metric quality. Individual level indicators are likely better able to capture more proximate-level issues for employees compared to organizational reports, and can be sampled without partnering with the organization.

The primary concern when including individual level measures is that some aspect of their subjectivity, such as their overall job satisfaction. If employees are simply in a bad mood when they fill out the survey, they may be likely to report everything negatively. To counteract these concerns, I include specifications that likely over control for this effect, by including controls for overall job satisfaction, turnover intentions, the meaning the respondent draws from their job, and their self-reported effort. These factors produce a strong test of the hypotheses, since understaffing, for example, likely impacts both job satisfaction and customer mistreatment. As a result, some of the overall association between understaffing and mistreatment will be captured by the halo effect controls.

Use of Single-Item Dependent Variables

In the quantitative analyses, I use three dependent variables, all of which are included as single items. This decision was based on two factors. First, single items allow for the collection of a broader range of concepts without unduly fatiguing the survey taker. Given the high attrition rate when surveying individuals via Facebook, where survey takers are “in the wild” (i.e. non-professional survey takers who are completing the survey without oversight from managers or researchers), it may be more appropriate to reduce the total amount of questions asked to the respondent. The use of single items, for instance, is becoming increasingly accepted in experience sampling designs, where researchers are concerned about respondent time and fatigue (Gabriel et al. 2019). In addition, single items are more appropriate when measuring specific activities, such as the experience of being yelled at, rather than psychological activities. Research is increasingly conducting studies showing that, for more objective, less psychological constructs, single items can appropriately substitute for multiple items (Bergkvist and Rossiter 2007; Fisher, Matthews, and Gibbons 2016; Matthews, Pineault, and Hong 2022).

Single items may also be more useful when previously developed scales do not accurately measure the construct being researched. When considering incivility scales developed for general incivility across contexts (Yao et al. 2022) and developed for customer incivility (Wilson and Holmvall 2013) or customer mistreatment (Wang et al. 2011) were capturing aspects of customer service that did not accurately reflect the setting and discussions from employees in the qualitative portion of the study. In addition, some items from Wilson and Holmvall's (2013) scale ask about customer’s complaining about the value of goods or service, which are outside the scope of what I am attempting to explain, or grumbling that there were too few employees working, which I am exploring as a *cause* of incivility, rather than an act of incivility in itself.

Given these argument, I focus on clear, objective single items: whether customers raised their voice, engaged in sexual harassment, or used racial/ethnic slurs. In pre-testing of 65 Starbucks and Walmart employees, shown in Table A-1 below, I did find that customer yelling is correlated to other, similar aspects of mistreatment.

Table A-1 About Here

APPENDIX B

Table B-1 below demonstrates the correlation matrix for variables included in the quantitative analysis. The correlation matrix shows that, while there is significant correlation between the independent variables, correlation is at acceptable levels, never exceeding a score of .2.

Table B-1 About Here

Table B-2 below includes the results mirroring those in Table 3 of the main text, using Ordinal Logistic Regression, rather than a linear model. Results show that both effect size and the significance of estimators follow similar patterns using this model.

Table B-2 About Here

APPENDIX C

Word embeddings are n-dimensional vectors that are used to represent a set of words in a text corpus. Embeddings are developed using machine learning. Developed by Mikolov et al. (2013), word embeddings use a neural network to iteratively use a word's context to predict a target word (defined as skip-grams, which are higher performing on small texts and lest frequently used words), or use a word to predict it's context (defined as continuous bag of words, performing better on larger corpora). In this prediction process, the authors rely on neural networks' "hidden layers" that reduce dimensionality by assigning each word in the corpus an n-dimensional vector, typically 300. These embedding models, and iterations on this technique, can then be used in a variety of contexts in natural language processing, such as predictive text. Many sets of word embedding models, and subsequent transformer models (which take into account the different use cases of words) have been developed using very large corpora of publicly or privately available text [cite].

These word embedding models are particularly useful to social science researchers because they embed meaning within them. In analogy tests, researchers find, for instance, that when taking the vector for King, subtracting the vector for Man, and adding the vector for Woman, the closest embedding vector is Queen. While publicly available embedding models can be useful for a variety of applications, especially when attempting to understand culture in general, it can be useful to develop embedding models with text directly related to a research question at hand. For instance, Storer (2022) sows that embeddings developed using Glassdoor data do a better job at finding relevant words for managers, customers, and employees in foodservice and retail contexts, than embeddings developed using large corpora of text. For these reasons, I use *Word2Vec* to develop embeddings drawing specifically on Yelp data using Skip Grams.

Recent advances in the Social Sciences have built upon the analogy technique described above, by developing Semantic Axes that capture key constructs like Race, Gender, and Affluence (Boutyline and Johnston 2023; Kozlowski, Taddy, and Evans 2019). Authors of these studies effectively show how concepts measured in semantic axes accurately capture social meaning, captured via surveys. As part of a simple explanation, Kozlowski et al. (2019) map different sports onto semantic axes of Affluence, showing that sports like Polo are associated with affluence, and camping or boxing associated with poverty.

I follow the same logic here, by producing race and gender axes in the Yelp data, and projecting different ways of discussing competency onto these axes. In order to control for variance in word choice, a key step is to average over a set of words. For the gender dimension, I considered the set of words most similar to "man" and "woman" using cosine distance, and chose appropriate word pairs. I performed the same operation for race/ethnicity, focusing on "black" and "white". I do caution, though, that the race dimension may be less accurate, because descriptions of race are not coded in speech to the same extent that gender is. For the race dimension, only "black" "african-american" "white" and "caucasian" could be used.

In order to measure the competency stereotype, I chose four appropriate word pairs for how customers might describe the competency of employees – "competent/incompetent," "efficient/inefficient," "helpful/unhelpful," and "fast/slow." For each word, I used cosine distance to search for the 20 most similar words, and average them together. Table C-1 below shows the 20 most similar words for each pair in the embeddings developed based on a subsample of 2 million Yelp reviews.

Table C-1 About Here

After identifying each word, I then project each word, along with each word in the vocabulary (the 10,000 most frequently used words in the corpus of text reviews), and identified where they fell on the gender and race axes. For ease of analysis, I present the Z-scores, where a score of 1 means that the concept is 1 standard deviation more feminine than the average word in the vocabulary. In the main text, I present this analysis using unigrams (one word) for a subsample of 2 million Yelp reviews. Embeddings based on the entire 7 million review dataset were too computationally intensive to collect, and multiple samples of 2 million embeddings replicated the same results. In sensitivity analyses presented here, I also perform the same sets of procedures using bigrams (a combination of two words, such as "slow_worker" when this combination appears frequently in the corpus). The figure below shows that results follow the same pattern.

Figure C-1 About Here

I also conduct the same analysis using embeddings constructed only with Yelp reviews of the 17 companies included in the sample (N=150,346). These models present a tradeoff, since on the one hand they represent only reviews of the companies in question, but the smaller number means that the word embeddings may be less accurate. The patterns here are generally similar, with each positive pair coming up as about 1 standard deviation more masculine than it's female counterpart, and between .2 and .6 standard deviations more White.

Figure C-2 About Here

It should be noted that the race dimension in the smaller sample of Yelp reviews may be less reliable, since the dimension of race is based on only a few words and not hard coded into speech in the same way gender is. Therefore, usage in the smaller dataset may be less common.

REFERENCES

Bergkvist, Lars, and John R. Rossiter. 2007. "The Predictive Validity of Multiple-Item versus Single-Item Measures of the Same Constructs." *Journal of Marketing Research* 44(2):175–84. doi: 10.1509/jmkr.44.2.175.

Boutyline, Andrei, and Ethan Johnston. 2023. "Forging Better Axes: Evaluating and Improving the Measurement of Semantic Dimensions in Word Embeddings."

Fisher, Gwenith G., Russell A. Matthews, and Alyssa Mitchell Gibbons. 2016. "Developing and Investigating the Use of Single-Item Measures in Organizational Research." *Journal of Occupational Health Psychology* 21(1):3–23. doi: 10.1037/a0039139.

Gabriel, Allison S., Nathan P. Podsakoff, Daniel J. Beal, Brent A. Scott, Sabine Sonnentag, John P. Trougakos, and Marcus M. Butts. 2019. "Experience Sampling Methods: A Discussion of Critical Trends and Considerations for Scholarly Advancement." *Organizational Research Methods* 22(4):969–1006. doi: 10.1177/1094428118802626.

Goel, Sharad, Adam Obeng, and David Rothschild. 2015. *Non-Representative Surveys: Fast, Cheap, and Mostly Accurate*. *Working Paper*. Stanford University.

Keeter, Scott, Nick Hatley, Courtney Kennedy, and Arnold Lau. 2017. "What Low Response Rates Mean for Telephone Surveys." *Pew Research Center Methods*. Retrieved July 1, 2021 (https://www.pewresearch.org/methods/2017/05/15/what-low-response-rates-mean-for-telephone-surveys/).

Kozlowski, Austin C., Matt Taddy, and James A. Evans. 2019. "The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings." *American Sociological Review* 84(5):905–49. doi: 10.1177/0003122419877135.

Matthews, Russell A., Laura Pineault, and Yeong-Hyun Hong. 2022. "Normalizing the Use of Single-Item Measures: Validation of the Single-Item Compendium for Organizational Psychology." *Journal of Business and Psychology* 37(4):639–73. doi: 10.1007/s10869-022-09813-3.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. "Efficient Estimation of Word Representations in Vector Space."

Schneider, Daniel, and Kristen Harknett. 2019. "What's to Like? Facebook as a Tool for Survey Data Collection." *Sociological Methods & Research* 51(1):108–40. doi: 10.1177/0049124119882477.

Storer, Adam. 2022. "Serving Like an Organization: How Foodservice and Retail Workers Interpret Their Interactions With Customers." *Work and Occupations* Advanced online publication:07308884221134582. doi: 10.1177/07308884221134582.

Wang, Mo, Hui Liao, Yujie Zhan, and Junqi Shi. 2011. "Daily Customer Mistreatment and Employee Sabotage Against Customers:Examining Emotion and Resource Perspectives." *Academy of Management Journal* 54(2):312–34. doi: 10.5465/amj.2011.60263093.

Wilson, Nicole L., and Camilla M. Holmvall. 2013. "The Development and Validation of the Incivility from Customers Scale." *Journal of Occupational Health Psychology* 18(3):310–26. doi: 10.1037/a0032753.

Yao, Jingxian, Sandy Lim, Cathy Yang Guo, Amy Y. Ou, and Jomel Wei Xuan Ng. 2022. "Experienced Incivility in the Workplace: A Meta-Analytical Review of Its Construct Validity and Nomological Network." *Journal of Applied Psychology* 107(2):193–220. doi: 10.1037/apl0000870.

TABLES AND FIGURES

Table A-1 Correlation of Customer Mistreatment Questions

		1	2	3	4
1	Customers Raise Voice	1			
2	Customers Act Like They're Better Than You	0.617***	1		
3	Customers Are Rude	0.683***	0.614***	1	
4	Customers Ask To Speak To a Manager	0.664***	0.452***	0.635***	1

Table B1 - Correlation Matrix

	1	2	3	4	5	6	7	8	9	10
1 Understaffing	1									
2 Inadequate Training	0.201***	1								
3 Misaligned Metrics	0.106***	0.104***	1							
4 Customers Raise Voice	0.198***	0.227***	0.182***	1						
5 Customers Sexually Harass	0.118***	0.168***	0.0909**	0.491***	1					
6 Customers Use Racial Slurs	0.107***	0.165***	0.0402	0.446***	0.568***	1				
7 Tenure	0.105***	0.147***	0.0467	0.0299	0.000653	0.0863**	1			
8 Education	-0.0175	0.0142	0.0653*	0.0686*	0.0476	0.0693*	0.105***	1		
9 Enrolled	0.0337	-0.0440	0.0667*	0.133***	0.0910**	-0.00399	-0.252***	0.0246	1	
10 ESL	-0.0223	-0.0633*	-0.0348	0.0520	0.0949***	0.130***	0.0123	0.0191	0.0716*	1
11 Kids	-0.0488	0.00875	-0.0867**	-0.147***	-0.116***	0.0206	0.208***	-0.0269	-0.300***	0.0357
12 Cohabitating	0.0325	-0.0279	-0.00207	0.0419	0.0563*	0.0670*	0.0550*	0.00510	-0.128***	0.0164
13 Cisgender Male	-0.0578*	0.0235	0.0161	-0.0456	0.0158	0.0674*	0.115***	0.107***	-0.0813**	0.0825**
14 White	-0.0356	-0.0227	-0.00355	-0.0523	-0.000366	-0.0814**	-0.0664*	-0.0274	-0.00385	-0.316***
15 Age	-0.00621	0.0799**	-0.0192	-0.225***	-0.213***	-0.0491	0.398***	0.0589*	-0.445***	-0.0508
16 Full-Time	0.0653*	-0.00113	0.0548	0.129***	0.100***	0.151***	0.273***	0.0307	-0.174***	0.0618*
17 Former Employee	-0.101***	-0.117***	-0.0283	-0.0854**	-0.0231	-0.0758**	-0.0791**	0.0348	0.0253	-0.0121
18 Manager	-0.0340	0.0488	-0.0177	-0.116***	-0.162***	-0.170***	-0.183***	-0.127***	0.0624*	-0.0374
19 Hourly Wage	0.0293	0.0339	0.0855**	0.0926***	0.0884**	0.0993***	0.282***	0.207***	-0.0634*	0.0490
20 Usual Hours	0.0634*	0.00778	0.0526	0.105***	0.113***	0.176***	0.200***	0.0663*	-0.145***	0.0491
21 Hourly Variability	-0.0236	0.000982	0.0147	0.103***	0.0732**	-0.0312	-0.212***	0.0139	0.236***	-0.133***
22 Prop. Customers Male	-0.0161	-0.0305	0.0312	0.0422	0.0426	0.0575*	-0.0117	-0.0329	0.0418	0.0137
23 Prop. Customers White	0.0202	0.0573*	0.0170	-0.0299	-0.0536	-0.130***	-0.0303	0.0612*	0.0872**	-0.0771**
24 Prop. Coworkers Male	0.00902	0.0706*	0.0255	-0.0139	0.0353	0.111***	0.140***	0.0568*	-0.0728**	0.0401
25 Prop. Coworkers White	0.0404	0.0275	0.00917	0.00607	-0.00168	-0.126***	-0.0737**	0.0159	0.0159	-0.114***
26 Customer Class Background	0.0247	0.0205	0.103***	0.0696*	0.0681*	-0.0348	-0.0551*	-0.0247	0.158***	-0.0416
27 Freq. Customer Interactions	0.0867**	-0.0265	0.0692*	0.129***	-0.00688	-0.0479	-0.0967***	-0.0244	0.0440	-0.0349
28 Company Doing Well	-0.0354	-0.119***	0.0363	-0.0157	-0.0434	-0.0779**	0.0406	0.0811**	0.0319	-0.00145
29 Job Satisfaction	-0.211***	-0.459***	-0.147***	-0.314***	-0.279***	-0.251***	-0.0979***	0.0133	0.00885	-0.0410
30 Turnover Intentions	0.166***	0.261***	0.128***	0.262***	0.252***	0.243***	-0.0614*	0.0363	0.103***	0.0837**
31 Job Meaningfulness	-0.165***	-0.341***	-0.104***	-0.218***	-0.214***	-0.169***	-0.0922***	0.0422	-0.0398	0.0364
32 Self-Reported Effort	0.0215	-0.124***	-0.0249	-0.0716*	-0.131***	-0.0622*	-0.00551	0.0278	-0.0553*	-0.0313

Table B1 - Correlation Matrix (Continued)

	11	12	13	14	15	16	17	18	19	20
1 Understaffing										
2 Inadequate Training										
3 Misaligned Metrics										
4 Customers Raise Voice										
5 Customers Sexually Harass										
6 Customers Use Racial Slurs										
7 Tenure										
8 Education										
9 Enrolled										
10 ESL										
11 Kids	1									
12 Cohabitating	0.269***	1								
13 Cisgender Male	-0.0454	0.0342	1							
14 White	-0.0293	0.00124	-0.0676*	1						
15 Age	0.528***	-0.00705	0.0730**	-0.0367	1					
16 Full-Time	0.134***	0.111***	0.144***	-0.0335	0.0525	1				
17 Former Employee	-0.0156	0.00110	0.0114	0.0543	-0.0831**	-0.0112	1			
18 Manager	-0.0956***	-0.144***	-0.0267	-0.00372	0.0190	-0.421***	0.0259	1		
19 Hourly Wage	0.00933	0.0860**	0.122***	-0.0458	-0.00866	0.328***	0.0614*	-0.361***	1	
20 Usual Hours	0.116***	0.0932***	0.184***	-0.0595*	0.0597*	0.667***	-0.0740**	-0.434***	0.352***	1
21 Hourly Variability	-0.183***	-0.0617*	-0.131***	0.0989***	-0.252***	-0.402***	0.0517	0.157***	-0.132***	-0.424***
22 Prop. Customers Male	0.0519	-0.0308	0.0138	-0.0690*	0.0000274	0.0412	0.0302	0.0720*	0.0621*	0.0288
23 Prop. Customers White	-0.0828**	-0.126***	-0.0301	0.0702*	-0.0274	-0.0516	-0.0495	0.0620*	-0.0111	-0.0250
24 Prop. Coworkers Male	0.0891**	-0.0137	0.213***	-0.0314	0.0899**	0.132***	-0.0137	0.0535	0.212***	0.183***
25 Prop. Coworkers White	-0.00664	0.00332	-0.0766**	0.223***	-0.0694*	-0.0318	0.0226	-0.0238	-0.0543	-0.0225
26 Customer Class Background	-0.190***	-0.137***	-0.0680*	0.0252	-0.220***	-0.106***	0.0267	0.101***	-0.0894**	-0.133***
27 Freq. Customer Interactions	-0.0427	-0.0284	-0.252***	-0.0274	-0.0323	-0.0626*	-0.0187	0.000523	-0.0905**	-0.0631*
28 Company Doing Well	-0.0418	-0.0658*	0.0347	-0.0142	0.0157	0.0258	0.100***	0.0121	0.111***	0.00852
29 Job Satisfaction	0.0292	0.00568	-0.00209	-0.0102	0.0383	-0.0455	0.231***	0.0389	0.00462	-0.0631*
30 Turnover Intentions	-0.125***	-0.0116	-0.00671	-0.0238	-0.226***	0.0106	-0.198***	-0.0417	-0.0464	0.0287
31 Job Meaningfulness	0.0586*	-0.00879	0.0105	-0.0175	0.0964***	-0.0313	0.128***	-0.0302	0.00448	-0.0503
32 Self-Reported Effort	0.141***	-0.00777	-0.102***	-0.0131	0.173***	0.0406	-0.0115	-0.106***	0.0518	0.0121

Table B1 - Correlation Matrix (Continued)

	21	22	23	24	25	26	27	28	29	30	31
1 Understaffing											
2 Inadequate Training											
3 Misaligned Metrics											
4 Customers Raise Voice											
5 Customers Sexually Harass											
6 Customers Use Racial Slurs											
7 Tenure											
8 Education											
9 Enrolled											
10 ESL											
11 Kids											
12 Cohabitating											
13 Cisgender Male											
14 White											
15 Age											
16 Full-Time											
17 Former Employee											
18 Manager											
19 Hourly Wage											
20 Usual Hours											
21 Hourly Variability	1										
22 Prop. Customers Male	-0.0407	1									
23 Prop. Customers White	0.0669*	0.0844**	1								
24 Prop. Coworkers Male	-0.139***	0.242***	-0.0531	1							
25 Prop. Coworkers White	0.0907**	-0.0286	0.488***	-0.107***	1						
26 Customer Class Background	0.131***	-0.0227	0.163***	-0.0169	0.0853**	1					
27 Freq. Customer Interactions	0.0900**	0.0367	0.0200	-0.194***	0.0606*	0.0694*	1				
28 Company Doing Well	0.0412	0.102***	0.0541	0.0368	0.0557*	0.0993***	0.0933***	1			
29 Job Satisfaction	-0.0722**	0.0371	-0.0141	-0.0342	-0.00445	-0.0827**	0.0279	0.152***	1		
30 Turnover Intentions	0.141***	0.00467	0.00216	0.00172	0.0200	0.115***	0.0187	-0.138***	-0.591***	1	
31 Job Meaningfulness	-0.0311	0.0356	-0.0128	-0.0534	-0.000658	-0.0764**	0.0354	0.141***	0.596***	-0.358***	1
32 Self-Reported Effort	-0.0639*	0.0176	-0.0205	-0.0771**	-0.0362	-0.0703*	0.158***	0.140***	0.246***	-0.149***	0.365***

Table B-2: Ordinal Logistic Regression Sensitivity Analysis

	Customers Yell		Customers Sexist		Customers Racist	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
	b/se	b/se	b/se	b/se	b/se	b/se
Understaffing	0.200***	0.169***	0.146**	0.116*	0.125***	0.096*
	(0.03)	(0.03)	(0.05)	(0.05)	(0.04)	(0.04)
Inadequate Training	0.321***	0.206***	0.300***	0.189***	0.242***	0.140**
	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
Misaligned Metrics	0.203***	0.171***	0.085	0.053	0.035	-0.002
	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)
Prop. Customers Male	0.101	0.116	0.073	0.076	0.219*	0.220*
	(0.09)	(0.09)	(0.09)	(0.09)	(0.11)	(0.11)
Prop. Customers White	-0.125	-0.118	-0.130	-0.111	-0.161*	-0.137
	(0.06)	(0.06)	(0.07)	(0.08)	(0.08)	(0.08)
Prop. Coworkers Male	0.024	0.016	0.104	0.091	0.173*	0.166*
	(0.08)	(0.08)	(0.09)	(0.08)	(0.08)	(0.08)
Prop. Coworkers White	0.034	0.046	0.014	0.018	-0.152*	-0.158*
	(0.04)	(0.05)	(0.05)	(0.05)	(0.08)	(0.08)
Customer Class Background	0.101	0.052	0.115	0.066	0.046	-0.004
	(0.09)	(0.09)	(0.09)	(0.10)	(0.07)	(0.07)
Freq. Customer Interactions	0.256*	0.278*	-0.024	0.001	0.005	0.008
	(0.13)	(0.13)	(0.12)	(0.13)	(0.12)	(0.13)
Company Doing Well	-0.013	0.012	-0.063	-0.027	-0.093*	-0.062
	(0.04)	(0.04)	(0.05)	(0.05)	(0.04)	(0.03)
ESL	0.368*	0.252	0.596**	0.466*	0.615**	0.471**
	(0.17)	(0.15)	(0.19)	(0.21)	(0.19)	(0.17)
White	-0.192	-0.256	0.177	0.122	-0.155	-0.194
	(0.18)	(0.17)	(0.18)	(0.17)	(0.19)	(0.17)
Cis-male	-0.200	-0.186	-0.074	-0.091	0.114	0.112
	(0.14)	(0.16)	(0.23)	(0.24)	(0.12)	(0.12)
Kids	-0.217	-0.229	-0.244	-0.223	0.003	0.019
	(0.14)	(0.15)	(0.15)	(0.15)	(0.15)	(0.16)
Cohabitating	0.257	0.264*	0.299*	0.297*	0.139	0.149
	(0.14)	(0.13)	(0.12)	(0.12)	(0.12)	(0.12)
Age	-0.021***	-0.018**	-0.022***	-0.018**	-0.010	-0.005
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Hourly Wage	-0.006	-0.004	-0.015	-0.011	-0.013	-0.008
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
Usual Hours	0.008	0.006	0.012	0.009	0.015*	0.012
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Job Satisfaction		-0.197***		-0.145**		-0.104
		(0.05)		(0.04)		(0.06)
Turnover Intentions		0.019		0.053		0.112*
		(0.03)		(0.04)		(0.05)
Job Meaningfulness		-0.048		-0.047		-0.027
		(0.05)		(0.05)		(0.06)
Self-Reported Effort		0.008		-0.072		-0.029
		(0.06)		(0.08)		(0.05)
Cut1	1.167	-0.162	0.376	-0.919	0.772	0.162
	(0.82)	(0.88)	(0.98)	(1.03)	(1.02)	(1.02)
Cut2	3.381***	2.091*	1.785	0.523	2.473*	1.908
	(0.82)	(0.88)	(0.99)	(1.03)	(1.04)	(1.04)
Cut3	4.248***	2.976***	2.492*	1.244	3.225**	2.675*
	(0.81)	(0.86)	(1.02)	(1.05)	(1.06)	(1.05)
Cut4	5.467***	4.227***	3.872***	2.644*	4.265***	3.726***
	(0.82)	(0.86)	(1.06)	(1.09)	(1.12)	(1.08)
N	1271	1271	1271	1271	1271	1271
Ind. Controls	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Company Controls	Yes	Yes	Yes	Yes	Yes	Yes
Halo Controls	No	Yes	No	Yes	No	Yes

Standard Errors Clustered by Company; Multiple Imputation for Item Non-Respose; Only selected variables from each control group displayed in each model

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C-1 Nearest Words for Core Concepts

	competent	incompetent	efficient	inefficient	fast	slow	helpful	unhelpful
1	skilled	inept	courteous	unorganized	quick	sloooow	accommodating	rude
2	professional	uncaring	polite	disorganized	speedy	slower	courteous	unfriendly
3	proficient	unprofessional	prompt	incompetent	prompt	inattentive	informative	dismissive
4	competant	ignorant	professional	unprofessional	quickly	sloooooow	polite	unaccommodating
5	customer-focused	unhelpful	expedient	undermanned	efficient	spotty	personable	condescending
6	efficient	arrogant	personable	ineffective	quicky	slooooow	knowledgeable	unprofessional
7	intelligent	unintelligent	punctual	mismanaged	superfast	sloooowwww	knowledgable	unresponsive
8	trustworthy	dishonest	effecient	inept	super-fast	glacially	accomodating	discourteous
9	courteous	uneducated	attentive	unhelpful	faster	slooooooow	welcoming	uncaring
10	well-trained	inconsiderate	speedy	inefficiently	lightning-fast	slowish	hospitable	disrespectful
11	personable	negligent	efficent	unfriendly	slowwwww	s-l-o-w	attentive	uncommunicative
12	caring	uncompassionate	curteous	unreliable	food-ish	slooow	helpfull	unsympathetic
13	knowledgeable	rude	competent	undependable	timely	understaffed	accomdating	unapologetic
14	efficent	unresponsive	curtious	unacceptably	sloooooow	laggy	non-pushy	uninformative
15	effecient	untrained	expeditious	dysfunctional	slooooooow	slowwww	knowledgeble	argumentative
16	polite	unqualified	helpful	unaccommodating	quicker	slowww	efficient	incompetent
17	customer-oriented	inefficient	mannerly	unresponsive	foodish	unattentive	super-helpful	arrogant
18	skillful	unmotivated	proficient	uncaring	semi-fast	forgetful	knowlegable	unwelcoming
19	compassionate	untrustworthy	competant	discourteous	fastest	slowwwwww	courtious	unknowledgeable
20	mannerly	disrespectful	attentitive	unmotivated	swift	non-attentive	curteous	impolite

Figure C-1 Yelp 2 Million Subsample Bigram Embeddings

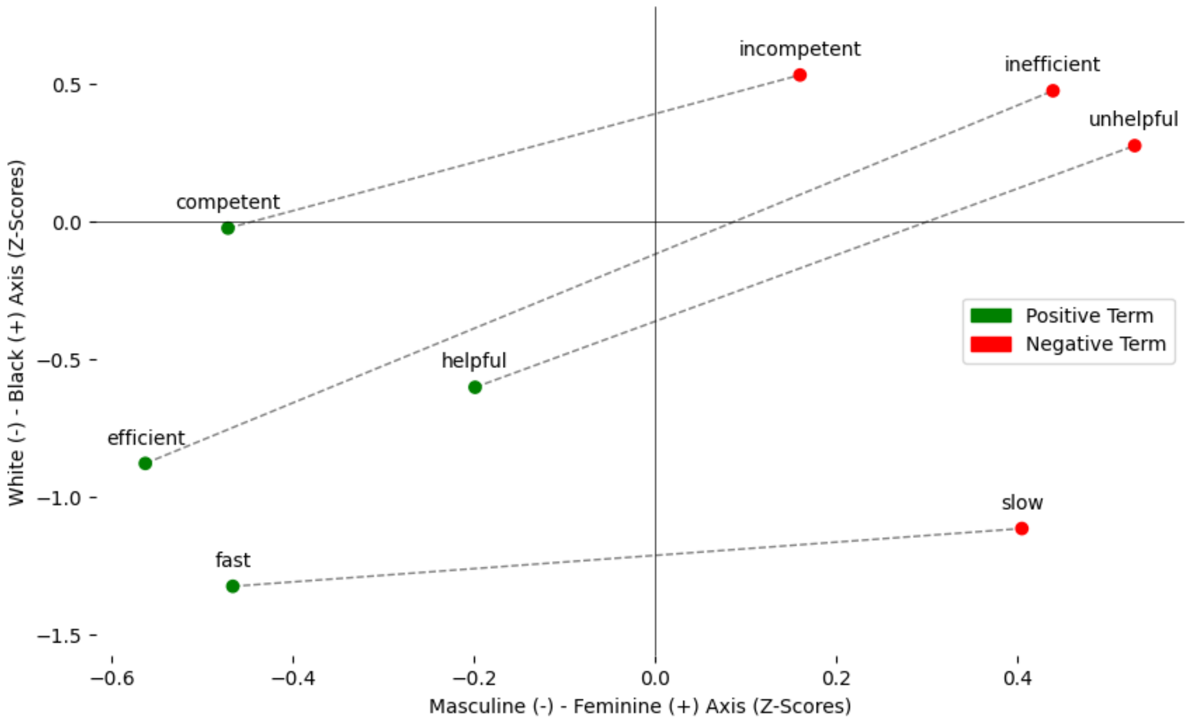


Figure C-2 Yelp Company-Specific Unigram Embeddings