Customer Discrimination in the Workplace: Evidence from Online Sales*

Erin Kelley,[†] Gregory Lane,[‡] Matthew Pecenco,[§] & Edward Rubin[¶]

August 21, 2024

Abstract

Many workers are evaluated on their ability to engage with customers. We measure the impact of gender-based customer discrimination on the productivity of online sales agents in sub-Saharan Africa. Using a novel framework that randomly varies the gender of names presented to customers without changing worker behavior, we find the assignment of a female-sounding name leads to 50 percent fewer purchases. Customers also lag in responding, are less expressive, and avoid discussing purchases. We show similar results for customers around the world and across workers. Removing customer bias, we find women would be more productive than their male coworkers.

*We are grateful to Fiona Burlig, Jesse Bruhn, Carlos Schmidt-Padilla, Reshmaan Hussam, Katy Bergstrom, Florence Kondylis, Jeremy Magruder, and Pascaline Dupas, as well as participants at various conferences and seminars for their helpful comments and suggestions. Funding for this project was graciously provided by PEDL. We also thank Borui Sun, Victoria Yin, Edwin Kasila, and Steven Wandera for their excellent research assistance. AEA RCT identification number: 0006698. This project received IRB approval from the University of Oregon (#IRB-08132018.010).

[†]Harris School of Public Policy, University of Chicago erinmkelley@uchicago.edu

[‡]Harris School of Public Policy, University of Chicago, NBER laneg@uchicago.edu

[§]Brown University matthew_pecenco@brown.edu

[¶]University of Oregon edwardr@uoregon.edu

1 Introduction

Extensive research in the social sciences has sought to measure the sources and extent of labor market discrimination. Tests of discrimination typically try to identify cases of direct discrimination (Hull et al., 2023) by examining whether there are differences in wages or hiring outcomes among individuals who exhibit equal levels of productivity (Parsons et al., 2011). A drawback to this approach is that it does not consider potential indirect forms of discrimination on the job that directly affect worker productivity (Glover et al., 2017; Sarsons, 2022; Egan et al., 2022). These indirect forms of discrimination can cause typical tests to underestimate the extent of discrimination.

Customers may be a significant source of indirect discrimination. Many workers are evaluated on their ability to engage with customers. If customers have explicit or implicit biases against workers of a certain gender or race, then workers of this identity will perform less well. While each instance of customer bias may be unremarkable or unobserved, the accumulation may take a toll.

Quantifying the magnitude of discriminatory customer behavior is required for understanding how an important source of discrimination shapes workers' labor market outcomes (Becker, 1957) and the extent to which standard tests fail to capture the full scope of inequity in the workplace. Crucially, the magnitude is an important input for firms that may want to promote the most talented workers rather than those least subject to bias. If customer behavior directly impacts worker productivity, employers in certain jurisdictions may also face regulatory liability for using this metric in deciding who to promote and how much to pay.[1] Finally, it is also relevant for regulators who may wish to identify whether policies like performance-based pay promote discriminatory outcomes.

This paper provides new evidence on the extent of customer discrimination by running a randomized field experiment on customer discrimination for workers within a firm. We answer this question in an actual labor market by partnering with an online travel agency whose offices are scattered across sub-Saharan Africa. The company sells flights and hotels and hires local sales agents to assist their customers. Specifically, we study over 2,000 customers from 70 countries

[1] For example, Title VII's Bona Fide Occupational Qualification in the United States makes exceptions in a very narrow set of circumstances.

(87% from Africa, 13% abroad) as they chat with online sales agents who answer their questions and help them make purchases. This context allows us to measure worker productivity through sales records and to document rich patterns of customer engagement, including bargaining and harassment, through chat transcripts.

We develop a framework for estimating the causal effect of customer-based discrimination. First, the names of workers—and implied genders—were randomized daily, providing plausible variation in customer beliefs about the gender of the agent with whom they were chatting.[2] Customers could only infer agents' gender from names as customers did not receive any other information about the agent. Second, workers were unaware of their assigned name due to a web plugin that masked the assigned name from their view. This step ensured that agents' behavior was not directly affected by their name assignment. Consequently, any change in consumer behavior towards sales agents could only occur if consumers responded to the randomly assigned names.

This research design overcomes challenges with two common experimental methods to study discrimination: audit studies and correspondence studies. Audit studies—in which actors, who are as similar as possible except on one dimension, engage in a task like applying to the same job—struggle to control for all other differences between the actors. Actors may also be subject to "demand effects" because they are aware of their treatment status and do not have real incentives to perform well. Correspondence studies—in which fictitious applications with different sounding names are sent to a possible discriminator like employers—can only measure indirect outcomes such as job application callback rates rather than actual job hiring (Bertrand and Duflo, 2017). In our setting, the daily name randomization eliminates omitted variable concerns, and the name masking alleviates concerns about demand effects. Furthermore, by studying real workplace interactions we can collect the ultimate outcome measures of interest—e.g., the likelihood of purchase—and specifics of the interaction—e.g., customer engagement with the sales agent.

The results are striking; we find that randomly assigned female names reduce the likelihood that customers make any purchase, the number of purchases, and the value of the purchases. Specifically, the likelihood of any purchase decreases by 3.8 percentage points, or 50% relative to the baseline purchase rate (7.6%). We observe similarly large reductions in the total number

[2] Changing the names of workers appears to be relatively common in online sales settings (e.g., LiveAgent).

of purchases, the total value of purchases, and the average purchase price conditional on any purchase. To confirm that our treatment effects result from the implied gender of agents' assigned names, we show that customers are aware of agents' names as they are frequently mentioned in conversation. We further show that these results are not driven by our choice of regression specification.

We consider three mechanisms why customers might discriminate against female workers, such as general customer disinterest, customers bargaining differently, or customers being overtly negative. Data from the evolution of agent-customer chat interactions suggest that customer disinterest in speaking with a female agent is the most likely discrimination source. Initially, we find that consumers respond more slowly to female agents—they only respond after receiving additional messages from the agent. Conditional on ever responding, customers are less expressive to workers with female names and are less likely to transition from initial topics of conversation (e.g. an inquiry about a specific hotel on a specific date) to discussions about purchases.

Additionally, the data do not support other possible mechanisms. We find no evidence that consumers differentially bargain when agents receive female-sounding names and no differences in hostile or harassing behavior, although any form of harassment is rare in this context. Together, the results appear to be most consistent with literature in sociology that highlights the importance of subtle and often unconscious gender bias in the workplace (Basford et al., 2014; DeSouza et al., 2017; Nordell, 2021) rather than more overt forms of discrimination such as bargaining (Castillo et al., 2013; Card et al., 2016; Rousille, 2021) or harassment (Georgieva, 2018; Folke and Rickne, 2022; Dupas et al., 2021).

What are the implications of these results for the labor market? While field experiments studying discrimination can measure the magnitude of discrimination by firms or customers, these studies are commonly criticized for not being able to assess whether this discrimination ultimately matters for workers (Heckman, 1998; Charles and Guryan, 2011). The seminal work of Becker (1957) shows that workers may sort away from discriminating firms or customers in equilibrium, and hence what matters is not whether some discrimination exists in the market but whether it is actually experienced in equilibrium. A unique feature of our experiment is that we can quantify the amount of discrimination experienced by employed workers. If workers could avoid discrimination, we should not see any women in these jobs that feature substantial customer discrimination. Surprisingly, the opposite is true. The vast majority of workers in this role are female, indicating important amounts of discrimination faced in equilibrium.

Why would the firm hire female workers who are being subjected to customer bias, and why would workers choose to stay? Comparing male and female agents in a non-experimental sample, we find they are equally productive on average (inclusive of customer bias), and hence the firm is not losing out by hiring female workers at similar wages to men. These results are consistent with the market outcome of a non-discriminatory firm that has internalized discriminatory customer preferences. It is only by running our experiment that we observe that female employees are more productive than their male counterparts absent customer bias. Female workers may choose to stay in these jobs for two reasons. First, if other occupations exhibit worse discrimination, they may be constrained in their options. Second, female workers may have a comparative advantage or derive non-pecuniary benefits from this occupation. Either way, our experiment points to important constraints that prevent workers from avoiding discrimination.

We view this paper as a proof of concept that customer discrimination represents an important source of indirect discrimination that can meaningfully affect worker productivity in ways that often go unnoticed by firms and econometricians alike. However, as a result of conducting our experiment in a real-world setting, the results may be influenced by the type of job, the customers, or the workers that comprise our sample. Nevertheless, we view our results as being more broadly representative for the following reasons. First, the customer sales jobs we study are typical in the service industry: the workers engage in straightforward, non-technical sales and customer assistance for an established firm. Second, we find similar rates of gender-based discrimination from customers in different parts of the world, both inside and outside Africa, suggesting common effects across different customer bases. Finally, we provide evidence that our results do not depend on worker characteristics: the effect for each worker is negative, and we cannot reject equality across workers.

While there are many benefits to conducting this real-world experiment with a large company, there are also some limitations. Only six female agents enrolled in the experiment and no male agents, which limits our ability to test whether male agents exhibit some trait that would change the impacts of customer behavior. An additional limitation is that we can only imperfectly show: similar effects on purchase values as purchase quantity, no large differences by customer gender, and that customers likely respond to worker gender rather than characteristics correlated with gender.

This paper makes three contributions. First, our results contribute to a growing literature that emphasizes the importance of studying indirect forms of discrimination. Hull et al. (2023) provides an analytical framework to describe indirect discrimination, specifically how discriminatory outcomes do not only arise from direct interactions but also from responses to past or expected future discrimination by other actors. Failing to account for these sources of discrimination could lead us to underappreciate the importance of discrimination in the market. In this paper, we find that discrimination by customers leads the firm to retain only the women who are more productive than their male counterparts after accounting for bias.[3] While the firm may not be aware of customer bias, the ultimate outcome is discriminatory because women have to be more productive to earn the same wage. This is consistent with Glover et al. (2017), which finds that manager bias makes black workers less productive on the job, and this discrimination ultimately causes the firm to hire relatively more productive black workers. Customer discrimination is likely a more difficult problem for a firm to solve as this outcome does not occur as a result of internal firm policies but rather external (customer) forces. Related work shows how discrimination can directly impact performance evaluations and how workers may endogenously react to this discrimination (Parsons et al., 2011; Hengel, 2022).[4]

Second, we provide causal evidence that customer discrimination lowers the measured productivity of female employees in the workplace by a meaningful margin and in ways they cannot avoid. While there is a growing body of work that finds evidence of at least some customer discrimination by directly studying customers (Nardinelli and Simon, 1990; Leonard et al., 2010; Bar and Zussman, 2017; Caselli and Falco, 2020) or by indirectly studying employers exposed to different customers (Kahn and Sherer, 1988; Holzer and Ihlanfeldt, 1998; Combes et al., 2016; Hurst et al., 2021; Kline et al., 2022), we are the first paper, to the best of our knowledge, that uses a randomized control trial to isolate customer discrimination rather than methodologies that may not fully control for differences in unobservable characteristics of workers and customers. This

[3] This is consistent with the Hull et al. (2023) definition of future-in-present discrimination.

[4] This paper also contributes to literatures on bias in student evaluations of teachers (MacNell et al., 2015; Mengel et al., 2019), by audience members for academic economists (Dupas et al., 2021), and by audiences for professional athletes (Caselli et al., 2023). Recent work has also shown the importance of patient racial preferences for doctor race (Alsan et al., 2019).

paper is also unique for studying gender. Moreover, we build on the existing literature by measuring worker productivity rather than relying on wages or hiring outcomes. This provides new insights on how customer discrimination affects workers and on how tests of discrimination may be biased by assuming fixed worker productivity.[5]

Relatedly, this paper also contributes to research that studies how customer discrimination affects goods sellers in marketplaces or product markets through audit or correspondence-type studies (List, 2004; Doleac and Stein, 2013; Ayres et al., 2015; Kricheli-Katz and Regev, 2016). While this literature has found more evidence of discrimination, the implications for how these results translate to workers and the labor market is unclear. Businesses or sellers may opt to offer products in industries where their customer base is likely to respond positively based on their recognized identity. Whether the discrimination faced by these fictional sellers would endure in the real labor market is an open question as real-world sellers might proactively choose industries where they believe they can overcome customer bias or find another form of comparative advantage. To the best of our knowledge, we are the first experimental study to enlist a real company with actual employees and customers to understand the impact of customer discrimination. This framework allows us to study a labor market setting among individuals who have selected into this job based on their comparative advantage and test how this discrimination relates to labor market sorting.

Finally, this paper contributes to a large literature investigating explanations for differences in earnings and employment between men and women in the labor market. Classic decompositions of gender earnings gaps split explanations into productivity-based factors (e.g., human capital, experience), discrimination, labor demand, and other drivers (Blau and Kahn, 2017; Caliendo et al., 2017; Gallen et al., 2017; Sin et al., 2020). Our results demonstrate that underlying customer-based discrimination can contribute to observed productivity differences, which would bias these decompositions.[6] Our work also complements important work investigating the effects of discrimination conditional on equal worker productivity (Goldin and Rouse, 2000; Sarsons, 2022). In contrast to these papers, our results imply that policies that seek to equalize outcomes for men

[5] Nardinelli and Simon (1990) is an exception. They define the productivity of baseball players to be their entertainment value, observed through the price of baseball cards.

[6] Blau and Kahn (2017) note that decompositions of discrimination may be understated if "some of the explanatory variables such as experience, occupation, industry, or union status have themselves been influenced by discrimination—either directly through the discriminatory actions of employers, coworkers, or customers, or indirectly through feedback effects."

and women of the same productivity (e.g. blind auditions as in Goldin and Rouse (2000), or equal pay for equal work policies) will not eliminate the gender gap and create a level playing field.[7]

Identifying the role of customer discrimination is relevant for any country with a booming service sector where customer-facing roles abound. As the share of women in the service industry continues to increase exponentially in sub-Saharan Africa, finding ways to prevent discrimination in customer-facing roles is particularly important. It is also particularly relevant in this context because barriers to female labor force participation are meaningful.[8] The gender inequality index is high in sub-Saharan Africa (UNDP, 2022) and issues of gender inequity in the workplace have become a central policy goal for governments across the continent and international institutions alike (World Bank, 2011; O'Donnell et al., 2020). Policymakers must be aware of this source of discrimination when considering different labor market policies.

The paper is organized as follows: section 2 details the context, the company we work with, and the data, section 3 describes the empirical strategy, section 4 presents the results, section 5 discusses the labor market implications, section 6 addresses external validity, and section 7 concludes.

2 Context

2.1 Service sector in sub-Saharan Africa

We study consumers' discriminatory behavior when engaging with customer sales representatives. These workers engage with customers to answer queries and make sales over phone and chat. This is a common job profile in the service industry. Most major companies have sales de-

[7] By showing that objective performance measures may not be a good reflection of true productivity, we also join a growing literature on the subjectivity of performance assessments and its impact on the gender gap (Benson et al., 2021).

[8] A variety of other barriers for female employment in low-income countries have been explored. For literature on norms and bargaining dynamics within the household see (Dean and Jayachandran, 2019; Bursztyn et al., 2020; Heath and Tan, 2020; Field et al., 2021; Lowe and McKelway, 2021; McKelway, 2021a,b); workplace attributes (Subramanian, 2021), safety during commutes (Borker, 2021), market demand (Hardy and Kagy, 2020), discrimination in the workplace (Duflo, 2012; Jayachandran, 2015; Sin et al., 2020; Delecourt and Ng, 2021). Delecourt and Ng (2021) uses an audit-study approach to show that customers do not discriminate against female-led small business vegetable sellers in India. Our study differs in three ways. First, our research design allows us to overcome some of the limitations of audit studies by fully controlling for all agent characteristics. Second, our experiment uses real workers in jobs they will continue to have after the experiment is over, which factors in the selection processes for hiring workers and worker incentives. Third, our experiment sheds light on the implications of customer discrimination in the labor market as a whole.

partments with customer service agents who assist customers from all over the world. Customer service industry jobs are dominated by women.[9]

In sub-Saharan Africa where our research is conducted, the number of customer-facing roles are increasingly common across the continent as service sector jobs increase. For example, the share of working-age individuals employed in services throughout sub-Saharan Africa rose 12% from 2011 to 2019 (World Bank, 2022). Women largely drove these trends: the share of working-age women employed in services increased by 16% over the same period—currently at 39.7%.

As internet connectivity spreads across the continent and service-sector jobs increasingly interface with customers online, we are likely to see more women in customer service jobs. In 2010, only 8.3% of the population in Africa had internet access. By 2017, internet access had increased to 22.3% (World Bank, 2022). Online shopping, in particular, has increased 18% annually between 2014 and 2017 (UNCTAD, 2018) and estimates suggest that almost 50% of digital buyers in Africa are female (Statista, 2019). The COVID-19 pandemic has likely accelerated these trends as consumers increasingly head online. Reflecting this growth and importance, in 2020 the value of African e-commerce was estimated at 20 billion USD—a 42% increase over 2019 (IFC, 2021).

2.2 Company and study details

One of the key features of this study is our collaboration with a real company. The ability to examine a real labor market is crucial as it enables us to determine whether discrimination endures in a real work environment and whether workers can proactively avoid it, as theory might suggest. Specifically, we work with an online travel agency with offices located across sub-Saharan Africa. We work primarily with their East African field office. While we cannot provide any identifiable information about the company for confidentiality reasons, we provide contextual information below about the location, the company, and the experiment.

Our results may be affected by the location and context where we operate, particularly the labor market conditions and social norms. Female labor force participation rates are relatively high (between 52-72%) in the countries where we conduct the research, and they are comparable to labor force participation rates for men (between 65-75%) (World Bank, 2022). Differences in the labor force participation rates between men and women (3-13%) are much smaller in our

[9] Zippia (2021) estimates that 70% of customer service industry jobs in the US are held by women.

context relative to other lower-middle income countries (which hover around 38% on average) (World Bank, 2024). Furthermore, based on the Global Gender Gap Index 2023 Report from the World Economic Forum, the countries where we operate rank within the top third in the economic participation and opportunity index for women (71-79%), compared to the global average of 59.8% across all countries in the sample (World Economic Forum, 2023)

While high rates of female labor force participation suggest a high degree of business integration between men and women, social norms tend to be relatively more conservative in the countries where we work. According to the UNDP's Gender Social Norm Index (GSNI), which is based on responses to seven questions from the World Values Survey about gender norms, the percentage of people with no bias against women is generally low, ranging from 1% to 4% in the countries in which we operate (with an average of 10% across the countries UNDP samples from). However, it is important to highlight that these norms are computed across the entire country, including rural areas where social norms are much more conservative (Yotebieng, 2021). In contrast, all our sales agents come from urban areas where it is common for women to work, run businesses, and regularly engage with men in both social and business contexts.

The company sells flights and hotels primarily for trips to different parts of the continent (the average price of a flight/hotel conditional on making a sale is 140 euros). Customers make purchases on the platform or engage with local sales agents to ask questions, make complaints, and assist with purchases (using phone calls and online chat interfaces). Customers initiated chats for several different primary reasons. The most common reason was the desire to make a general inquiry about the availability of flights or hotels matching the customer's travel requirements (47% of chats). However, other common objectives were to make a specific booking (21%), ask for the price on a specific flight or hotel (11%), or to cancel a booking (7%).

The company enlisted 6 agents who were working across two of their office locations in the experiment. All of the enlisted agents were female. While we would have liked to expand this study to additional firms, this form of experimentation within a firm typically requires extensive relationship building that isn't easily replicable to other firms. In the company overall, approximately two-thirds of the sales agents are women, and these women account for 83% of chats. This gender breakdown reflects broader industry trends. The sales agents are full-time company employees, and receive an annual wage.[10] Their job description resembles a common customer sales role insofar as sales representatives engage with customers and search for products in their company's catalogue to match customers' requests.

The company provides sales agents with a chat interface to interact with customers. Customers can initiate interactions with sales agents by clicking on a chat button at the bottom of the webpage. Clicking the chat button reveals a chat window displaying the agent's first name and automatically sends a short, formulaic greeting message from the agent to the customer. Thus, agents always send the first message; either the agent or the customer can send subsequent messages as the conversation evolves.

The company was keen to partner with the research team to investigate whether they could optimize this chat/sales interface. This particular test aimed to identify how customer behavior changed with respect to agents' identities—specifically when agents were assigned male- versus female-sounding names. To this end, the company needed to (1) randomize whether the name appearing in the chat implied a male or female identity and (2) ensure agents were unaware of their assigned names. The randomization was correctly implemented: a software program randomly pulled one name per sales agent per day from an existing list (with daily replacement) and assigned the randomized name to the agents' chat interface. A local field team compiled the list of names by drawing 1,198 names from local school yearbooks and assigned each name an implied gender, ethnicity, and whether it is a common English first name.[11] To limit the customers' inference of other dimensions of agents' identities besides gender, the interface only included agents' (randomly assigned) first names.

Next, to ensure that agents could not see the names that were assigned to them, a web plugin was designed to omit the agent's name from the agent-facing interface. The company installed the plugin on each agent's internet browser with oversight from our field team. The plugin symbol was removed from the list of visible extensions—appearing as a light grey square when all browser extensions were listed. The plugin worked in the following way. Consider a day when agent

[10] We do not have access to other information about agents' demographics or wages within the company.

[11] Of the 1,198 names, there are 1,196 unique full names, and 579 unique first names. Of these, 267 were female-sounding and 322 male-sounding. Table A1 lists 20 example names, by gender and whether they are English-sounding or not (59% of male names were non-English sounding, while 47% of female names were non-English sounding). We show the difference in name characteristics across gender does not affect our estimation of customer gender bias in section 4. We first discuss how we correlate these names with other characteristics in subsection 2.3 and discuss the results in subsection 4.3.

Mary (real name) was assigned the name *Steve*. Whenever the customer typed "Steve" into the chat, Mary would only see "Agent" in her chat window. In contrast, the customer would still see "Steve." This name masking included any references to the agent's assigned name in the chat transcript.[12]

The experiment was launched in January 2019 and the name assignment continued until October 2019. The experiment occurred in two offices, one in West Africa and one in East Africa. All agents who were consistently working in these offices at the launch time participated in the study, which included six agents in total. The company informed agents that it was interested in learning more about how customers respond to different agents and may change agents' display names in the chat. No further information about the nature or the objectives of the experiment was provided by the company to the agents, including the focus on gender. Agents did not ask any follow-up questions throughout the duration of the experiment, and the company made no additional requests of agents (in terms of protocols/procedures to follow). It is unlikely that knowledge of the experiment would have affected agents' behavior as the study was never discussed in any subsequent team meetings. Even if agents thought the experiment was about gender, it is unclear how this would affect their behavior as the name assignment changed daily without their knowledge.

2.3 Data

The analysis relies on two primary sources of administrative data. The first dataset records every purchase made by customers, including the sale amount. The second dataset contains the agent-customer chat interactions: the full chat transcript, a timestamp for each message, and the customer's country.[13] The sales data were matched to the chats using date and customers' IP address.

To measure overall purchases, we include purchases directly made by customers and purchases made by agents on behalf of customers. When customers purchase a product themselves, the sale is recorded in administrative sales records. When agents input customer details and pur-

[12] The vast majority of interactions occurred in English, limiting concerns about gendered identifiers. Other gendered identifiers like "Sir" or "Miss" are rarely used in 1.3% of chats, and we show the results are unaffected by excluding any days when this occurs.

[13] We only know customers' approximate locations—we do not have access to any customer demographic data. However, estimates from other sources suggest that in 2019 nearly 50% of digital buyers in Africa were female (Statista, 2019).

chase products on their behalf, we can only capture the sale by reading through the chat records and flagging instances when agents send final purchase confirmation details to customers. Customers then pay separately or at the time of receiving the order. When agents make purchases, we cannot measure purchase values. We, therefore, only measure the total value of purchases using the administrative sales records.[14]

From the chat transcripts, we create objective and subjective outcome measures.[15] Objective measures do not require human interpretation—for instance, whether a purchase occurred. Subjective measures represent outcomes that require human interpretation of the chat content and include an assessment of the primary and secondary purpose of the chat, the overall tone, whether customers bargained with agents, or whether customers harassed agents. Enumerators familiar with the cultural context hand-coded these subjective outcomes.[16] They were instructed to read through the entire chat transcript and identify the set of positive and negative words that were used, as well as the set of words that could reasonably be associated with bargaining (budgets/vouchers/discounts/cheaper) and harassment (comments on physical appearance/request for dates). From there, the enumerators identified whether the overall tone was neutral or non-neutral (including angry, sad, happy, ecstatic, impatient) and whether any bargaining or harassment occurred. 20% of the observations were double coded to ensure consistent measurement.[17] Enumerators did not know the gender assignment of the agent.

Agents' jobs involved several sales-related activities, including assisting customers via online chat and phone. Each agent worked the chat interface six weekdays per month on average. Table 1 shows that on days when agents responded to chats, they spent 2.6 hours on the online sales interface with customers, engaging in approximately 8 unique chat conversations per day. Additionally, the average chat lasted 22 minutes and contained 73 words. The sales agents did not all work during the full study period for institutional reasons, although they all worked a majority of the time.

[14] Our sales measure also includes the small set of chats (3%) that were transferred to the phone to complete a sale. There are no discernable patterns among the set of chats that are transferred: they are equally balanced across female and male-sounding agent names, and they do not appear to be transferred for one particular reason over another.

[15] The variables are further described in Table A2.

[16] We hand-coded these conversations to create clear and interpretable outcomes (e.g., chat purpose, bargaining) in our moderately-sized sample, rather than employing natural language processing models which are especially useful for large corpuses of text.

[17] Any discrepancies across these 20% were discussed and harmonized. Reading through the chats was a time-consuming process and budget considerations prevented double coding all observations

We restrict our sample in two ways. First, we limit our sample to normal working days to avoid time periods associated with testing and developing new chat and purchase features at the company. Second, we restrict our sample to observations with five or fewer previous purchases as some customers may have accessed the site using public areas or businesses with common IP addresses, which prevents us from linking a specific individual from that location to their purchase.[18]

3 Empirical strategy

The design of this study overcomes two major challenges to identifying the causal effect of customer-based gender discrimination. First, daily randomization of agents' names ensured customers were randomly exposed to female- or male-sounding names. This separates unobserved factors that correlate with gender from customers' perceptions of gender. Second, agents were not aware of the name consumers see—any revelation of the agent's name during the chat was masked automatically by a computer program and was not seen by the agent.[19] Therefore, agents' behavior cannot directly respond to the randomized name—only to customers' responses to these names. Together, these elements allow us to test for customer-based gender discrimination.

Treatment assignment occurred as follows. Agents were randomly assigned 'male' or 'female' each day (with daily replacement) with equal probability.[20] Given the selected gender, a specific name from the name database was randomly selected. This procedure occurred every day of the study period. The number of agents working varied daily. Some days, only one agent worked; other days, multiple agents operated the chat. Customers were allocated programmatically to agents; neither agents nor customers had any choice with respect to their matches.

Using this randomization, we estimate the effect of customer discrimination on worker productivity. Our main specifications take the following form:

$$y_{iadm} = \beta\, \mathbb{1}[\text{Assigned female}]_{adm} + \gamma_{am} + \varepsilon_{iadm}$$

[18] Only 2% of observations come from IP Addresses with five or more purchases.

[19] See subsection 2.2 for details.

[20] Due to variation in the number of customers arriving on a given day and sampling variability in treatment assignment, 47.6% of customer-agent interactions occurred with a female sounding name.

where y_{iadm} is the outcome of interest for customer i working with agent a on date d in month m. The indicator $\mathbb{1}[\text{Assigned female}]_{adm}$ is 1 if agent a (matched to customer i) is assigned female on that date d. The term γ_{am} represents agent by month-of-sample fixed effects. We augment this regression specification to estimate individual-agent treatment effects and heterogeneity across customer characteristics.[21]

The coefficient β identifies the *absolute* customer bias workers face in their job, which is the key parameter of interest when considering the impact of customer discrimination. We are limited in our ability to test for *relative* customer bias depending on customer characteristics which has been a focus of other papers (Leonard et al., 2010; Combes et al., 2016; Bar and Zussman, 2017). For the customer characteristics we do observe, these tests are in section 6.

Agents worked different times, in different locations, and some only worked part of the sample period, implying that agents themselves are not explicitly randomly assigned to customers. While not strictly necessary, since treatment assignment is uncorrelated with customer type, our main analysis restricts comparisons between similar customers using agent by month-of-sample fixed effects. Specifically, the research design compares (1) a consumer who chats with agent a in month m on a day when the agent was assigned female to (2) a consumer chatting with the same agent in the same month when the agent was assigned male. We show our main results are robust to a range of alternative model choices detailed in subsection 4.3.[22]

Customers may have multiple interactions with agents on the same day if they are disconnected or return to ask additional questions. We account for this possibility in two ways. First, we two-way cluster our standard errors at the agent-day (the level of randomization) and customer-day levels. Second, we assign the customer the treatment status of their first chat of the day. This circumvents the possibility that customers can affect their treatment status by returning to chat with an agent of a different gender.

We validate the randomization procedure on the analysis sample in Table 1. In column (3) of this table, we regress observable customer characteristics (e.g., number of past purchases), agent characteristics (e.g., number of daily chats), and environment characteristics (e.g., day of week,

[21] Note the regression can analogously be run at the (grouped) agent-day level after collapsing and reweighting to match our customer (microdata) approach (Angrist and Pischke, 2008). The focus of the paper on customer behavior and the parsimony of the current approach motivates the customer-level analysis.

[22] Including agent and month fixed effects separately does not necessarily produce clear comparisons because not all of the agents worked every month.

total agents across all offices working on this date) on an indicator for whether the agent received a female name, using our main specification.[23] Female assignment does not correlate with any customer, agent, or environmental characteristics at the 10% level. We fail to reject the joint null hypothesis that each of these effects is zero ($p = 0.61$). Table 1 includes an additional row that identifies whether the customer mentioned the agent's actual name. This event occurs very rarely (mean is <0.01), validating the name assignment procedure, and likely results from agents' names coincidentally matching something said in the chat.

4 Results

4.1 Main purchase outcome

The experiment aims to identify the impact of gender on consumer behavior. This strategy requires that consumers pay attention to agents' assigned names. We confirm that customers notice agents' names by measuring how often consumers use agents' assigned names in chats. This test provides a lower bound for consumers' awareness of agents' names—and likely the names' implied genders. In our study sample, customers use agents' assigned names in 11% of chats in which consumers ever respond to agents' initial messages. We interpret this as a relatively high share of customer awareness as many chats are brief and mentioning a person's name is unnecessary.[24] Thus, agent names are indeed salient in chat interactions and could affect customers' behavior.

Table 2 presents the estimated effects of female-name assignment on outcomes related to customer purchases. As our main outcomes, we measure purchases within 24 or 48 hours of the chat to capture behavior plausibly related to the chat interactions rather than unrelated interactions that happen later. We construct purchases in two ways: making any purchase and the number of distinct purchases.

We find that consumers assigned to agents with female names are less likely to purchase products on the website. Column (1) shows that female-name assignment decreases the probability

[23] The number of daily chats by an agent in a day could potentially be affected by name assignment if labor supply or hours worked changes. In practice, since customer allocation is done programmatically there seems to be little room for endogenous response along this margin or more simply, labor supply may be unaffected.

[24] Additionally, the first name of the worker is shown at the top of the chat box and is the only information provided to the customer about the worker.

that any purchase occurs (within 48 hours) by 3.8 percentage points ($p = 0.003$). The likelihood that a chat results in any purchase in the control group (male-sounding names) is only 7.6%.[25] Thus, the point estimate implies a 50% reduction in the likelihood of making a sale. Column (2) shows that consumers also purchase 0.038 fewer total products ($p = 0.005$) when interacting with female-sounding names. Columns (3-4) show very similar results for the same outcomes measured in a 24-hour window after the chat.[26]

The results do not vary based on the evaluation period. Figure 1 plots the effect of female name assignment on total number of sales for 12 hours post-chat, and each 24-hour period up to 7 days. After 24 hours, when most purchases are made, the estimates are quantitatively and qualitatively similar for the next six days.[27] Therefore, the treatment does not selectively delay purchases to a later time. Instead, it prevents purchases from happening altogether.

Female name assignment also translates into lost revenue. Table A3 shows customers assigned to agents with female names reduced the total value of their purchases by 60% (column 1), stemming from fewer purchases (discussed previously) and a 36% reduction in the average purchased price conditional on any purchase (column 2).[28] The reduction in average purchased price conditional on any purchase is another interesting potential indication of discrimination, although this result could be driven by either high-value purchasers being less likely to purchase or similar customers purchasing lower price products.

The closest literature to compare the magnitude of our estimates is the literature on customer discrimination for independent sellers in goods markets. This comparison is somewhat challenging. For example, List (2004) and Doleac and Stein (2013) find reductions of 3-30% and 11% of the price offered for the product to the minority seller, respectively, which is much smaller than the 51% reduction we observe in the probability a transaction occurs. The differences in outcomes,

[25] We calculate control group means accounting for agent-month fixed effect cells $c \in \mathscr{C}$ as $\sum_{c \in \mathscr{C}} (E[Y|C = c, D = 0) w_c$ for weights w_c, cell C, and treatment status D. We do so to correspond to the OLS estimand, $\beta^{OLS} = \sum_{c \in \mathscr{C}} (E[Y|C = c, D = 1] - E[Y|C = c, D = 0) w_c$.

[26] We also test for dynamic effects of female name assignment. We do not find evidence for this; the p-value of the joint test of the assignment to a female name in the previous two working days does not reject the null hypothesis of no effect either individually or jointly ($p = 0.516$).

[27] We find 81% of sales occur within 24 hours and 88% occur within 48 hours, relative to the total number of sales over 7 days.

[28] The total-value and price measures only include purchases by customers and not purchases by agents on behalf of customers. Interpreting these measures is consequently more challenging because the mode of purchase is potentially endogenous to treatment. In practice, we find treatment status affects purchases by agents on behalf of customers and purchases made directly by customers in the same magnitude and direction (see Table A6). This means we likely *underestimate* the coefficient on total price.

market types, or contexts could all potentially reflect some of the differences.

These results highlight the importance of customer-side discrimination in productivity differences between women and men in the workplace (for consumer-facing roles). Prior research on the gender wage gap suggests women receive lower pay partly because they are less productive (Blau and Kahn, 2017; Caliendo et al., 2017; Gallen et al., 2017; Sin et al., 2020). We show that discriminatory behavior—on the part of consumers—can drive these productivity differences. In our context, for women and men to have similar productivity levels, women would need to overcome significant barriers created by consumers' behavior. These results also suggest that piece-rate wage structures—i.e., rewarding employees for their output levels—could further workplace inequality.

4.2 Mechanisms

Why do customers discriminate? We use the richness of our data to explore three potential mechanisms that may explain why purchases fall when consumers chat with female agents. First, customers may be hesitant to engage with female sales agents. For instance, customers may dislike working with women or believe women are less efficient at helping with purchases. Second, recent work suggests women are more likely to face harassment and verbal abuse on the job (Georgieva, 2018; Folke and Rickne, 2022; Dupas et al., 2021) and, third, may face different bargaining processes (Ashraf, 2009; Castillo et al., 2013; Card et al., 2016; Vesterlund, 2018; Rousille, 2021).

Engagement The data suggests that the first mechanism is more consistent with the results: customers are hesitant to engage with female agents. We investigate this along the extensive and intensive margin of the conversation. On the extensive margin, we investigate whether the customer engages with an agent at all, as some consumers may be hesitant to chat with female agents or may entirely avoid female agents. On the intensive margin, we investigate what customers discuss, and the tones they use to express themselves. Note that we use these two measures as imperfect proxies because customer engagement is impossible to measure directly. This means we are likely to miss some changes in customer engagement and we do not expect that the magnitude of our measured effect on this mechanism will be able to explain the entirety of the main sales effect.

Columns (1-2) in Panel A of Table 3 show the effect of female-name assignment on *extensive margin* consumer interactions. Mechanically, agents always send the first message due to the website sending automated messages, and the conversations begin afterwards. In column (1), female assignment leads to a negative, albeit statistically insignificant effect, on the likelihood the customer ever responds ($p = 0.322$). However, agents can send multiple messages to customers to encourage their response, which means that measuring a binary variable of any response by the customer may not fully capture a lack of engagement. Column (2) shows that female-assigned agents send more messages before receiving a response ($p = 0.034$), consistent with lower customer engagement (higher hesitance).[29]

We further investigate customer hesitancy along the *intensive margin* in two ways. First, we analyze the conversations' tones. While specific tones are likely imperfect proxies for genuine emotions, whether a customer expresses any tone may reflect a customer's overall level of engagement with the agent. To this end, research assistants constructed a measure for any non-neutral tone detected in the conversation. Column (3) in Panel A of Table 3 indicates a 3.1 percentage point reduction in the probability of any tone used when customers engage with female-assigned agents, a 35% reduction relative to the control-group mean ($p = 0.032$).[30]

Second, we investigate whether the topic or purpose of the chat changes when customers believe they are talking to a female agent. The results are presented in Panel B of Table 3. As outlined in subsection 2.3, we hand-coded every chat transcript and categorized the primary and secondary purpose of the chat into 4 groups (general inquiry, price inquiry, making a booking, or other).[31] Column (3) shows that female-name assignment does not affect the probability that the primary purpose of the chat is to make a booking.[32] However, column (4) indicates the probability that the secondary purpose of the chat transitions to discussing making a booking falls by 5.7 percentage points (22%, $p = 0.07$). Column (5) shows this effect is much more pronounced among customers

[29] Note that this should not be interpreted as differential worker behavior based on treatment status (which they do not know). Instead, this should be interpreted as agents following up when customers do not engage, which occurs more often when they are assigned a female name.

[30] Table A11 presents results separately for each tone. We find customers are less likely to express happiness, more likely to express sadness, and also less likely to be impatient. They are not more likely to express anger or be ecstatic. Since these forms of engagement operate across various emotional dimensions, we aggregate them into a single measure of overall tone or engagement.

[31] Other includes less frequent events like confirming, cancelling, changing a booking, reporting a complaint, and random/unknown reasons.

[32] The treatment status also does not affect the likelihood of other initial topics (see Table A4).

who were coming to talk to agents about general or price inquiries than other topics: the likelihood of transitioning to making a booking falls by 10 (14.5) percentage points for customers who initially had a general (price) inquiry. Instead of discussing a booking, conversations with female-name agents simply never transition to another topic; column (6) shows conversations with agents assigned female names have similar-sized reductions in the likelihood of any secondary conversational purpose. Together, the results provide evidence for initial hesitance, less expression, and a lower likelihood of transitioning towards the ultimate goal of customers making purchases.

The finding that customers are less likely to make purchases after simple price inquiries about specific hotels on specific dates suggests some difficulty in information transferal between worker and customer. This could reflect different beliefs in the quality of this information or differences in the acquisition of information based on the gender of the person providing the information (Conlon et al., 2021).

Other Mechanisms We also investigate whether customers are more abusive or more likely to bargain with women.[33] The results in column (4-6) of Panel A in Table 3 suggest these mechanisms do not explain the differences in sales in our setting. Column (4) measures whether any language is classified as harassment within the chat. The data contain few instances of harassment: 0.3% of conversations for the male-assigned (control) sample indicate harassment. The rate in the female-assigned sample is practically identical to the male-assigned sample and does not differ statistically. Using a broader definition of any negative wording, we find no evidence for this mechanism either (column 5). Finally, column (6) tests whether customers bargain more often with female sales agents. While 14% of chats exhibit some bargaining behavior—for example, asking for discounts on the listed price—we find no significant effect of female-name assignment on the likelihood of bargaining. This null result rules out meaningful changes in the amount of bargaining faced by women. Therefore, in this context, differential bargaining does not appear to drive the observed productivity differences.

[33] While we consider abuse to be a mechanism affecting worker productivity, we acknowledge that it could also be an outcome if customers abuse women who they perceive to be less productive (Folke and Rickne, 2022). Here we view abuse as plausibly affecting productivity by decreasing worker morale and their ability to sell to *other* customers.

4.3 Robustness and interpretational confounds

Robustness Our main results are robust to various analysis choices. First, the results are almost identical when we aggregate to the customer-day level to account for customers returning to the site multiple times within the same day (Table A5). Second, the results are robust to how we define our outcomes (Table A6). Specifically, the effects are remarkably similar when looking at purchases from chat-based records (via agents, column 3) or from administrative records (via customers, columns 4-5). Third, our results are also unaffected if we exclude any of the relatively few days when customers use a gendered identifier (e.g. Sir or Miss), which could potentially reveal the treatment assignment to the agent (columns 6-7).

Fourth, we show our results are robust to choices over alternative regression specifications in Table A7. The results are quantitatively and qualitatively similar when we include additional customer controls for past purchases, customer location, and customer chat history (column 3), add day-of-week (column 4) or week fixed effects (column 5), include agent and month fixed effects separately (column 6), only have date fixed effects (column 7), and remove all fixed effects and controls (column 8). The effect of female status leads to proportional reductions of 41-58% across all specifications.

Name confounds A common concern in studies that randomly assign names to measure discrimination along a particular margin is that the chosen names may correlate with other characteristics that could drive the results (Bertrand and Mullainathan, 2004; Charles and Guryan, 2011). This concern is more likely to be relevant for studies that select names based on some specific criteria (e.g., "racially distinct" names), such that the names themselves may be associated with other characteristics that may not reflect those of the average person in that subgroup (Fryer Jr and Levitt, 2004). This can impact the internal validity of the study. In contrast, we draw all names from a local yearbook to approximate the actual distribution of names in the population since first names already typically confer gender. In the rest of this section, we delve into the importance of other features of the assigned names.

We use a number of strategies to assess the extent to which the displayed first names correlate with other characteristics besides gender that may impact customer behavior. As in Bertrand and

Mullainathan (2004); Kline et al. (2022), we jointly test whether the effects of each different first name impacts customer discrimination, controlling for the agent's assigned sex. We use a less demanding specification that controls for agent and month fixed effects instead of agent-month fixed effects for statistical power. In the 66% of the sample featuring an assigned name that was repeated across dates, we fail to reject that first names have no causal effect on customer behavior within assigned sex ($p = 0.39$).

Finally, we hand-coded additional characteristics associated with name. A potentially salient feature of customer behavior in this context could be ethnicity. A priori names are unlikely to confer much information about ethnicity because only first names are shown. To further support this, column (2) in Table A7 shows that our results are unaffected by directly controlling for name ethnicity as fixed effects.[34] We also hand-coded whether names are traditionally English-sounding names. We show in section 6 that controlling for this covariate also does not change the results.[35]

Interpretational confounds Agents may have certain gender-specific language that could appear strange to consumers when assigned the opposite gendered name. For example, a male agent may use specific language that could confuse a customer who assumes they are speaking with a woman because of their female-sounding name—and this may reduce the chance of a sale. In practice, this is unlikely given the short and straightforward conversations we study. Still, could customers be responding to a mismatch between the gender implied by agents' assigned name and the agents' actual gender? It is unlikely that this confound explains our results because every agent in our experimental sample is female and could only potentially confuse a customer with their language when they are assigned a male-sounding name. However, because we find

[34] Ethnicity is coded by a field team based on full name. We assign name ethnicity based on the full name although only first names were shown. There are 17 ethnicities in the data.

[35] As an additional check we link the assigned names to other characteristics that may be important to customers. Customers may infer something about worker productivity based on their beliefs about worker age or socioeconomic status. We use voter registry data from the United States to identify characteristics besides gender associated with the displayed first names. Specifically, we follow Norris et al. (2021) to create measures of the average age, racial, socioeconomic status associated with a particular name. We acknowledge that first name associations might differ across continents, so data from the United States may not represent all customer inferences in our setting. However, since we have a global customer base, any dataset from a single region would face the same challenge, and we are unaware of similar comprehensive data from other sources. Despite its limitations, we find this exercise to be useful for ruling out certain confounds. Table A8 presents a number of specifications relating name assignment and the likelihood of any customer purchase. Across the name characteristics of female, age, neighborhood % black, and neighborhood % below the poverty line, we find only female name assignment affects purchases, and it does not appear that the impact of female names is driven by correlations with names that are associated with older individuals or with higher socioeconomic status.

that being assigned a female name reduces the likelihood of a sale, any confusing behavior from a male-sounding name may attenuate our estimates.[36]

5 Labor Market Implications

In the previous sections, we found customers are less likely to make purchases when working with agents with female-sounding names. While discrimination that reduces female worker productivity in this occupation suggests negative impacts on worker well-being, how this observed bias affects workers ultimately depends on the labor market (Becker, 1957). If workers can sort away from the customer bias in this occupation into another equivalent job without customer contact, for example, there may be no consequences for worker welfare. In this firm, as is common in the customer service industry, the vast majority of workers in this role are female. Two thirds of workers at this firm are female, and women handle 85% of chat conversations. This indicates that equilibrium labor market sorting is not reallocating workers away from this observed, important source of discrimination.

Why might this be? Sources for the lack of sorting could arise either from the firm side or from the worker side. It is possible that firms are unaware of these productivity differences and simply hire female workers who are lower productivity as a result of customer bias. It is equally possible that female workers face other constraints to taking outside occupations.

To try to understand whether female workers at this firm are less productive than their male counterparts as a result of customer bias, we compare the results from our experimental research design to a simpler non-experimental comparison of male and female agents. The non-experimental results measure correlations between chat purchases and agents' actual gender using chats with over 7,000 customers *outside* the experimental sample.[37] We include office by month-of-sample fixed effects in the non-experimental regression models to limit comparisons between male and female workers working in the same location, with similar customers, over the same period.

Table 4 shows correlations between agents' actual gender and sales in Panel A and experimental estimates of female name assignment in Panel B. In Panel A, we find no economically or

[36] Furthermore, our results are unlikely to be different for a sample of male agents as we find customers have similar proportional reductions in their purchases even if they have short conversations and hence little interaction with the agents (this table is presented in the working paper version of this manuscript.)

[37] The experimental sample is not representative of all sales at the company. Therefore, this exercise is only suggestive.

statistically significant differences between male and female agents across any of the purchase outcomes. In Panel B, we reproduce our main experimental estimates for comparison—we find female name assignment leads to statistically significant reductions in sales across all purchase outcomes. To compare the effect of female-name assignment in the experimental sample to the effect of being female in the correlational sample, we use seemingly unrelated regression. A test of equality across the two female coefficients rejects the null hypothesis at the 5% level for any purchase and number of purchases within 48 hours and at the 10% level for outcomes within 24 hours.[38]

The difference between the experimental and correlational estimates sheds light on gender-based selection into this occupation. In particular, the difference suggests women in these jobs may be more productive than their male counterparts in the absence of customer discrimination. This is consistent with an equilibrium outcome in which males and females are paid similar wages, with female employees being taxed by customer bias.

Consequently, it does not appear that the firm faces a trade-off when hiring male or female workers, but female workers experience significant discrimination from customers in this job. Given that female workers do not sort away, there may be similar or more discrimination in other jobs they are able to obtain, or potentially non-pecuniary benefits and comparative advantage women experience in their current jobs may deter them from leaving. Together, these results highlight important amounts of market-level discrimination and suggest that simple comparisons of job or industry composition may not be good indications of bias faced by workers.

6 External Validity

These results are important in their own right as they provide a proof of concept that customers can be a meaningful source of labor market discrimination. Nevertheless, we can leverage our data to explore whether the impacts of customer discrimination are moderated by the context where our research is conducted.

[38] Additionally, the rates of sales for agents assigned to a female name in the experimental sample are comparable to sales by female agents in the non-experimental sample (e.g., .042 vs .049 in column 2, respectively).

Customers Are the results specific to certain types of customers? We can test this in a number of ways. First, we investigate whether the results are specific to customers purchasing from Africa versus abroad.[39] Second, we explore whether customers discriminate along another salient margin: whether the name of the agent they are talking to is English-sounding or not. Table 5 presents these results for the full sample (columns 1-2), customers purchasing from Africa (columns 3-4), and for customers purchasing from abroad (columns 5-6). Focusing on columns (1, 3, 5), which estimate the effect of female-name assignment across customer locations, we see that our results are remarkably consistent. If anything, the results are slightly larger for customers from abroad, who reduce their purchases by 5.2 percentage points, relative to those from Africa, who reduce their purchases by 3.5 percentage points (though the impacts are not statistically different from one another, $p = 0.51$).[40]

Next, we investigate customer bias along other margins in columns (2, 4, 6). The most prominent other dimension in this setting is related to race, which we investigate based on whether the names are common English-sounding names. African customers are no less likely to purchase products for agents with non-English names (column 4, $p = 0.449$), while customers outside Africa, who may be less familiar and more biased towards these names, reduce their purchases by 4.2 percentage points (column 6, $p = 0.097$). We reject the effects are equal across these customers ($p = 0.07$); therefore, this type of race-based customer discrimination is differential and limited to certain types of customers. This stands in contrast to the gender-based discrimination we observe, which we detect across all customers.

Finally, customer behavior may be driven by homophily—a customer's preference to interact with an agent of the same gender. Directly evaluating this is challenging because customers do not fill out a survey or provide demographic information before speaking with their agent. We observe a small number of interactions (135) in which a customer reveals their name (and hence their gender) when they introduce themselves or during the conversation. We hand-coded these conversations and find that 76% of customers reveal their gender to be male.

Using instances where customers reveal their names in our analysis is challenging because of low statistical power and the population who reveals their name are likely different from the full

[39] The majority of customers abroad are in North America or Europe.
[40] We also find quantitatively similar results when splitting across East and West Africa.

population in important ways. Keeping these caveats in mind, Table A9 presents a number of analyses relating to this sample. We use a less demanding specification with agent and month fixed effects to increase power. Column (1) shows the female name assignment is uncorrelated with whether the customer reveals their name.[41] While customers who reveal their name are much more likely to purchase a product on average (23% vs 7%), the impact of female name assignment in the full population (column 2) and the customer-name-reveal sample (column 3) are qualitatively similar. Both exhibit decreases in purchases (41% vs 47% relative reduction), but the effect in the customer reveal sample is statistically insignificant ($p = 0.12$). Columns (4-5) present the results separately for males and females. In both, we find statistically insignificant but comparable parameter estimates (-.12 and -.095). While the point estimates are both appreciably negative, the sample sizes are small and we cannot reject model effects in the opposite direction. Taking the point estimates literally suggests homophily may not be an important factor driving these results.

Workers While this paper tests customer discrimination across thousands of customers, we have relatively few workers engaging with customers. This would be a concern for external validity if we thought the results were unique to the specific set of workers with which we engage. Given that all agents in our sample are assigned to both treatment statuses, we can directly test for heterogeneity across workers. Table A10 shows we cannot reject that the treatment effects are the same across all agents, in levels ($p = 0.69$) and when calculated proportionally as a fraction of the agent's sales when assigned a male name ($p = 0.58$).[42] Additionally, the agent with the smallest proportional reduction still experiences 25% fewer customer purchases when assigned a female name (relative to their sales rate when not assigned a female name), demonstrating that this customer discrimination is economically significant for all workers.

Further institutional reasons suggest these workers are unlikely to be atypical of other workers at the firm. As discussed in subsection 2.2, the sample of workers in the experimental arm included all employees working consistently in the two branch offices when the experiment launched.

[41] This says that customer responses are balanced across female name assignment, but this does not reveal whether the two genders are equally likely to reveal their name based on treatment assignment. This is unobservable because we do not observe the true share of male and female customers in the population.

[42] This table includes 5 rather than 6 agents given that one agent has few observations and we cannot consistently estimate their individual-specific treatment effect.

Consequently, worker selection was not based on specific features of the workers themselves. Moreover, all workers being female in our experimental sample reflects the high rates of female workers in this profession more broadly.

Alternatively, our external validity may be threatened if the name-masking procedure affects workers' productivity such that they are unable to express their identity as they otherwise would. This concern does not threaten our identification of the effect of customer bias since it is equally true when assigning male- or female-sounding names, but it may create interactions that are less reflective of reality. In practice, this seems unlikely as female workers appear to sell similar amounts when assigned a female name in the experiment compared with female workers who are not in the experiment (see section 5 for more details).

Jobs Finally, while our results may be specific to the types of customer sales roles we focus on, these jobs are comparable to the majority of jobs in this industry. Most companies have customer service departments, and hire sales agents to engage customers over the phone or online, answering their queries and assisting with sales. Amazon alone employs customer service associates in more than 130 locations in over 40 countries around the world. These sales agents must know the products on offer and be able to satisfy the demands of customers who may come from different parts of the world. Also, the products being sold are non-technical and not experience-based (i.e. the sales agents are not expected to have tried the products or to have provided recommendations based on their experience), which is common among these jobs.

7 Conclusion

This paper demonstrates that customer-based discrimination negatively affects female worker productivity. When sales agents randomly receive female-sounding names, the probability a customer makes a purchase falls by 50%. Consumers also purchase fewer total products, and the total value of their purchases decreases. We find supporting evidence that customers' lack of interest is driving these effects. Customers lag in responding to female agents and are less likely to transition from their initial inquiry into a discussion about purchasing. In contrast, we do not find evidence that harassment or differential bargaining are important in this context.

Interestingly, these experimental results reveal significant levels of customer discrimination in a market where actual sales rates are equal between men and women. The lack of observed productivity differences between genders would naturally arise if customer discrimination was internalized by the firm: the firm only retains women who can overcome customer discrimination, such that their observed productivity is the same as men's. It is only by running our experiment that we can observe that female employees are more productive than their male counterparts absent customer bias.

Identifying the existence and extent of customer discrimination in a real-world setting is particularly relevant for two reasons. First, seminal work by Becker (1957) suggests that customer-based discrimination will not be competed away in equilibrium because firms internalize these preferences—exactly the dynamic we observe. Secondly, hypotheses that workers may be able to avoid from industries in which they face customer discrimination, thereby limiting its impact, do not appear to hold in our setting (Heckman, 1998). More broadly, if women are unable to avoid customer discrimination through sorting, this may present a barrier to female labor force participation. Women who would otherwise be productive employees will not be hired due to customer preference. Therefore, the customer discrimination we document potentially presents an important and persistent labor market distortion. This is especially relevant in the context where we work: significant gender disparities exist in formal-sector employment across sub-Saharan Africa, where less than 15 percent of women work full-time for an employer (World Bank, 2013; Klugman and Twigg, 2016).

From a policy perspective, the most direct approach to tackling this problem is to change customer norms around women in the workplace. Governments may use programs that increase the representation of women in positions of power, exposing the general population to women as authority figures.[43,44] Similarly, if firms believe they could capture future benefits by sensitizing customers (perhaps through market power) or choose to because of pro-social intentions, these same firms may seek to change customer norms themselves. Nevertheless, norm change is likely to be a difficult and slow process, and since many customer services positions are already pre-

[43] This is similar to Beaman et al. (2009) who show that prior exposure to a female politician improves positive perceptions of female leaders and relates more broadly to a literature on the contact hypothesis.

[44] Relatedly, Webb (2023) shows neighborhood discussions can reduce customer discrimination for transgender workers.

dominantly female, simply hiring female sales representatives is unlikely to lead to norm change.

Absent norm change, a second-best approach may be to limit the consequences of customer-based discrimination on female employees. This consideration is particularly relevant for industries which tie employee pay to productivity/output (e.g., number of sales) through piece-rate wages. Our results suggest that such individual-based incentivized pay schemes may increase the impact of customer discrimination on worker pay.[45] Second, some companies have found that obscuring identities makes the job easier for their customer service representatives (Chan, 2022). In our setting, discrimination's effects might be eliminated by agents using gender-neutral names (or avoiding names altogether). While such measures could reduce inequality, they also potentially perpetuate the bias that creates these inequalities in the first place.

[45] For example, one way to limit this source of discrimination is for employers to 'pool' performance-based bonuses—a common practice for sharing tips in the restaurant industry.

Data Availability

Code replicating the tables and figures in this article can be found in Kelley et al. (2024) in the Harvard Dataverse, https://doi.org/10.7910/DVN/V2DR.

References

ALSAN, M., O. GARRICK, AND G. GRAZIANI (2019): "Does Diversity Matter for Health? Experimental Evidence from Oakland," *American Economic Review*, 109, 4071–4111.

ANGRIST, J. D. AND J.-S. PISCHKE (2008): *Mostly harmless econometrics*, Princeton university press.

ASHRAF, N. (2009): "Spousal control and intra-household decision making: An experimental study in the Philippines," *American Economic Review*, 99, 1245–77.

AYRES, I., M. BANAJI, AND C. JOLLS (2015): "Race effects on eBay," *The RAND Journal of Economics*, 46, 891–917.

BAR, R. AND A. ZUSSMAN (2017): "Customer discrimination: evidence from Israel," *Journal of Labor Economics*, 35, 1031–1059.

BASFORD, T. E., L. R. OFFERMANN, AND T. S. BEHREND (2014): "Do You See What I See? Perceptions of Gender Microaggressions in the Workplace," *Psychology of Women Quarterly*, 38, 340–349.

BEAMAN, L., R. CHATTOPADHYAY, E. DUFLO, R. PANDE, AND P. TOPALOVA (2009): "Powerful Women: Does Exposure Reduce Bias?" *The Quarterly Journal of Economics*, 124, 1497–1540.

BECKER, G. S. (1957): "The Economics of Discrimination," *University of Chicago Press Economics Books*.

BENSON, A., D. LI, AND K. SHUE (2021): "Potential and the gender promotion gap," *Unpublished Working Paper*.

BERTRAND, M. AND E. DUFLO (2017): "Field experiments on discrimination," *Handbook of economic field experiments*, 1, 309–393.

BERTRAND, M. AND S. MULLAINATHAN (2004): "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination," *American economic review*, 94, 991–1013.

BLAU, F. D. AND L. M. KAHN (2017): "The Gender Wage Gap: Extent, Trends, and Explanations," *Journal of Economic Literature*, 55, 789–865.

BORKER, G. (2021): "Safety First: Perceived Risk of Street Harassment and Educational Choices of Women," *Working Paper*.

BURSZTYN, L., A. L. GONZÁLEZ, AND D. YANAGIZAWA-DROTT (2020): "Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia," *American Economic Review*, 110, 2997–3029.

CALIENDO, M., W.-S. LEE, AND R. MAHLSTEDT (2017): "The Gender Wage Gap and the Role of Reservation Wages: New Evidence for Unemployed Workers," *Journal of Economic Behavior & Organization*, 136, 161–173.

CARD, D., A. R. CARDOSO, AND P. KLINE (2016): "Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women," *The Quarterly journal of economics*, 131, 633–686.

CASELLI, M. AND P. FALCO (2020): "As long as they are cheap: Experimental evidence on the demand for migrant workers," *Available at SSRN 3685873*.

CASELLI, M., P. FALCO, AND G. MATTERA (2023): "When the stadium goes silent: How crowds affect the performance of discriminated groups," *Journal of Labor Economics*, 41, 431–451.

CASTILLO, M., R. PETRIE, M. TORERO, AND L. VESTERLUND (2013): "Gender Differences in Bargaining Outcomes: A Field Experiment on Discrimination," *Journal of Public Economics*, 99, 35–48.

CHAN, W. (2022): "The AI Startup Erasing Call Center Worker Accents: Is It Fighting Bias – or Perpetuating It?" *The Guardian*.

CHARLES, K. K. AND J. GURYAN (2011): "Studying discrimination: Fundamental challenges and recent progress," *Annu. Rev. Econ.*, 3, 479–511.

COMBES, P.-P., B. DECREUSE, M. LAOUENAN, AND A. TRANNOY (2016): "Customer discrimination and employment outcomes: theory and evidence from the French labor market," *Journal of Labor Economics*, 34, 107–160.

CONLON, J. J., M. MANI, G. RAO, M. W. RIDLEY, AND F. SCHILBACH (2021): "Learning in the Household," Tech. rep., National Bureau of Economic Research.

DEAN, J. T. AND S. JAYACHANDRAN (2019): "Changing Family Attitudes to Promote Female Employment," *AEA Papers and Proceedings*, 109, 138–142.

DELECOURT, S. AND O. NG (2021): "Does gender matter for small business performance? Experimental evidence from India," .

DESOUZA, E. R., E. D. WESSELMANN, AND D. ISPAS (2017): "Workplace Discrimination against Sexual Minorities: Subtle and not-so-subtle," *Canadian Journal of Administrative Sciences / Revue Canadienne des Sciences de l'Administration*, 34, 121–132.

DOLEAC, J. L. AND L. C. STEIN (2013): "The visible hand: Race and online market outcomes," *The Economic Journal*, 123, F469–F492.

DUFLO, E. (2012): "Women empowerment and economic development," *Journal of Economic literature*, 50, 1051–79.

DUPAS, P., A. SASSER MODESTINO, M. NIEDERLE, J. WOLFERS, AND T. S. D. COLLECTIVE (2021): "Gender and the Dynamics of Economics Seminars," Working Paper 28494, National Bureau of Economic Research.

EGAN, M., G. MATVOS, AND A. SERU (2022): "When Harry fired Sally: The double standard in punishing misconduct," *Journal of Political Economy*, 130, 1184–1248.

FIELD, E., R. PANDE, N. RIGOL, S. SCHANER, AND C. TROYER MOORE (2021): "On Her Own Account: How Strengthening Women's Financial Control Impacts Labor Supply and Gender Norms," *American Economic Review*, 111, 2342–2375.

FOLKE, O. AND J. RICKNE (2022): "Sexual harassment and gender inequality in the labor market," *The Quarterly Journal of Economics*, 137, 2163–2212.

FRYER JR, R. G. AND S. D. LEVITT (2004): "The causes and consequences of distinctively black names," *The Quarterly Journal of Economics*, 119, 767–805.

GALLEN, Y., R. V. LESNER, AND R. VEJLIN (2017): "The Labor Market Gender Gap in Denmark: Sorting out the Past 30 Years," 41.

GEORGIEVA, K. (2018): "Changing the Laws That Keep Women out of Work," https://www.worldbank.org/en/news/opinion/2018/03/29/changing-the-laws-that-keep-women-out-of-work.

GLOVER, D., A. PALLAIS, AND W. PARIENTE (2017): "Discrimination as a self-fulfilling prophecy: Evidence from French grocery stores," *The Quarterly Journal of Economics*, 132, 1219–1260.

GOLDIN, C. AND C. ROUSE (2000): "Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians," *American Economic Review*, 90, 715–741.

HARDY, M. AND G. KAGY (2020): "It's getting crowded in here: experimental evidence of demand constraints in the gender profit gap," *The Economic Journal*, 130, 2272–2290.

HEATH, R. AND X. TAN (2020): "Intrahousehold Bargaining, Female Autonomy, and Labor Supply: Theory and Evidence from India," *Journal of the European Economic Association*, 18, 1928–1968.

HECKMAN, J. J. (1998): "Detecting discrimination," *Journal of economic perspectives*, 12, 101–116.

HENGEL, E. (2022): "Are Women Held to Higher Standards? Evidence from Peer Review," *The Economic Journal*.

HOLZER, H. J. AND K. R. IHLANFELDT (1998): "Customer discrimination and employment outcomes for minority workers," *The Quarterly Journal of Economics*, 113, 835–867.

HULL, P., A. IMAS, AND J. A. BOHREN (2023): "Systemic Discrimination: Theory and Measurement," *Working Paper*.

HURST, E., Y. RUBINSTEIN, AND K. SHIMIZU (2021): "Task-based discrimination," Tech. rep., National Bureau of Economic Research.

IFC (2021): "Women and E-Commerce in Africa," *International Finance Corporation*, 74.

JAYACHANDRAN, S. (2015): "The roots of gender inequality in developing countries," *economics*, 7, 63–88.

KAHN, L. M. AND P. D. SHERER (1988): "Racial differences in professional basketball players' compensation," *Journal of Labor Economics*, 6, 40–61.

KELLEY, E., G. LANE, M. PECENCO, AND E. RUBIN (2024): "Replication Code for Customer Discrimination in the Workplace: Evidence from Online Sales," *Harvard Dataverse* *https://doi.org/10.7910/DVN/V2DRAR*.

KLINE, P., E. K. ROSE, AND C. R. WALTERS (2022): "Systemic discrimination among large US employers," *The Quarterly Journal of Economics*, 137, 1963–2036.

KLUGMAN, J. AND S. TWIGG (2016): "Gender at Work in Africa: Legal Constraints and Opportunities for Reform," *African Journal of International and Comparative Law*, 24, 518–540.

KRICHELI-KATZ, T. AND T. REGEV (2016): "How many cents on the dollar? Women and men in product markets," *Science advances*, 2, e1500599.

LEONARD, J. S., D. I. LEVINE, AND L. GIULIANO (2010): "Customer discrimination," *The Review of Economics and Statistics*, 92, 670–678.

LIST, J. A. (2004): "The nature and extent of discrimination in the marketplace: Evidence from the field," *The Quarterly Journal of Economics*, 119, 49–89.

LIVEAGENT, W. (2022): "Live Agent," https://www.liveagent.com/customer-support-glossary/agent-alias/.

LOWE, M. AND M. MCKELWAY (2021): "Coupling Labor Supply Decisions: An Experiment in India," *Working Paper*.

MACNELL, L., A. DRISCOLL, AND A. N. HUNT (2015): "What's in a name: Exposing gender bias in student ratings of teaching," *Innovative Higher Education*, 40, 291–303.

MCKELWAY, M. (2021a): "How Does Women's Employment Affect Household Decision-Making? Experimental Evidence from India," *Working Paper*.

——— (2021b): "Women's Employment in India: Intra-Household and Intra-Personal Constraints," *Working Paper*.

MENGEL, F., J. SAUERMANN, AND U. ZÖLITZ (2019): "Gender bias in teaching evaluations," *Journal of the European economic association*, 17, 535–566.

NARDINELLI, C. AND C. SIMON (1990): "Customer racial discrimination in the market for memorabilia: The case of baseball," *The Quarterly Journal of Economics*, 105, 575–595.

NORDELL, J. (2021): "The End of Bias: A Beginning," https://us.macmillan.com/books/9781250186188/theendofbiasabeginning.

NORRIS, S., M. PECENCO, AND J. WEAVER (2021): "The effects of parental and sibling incarceration: Evidence from ohio," *American Economic Review*, 111, 2926–2963.

O'DONNELL, M., U. NWANKWO, A. CALDERON, C. STRICKLAND, ET AL. (2020): "Closing Gender Pay Gaps," .

PARSONS, C. A., J. SULAEMAN, M. C. YATES, AND D. S. HAMERMESH (2011): "Strike three: Discrimination, incentives, and evaluation," *American Economic Review*, 101, 1410–35.

ROUSILLE, N. (2021): "The Central Role of the Ask Gap in Gender Pay Inequality," *Working paper*.

SARSONS, H. (2022): "Interpreting signals in the labor market: evidence from medical referrals," *Job Market Paper*, 141–145.

SIN, I., S. STILLMAN, AND R. FABLING (2020): "What Drives the Gender Wage Gap? Examining the Roles of Sorting, Productivity Differences, Bargaining and Discrimination," *The Review of Economics and Statistics*, 1–44.

STATISTA (2019): "Online Shoppers in Africa by Gender," https://www.statista.com/statistics/1190608/online-shoppers-in-africa-by-gender/.

SUBRAMANIAN, N. (2021): "Workplace Attributes and Women's Labor Supply Decisions: Evidence from a Randomized Experiment," *Working Paper*.

UNCTAD (2018): "UNCTAD B2C E-COMMERCE INDEX 2018: FOCUS ON AFRICA," Tech. rep.

UNDP (2022): "Gender Inequality Index," Tech. rep., United Nations.

VESTERLUND, L. (2018): "Knowing When to Ask: The Cost of Leaning-in," Working Paper 6382, Department of Economics, University of Pittsburgh.

WEBB, D. (2023): "Silence to Solidarity: Using Group Dynamics to Reduce Anti-Transgender Discrimination in India," .

WORLD BANK, G. (2011): *World development report 2012: Gender equality and development*, The World Bank.

——— (2013): "Gender at Work: A Companion to the World Development Report on Jobs," https://www.worldbank.org/content/dam/Worldbank/document/Gender/GenderAtWork_web.pdf.

——— (2022): "World Development Indicators | DataBank," https://databank.worldbank.org/source/world-development-indicators.

——— (2024): "Gender Data Portal | DataBank," https://genderdata.worldbank.org/.

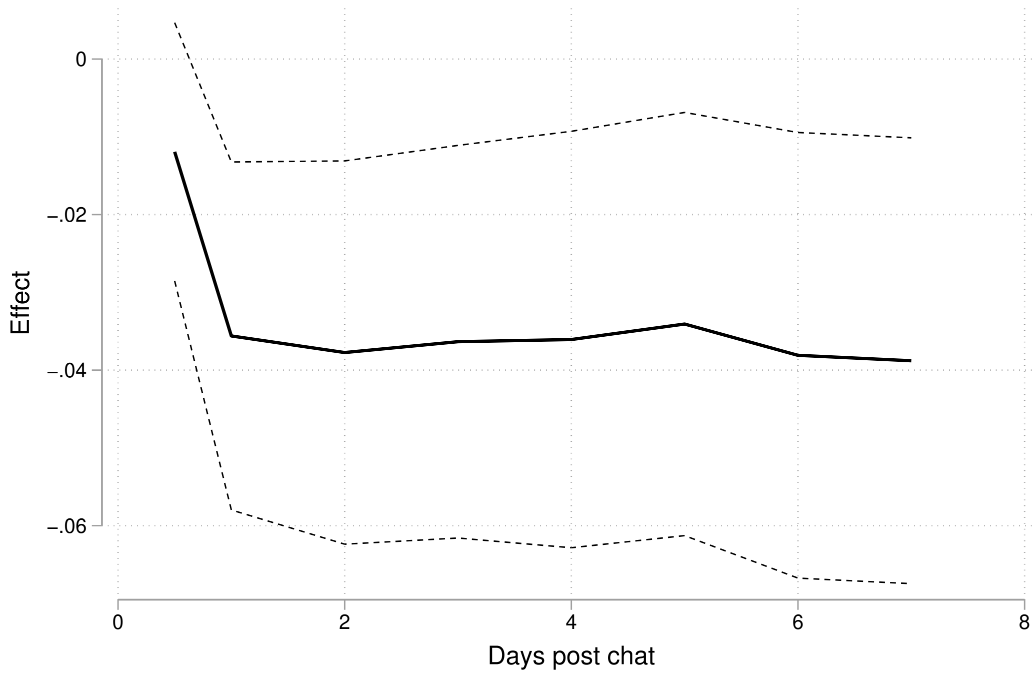
WORLD ECONOMIC FORUM, W. (2023): "Global Gender Gap Report 2023," https://www.weforum.org/publications/global-gender-gap-report-2023/.

YOTEBIENG, K. (2021): "What we know about persistent social norms that serve as barriers to girls' access, participation, and achievement in education in eight sub-Saharan African countries," https://www.ungei.org/publication/what-we-know-and-dont-know-about-persistent-social-norms-serve-barriers-girls-access.

ZIPPIA (2021): "Customer Service Representative Demographics and Statistics [2023]: Number Of Customer Service Representatives In The US," https://www.zippia.com/customer-service-representative-jobs/demographics/.

Figures

Figure 1: Effects of female name assignment over time



The figure shows rates of purchases and treatment effects over time. Panel (a) show the number of purchases per conversation X days from the time of the chat. The first observation corresponds to 0.5 days from the time of the chat and continues up to 7 days. Panel (b) shows the effect of female name assignment on the number of purchases, up to X days from the time of the chat. Regressions in each period are estimated separately and include the baseline set of fixed effects: agent-month fixed effects. Dotted lines represent 95% confidence intervals clustered at the agent-day and IP-address level.

Tables

Table 1: Balance tests for female assignment

	N	Var. Mean	Female
Customer mention agent true name	2655	.00	-.00128
			(.00234)
Customer amount of past chats	2655	.38	-.0667
			(.0439)
Customer amount of past purchases	2655	.29	-.0166
			(.0458)
Agent first message length	2655	5.47	-.00066
			(.0041)
Day of week	2655	3.08	-.316
			(.213)
Agent chats (daily)	337	7.76	-.106
			(.596)
Agent worked previous day (daily)	337	.54	-.0827
			(.0583)
Agent hours worked (daily)	337	2.57	-.0291
			(.172)
Total working agents (daily)	337	8.23	.0411
			(.169)
Joint p-value			.61

This table shows customer and agent outcome means in column (2) and correlation between female name assignment and outcomes in column (3). The number of chats, whether worked on previous day, hours worked by agents, and total working agents are at the day level, while the other variables are at the chat level. Female indicator determined in customer's first chat of the day. Controls include agent-month fixed effects. Standard errors in parentheses and clustered at agent-day level. Joint p-value tests equality of all coefficients with zero. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: Effect of female assignment on purchase outcomes

	Purchases (48h)		Purchases (24h)	
	(1)	(2)	(3)	(4)
	Any	Total	Any	Total
Female	-.038***	-.038***	-.036***	-.035***
	(.013)	(.014)	(.011)	(.012)
Control Mean (wt)	.076	.081	.070	.073
N	2655	2655	2655	2655

This table shows the effect of female name assignment on purchase outcomes. Any represents any purchase and Total represents number of purchases. Any purchases and total purchases combine purchases by customer and by agent. Purchases are measured within 24 or 48 hours of the start of the chat. Female indicator determined in customer's first chat of the day. Controls include agent-month fixed effects. The control group mean is reweighted by fixed effect cells to match the implied agent-month fixed effect OLS weights. Standard errors two-way clustered at the agent-day and customer-day level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Effect of female name assignment on chat response, outcomes, and purpose

	Extensive margin		Intensive margin			
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Chat response and outcomes						
	Ever respond	Msgs to response	Any tone	Harass	Any negativity	Bargaining
Female	-0.021	0.009**	-0.031**	0.000	0.004	0.006
	(0.021)	(0.004)	(0.014)	(0.003)	(0.004)	(0.018)
Control Mean (wt)	0.676	1.009	0.089	0.003	0.008	0.141
Observations	2,655	2,655	1,745	1,745	1,745	1,745
Panel B: Chat purpose						
			Initial purpose: Booking	Secondary purpose: Booking	Secondary purpose: Booking	Secondary purpose: Any
Female			-0.009	-0.057*		
			(0.025)	(0.031)		
Female X Initial general inquiry					-0.100**	-0.084*
					(0.045)	(0.045)
Female X Initial price inquiry					-0.145*	-0.108*
					(0.076)	(0.060)
Female X Initial make booking					-0.015	-0.026
					(0.020)	(0.042)
Female X Initial other purpose					-0.023	-0.029
					(0.026)	(0.034)
Control Mean (wt)			0.150	0.253	0.378	0.471
Observations			1,745	1,745	1,745	1,745

This table shows the effect of female name assignment on chat responses and outcomes. Columns (1-2) report extensive margin chat responses, while columns (3-6) report intensive margin chat behavior (conditional on any customer response). Panel A shows chat reponses and chat outcomes. Ever respond in column (1) is a 1 if the customer ever responded. Msgs to response in column (2) is the number of messages sent by agent before customer first response. Column (3) measures any non-neutral chat tone, column (4) measures any harassment of the agent, column (5) measures any negative words or phrases, and column (6) measures any bargaining. Panel B shows information on primary and secondary chat purposes. Column (3) measures whether the visitor's initial purpose of the chat was to make a booking. Columns (4-5) indicate whether the visitor's secondary purpose of the chat was to make a booking. Column (6) measures whether the conversation preceded to have any secondary purpose. Female indicator determined in customer's first chat of the day. Variables of the form Female X Initial ... are an interaction between the female indicator and an indicator for the initial purpose of the chat being one of the following: general inquiry, price inquiry, make a booking, or other purpose. Controls include agent-month fixed effects and also initial chat purpose for the heterogeneity analysis (Panel B, columns 5-6). The control group mean is reweighted by fixed effect cells to match the implied agent-month fixed effect OLS weights. Standard errors two-way clustered at the agent-day and customer-day level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Correlational relationship between female agent and administrative sales outcomes

	Purchases (48h)		Purchases (24h)	
	Any	Total	Any	Total
	(1)	(2)	(3)	(4)
Panel A: Correlational estimates				
Female	-0.003	-0.004	-0.002	-0.002
	(0.011)	(0.012)	(0.011)	(0.012)
Control Mean (wt)	0.049	0.053	0.046	0.048
Observations	8,867	8,867	8,867	8,867
Panel B: Experimental estimates				
Female	-0.038***	-0.038***	-0.036***	-0.035***
	(0.013)	(0.014)	(0.011)	(0.012)
Control Mean (wt)	0.075	0.080	0.070	0.072
Observations	2,655	2,655	2,655	2,655
Equality b/w Exp/Non-exp (p)	0.035	0.053	0.033	0.049

This table shows correlational and causal effects of female agent on sales outcomes from administrative records. Panel A shows correlational estimates using the non-experimental sample, while Panel B shows causal estimates from the experimental sample. Any represents any sale, Total represents number of sales, and Total price is the cumulative price of all sales in EUR. Sales are measured within 24 or 48 hours of the start of the chat. Female indicator determined in customer's first chat of the day. Controls include office-month fixed effects in Panel A, and agent-month fixed effects in Panel B. The control group mean is reweighted by fixed effects cells to match the implied OLS weights. For the correlational estimates, the control mean is reweighted by office-month, while in the experimental estimates it is reweighted by agent-month. Standard errors two-way clustered at the agent-day and customer-day level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Effect of female and non-English name assignment on purchase outcomes

	Any purchase (48h)					
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-.038***	-.037***	-.035**	-.034**	-.052**	-.053**
	(.013)	(.013)	(.014)	(.015)	(.026)	(.026)
Non-English name		.0044		.011		-.042*
		(.013)		(.015)		(.025)
Control Mean (wt)	.076	.076	.075	.075	.078	.078
F=Non-E (p)		.00		.01		.74
F=F C4,6 (p)						.51
Non-E=Non-E C4,6 (p)						.07
Sample	Full	Full	Africa	Africa	Not Africa	Not Africa
N	2655	2655	2321	2321	322	322

This table shows the effect of female and non-English name assignment on purchase outcomes, overall and by region. The outcome measures any purchase within 48 hours of the start of the chat. Female indicator determined in customer's first chat of the day. Non-english name indicator determined in customer's first chat of the day. Controls include agent-month fixed effects. The control group mean is reweighted by fixed effect cells to match the implied agent-month fixed effect OLS weights. F=Non-E (p) refers to the p-value from a test of equality between the Female and Non-English coefficients within the same model. F=F C4,6 (p) refers to the p-value from a test of equality between the Female coefficients across columns 4 and 6. Non-E=Non-E C4,6 (p) refers to the p-value from a test of equality between the Non-English coefficients across columns 4 and 6. Standard errors two-way clustered at the agent-day and customer-day level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Online Appendix for *Customer Discrimination in the Workplace: Evidence from Online Sales*

Tables

Table A1: Example names used in name assignment, by gender and non-English status

Female, English	Male, English	Female, Non-English	Male, Non-English
Emmily	Gabriel	Sekinat	Mohammad
Vivian	Kelvin	Khadijah	Dinah
Abigail	Elias	Genevieve	Jediel
Judith	Stanley	Habeeb	Hamza
Laura	Thomas	Chioma	Ebiason
Helen	Harrison	Fridah	Gideon
Joy	Sammy	Linet	Abuzarin
Eunice	Philip	Sherifa	Abayomi
Jayne	Isaiah	Consolata	Ezekiel
Marian	Samuel	Peace	Taiwo
Carolyne	Anthony	Bayu	Farouk
Josephine	Charles	Oluwafunmibi	Erastus
Lizzy	Denis	Zaida	Mories
Sharon	Simon	Aderonke	Umaru
Annmarie	Henry	Nduta	Brightone
Bianca	Dennis	Sadiya	Bashir
Clare	Lawrence	Nafisha	Alphonce
Deborah	Dick	Flavian	Halilu
Susan	Antony	Habu	Abdulrahman
Stephanie	Edwin	Staline	Adewale

This table shows a random set of names drawn from the dictionary of possible names to be assigned. Twenty names are presented for both male and female, and for both English and non-English coded names.

Table A2: Outcome variable descriptions

Purchases	
Any Purchases	Whether customer made any purchase 24 or 48 hours after the chat
Total Purchases	The total number of purchases that were made by the customer 24 or 48 hours after the chat
Total Price	The cumulative price of all purchases in EUR that were made by the customer 24 or 48 hours after the chat
Chats	
Ever Respond	$= 1$ if the customer ever responded
Messages to Response	Number of messages sent by agent before customer first response
Tone	
	We employed research assistants based in sub-Saharan Africa to read through all of the chats and categorize them by overall tone of the conversation and flag any instances of harassment, negativity, or bargaining.
Any	Measures any non-neutral chat tone. Chats were coded neutral or non-neutral tone (including angry, sad, happy, ecstatic, impatient)
Harassment	Measures any harassment of the agent
Any negativity	Measures whether any negative words or phrases were used by the customer
Bargaining	Measures any bargaining with the agent. This includes asking for discounts, or better prices
Purpose	
Initial / Secondary Purpose	Each chat was hand coded to capture the initial and secondary purpose (if any) the customer had when initiating the conversation with the chat agent. The initial purpose is defined as the first issue the customer raises with the agent while the secondary purpose is any subsequent topic after the primary issue was resolved.
General Inquiry	Customer requested general information about the platform or general service availability
Make a Booking	Customer asked for help making a specific booking
Price Inquiry	Customer asked about the price for a specific hotel or for a general category of hotels in an area on specific dates
Confirm Booking	Customer asked to confirm that their booking request had been received and processed

Other Purpose	Captures other less common reasons including complaints, date changes, and cancellations

Table A3: Effect of female assignment on purchase prices

	Purchase prices (48h)		Purchase prices (24h)	
	(1)	(2)	(3)	(4)
Female	-3.2***	-50**	-3.2***	-58***
	(1.2)	(21)	(1.1)	(20)
Control Mean (wt)	5.3	140	4.9	146
Sample	All	Sales	All	Sales
N	2655	68	2655	58

This table shows the effect of female name assignment on purchase prices. Price is the cumulative price of all purchases in EUR. Prices are based only on customer purchases (via admin-based records). Purchases are measured within 24 or 48 hours of the start of the chat. Odd columns include the full sample, while even columns only include the sample with any customer purchase within the relevant time frame. Female indicator determined in customer's first chat of the day. Controls include agent-month fixed effects. The control group mean is reweighted by fixed effect cells to match the implied agent-month fixed effect OLS weights. Standard errors two-way clustered at the agent-day and customer-day level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Effect of female assignment on initial chat purpose

	Initial chat purpose			
	(1) General Inquiry	(2) Make a Booking	(3) Price Inquiry	(4) Other
Female	.02	-.009	.0016	-.013
	(.03)	(.025)	(.016)	(.019)
Control Mean (wt)	.458	.222	.101	.219
N	1745	1745	1745	1745

This table shows the effect of female name assignment on initial chat purpose. The outcome variables are binary variables that are 1 if the initial purpose is about a general inquiry (column 1), about making a specific booking (column 2), an inquiry about a price (column 3), and other less common initial conversation purposes including complaints, confirmations, changes, cancellations, or unknown reasons (column 4). Female indicator determined in customer's first chat of the day. Controls include agent-month fixed effects. The control group mean is reweighted by fixed effect cells to match the implied agent-month fixed effect OLS weights. Standard errors two-way clustered at the agent-day and customer-day level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Effect of female assignment on purchase outcomes (customer-day level)

	Purchases (48h)		Purchases (24h)	
	(1)	(2)	(3)	(4)
	Any	Total	Any	Total
Female	-.04***	-.041***	-.038***	-.037***
	(.013)	(.014)	(.011)	(.013)
Control Mean (wt)	.077	.083	.072	.074
N	2172	2172	2172	2172

This table shows the effect of female name assignment on purchase outcomes. The data is aggregated to the customer-day level. Any represents any purchase and Total represents number of purchases. Any purchases and total purchases combine purchases by customer and by agent. Purchases are measured within 24 or 48 hours of the start of the chat. Female indicator determined in customer's first chat of the day. Controls include agent-month fixed effects. The control group mean is reweighted by fixed effect cells to match the implied agent-month fixed effect OLS weights. Standard errors clustered at the agent-day level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: Effect of female assignment on purchases, by outcome source and non-gender reveal sample

	Purchases: All		Purchases: Chat-based	Purchases: Admin-based		Purchases: Non-gender ID	
	(1) (48h)	(2) (24h)	(3) All	(4) (48h)	(5) (24h)	(6) (48h)	(7) (24h)
Female	-.038***	-.036***	-.018***	-.019*	-.017**	-.04***	-.038***
	(.013)	(.011)	(.0066)	(.0099)	(.0087)	(.013)	(.012)
Control Mean (wt)	.076	.070	.033	.043	.037	.078	.073
N	2655	2655	2655	2655	2655	2274	2274

This table shows the effect of female name assignment on purchase outcomes across purchase types and samples. All purchases refers to purchases from any source. Chat-based purchases refers to any purchases that were captured by the hand-coded chat data. Admin-based purchases refers to any purchase that were captured from sales data. Non-gender ID sample only includes observations from days when a customer did not use a gendered identifier in any chat to that agent. Gendered identifiers include: sir, maam, ma'am, brother, sister, miss. Purchases are measured within 48 or 24 hours of the start of the chat. Female indicator determined in customer's first chat of the day. Controls include agent-month fixed effects. The control group mean is reweighted by fixed effect cells to match the implied agent-month fixed effect OLS weights. Standard errors two-way clustered at the agent-day and customer-day level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: Effect of female assignment on purchase outcomes, robustness to alternative specifications

	Any purchases (48h)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	-.038***	-.044***	-.039***	-.035***	-.033***	-.031***	-.029**	-.025**
	(.013)	(.012)	(.013)	(.012)	(.011)	(.011)	(.012)	(.012)
Control Mean (wt)	.076	.076	.076	.076	.076	.066	.056	.062
Proportional effect	-.50	-.58	-.51	-.46	-.43	-.47	-.52	-.41
Agent-month FE	X	X	X	X	X			
Name ethnicity FE		X						
Customer controls			X					
DOW FE				X				
Week FE					X			
Agent FE						X		
Month FE						X		
Date FE							X	
N	2655	2654	2634	2655	2654	2655	2648	2655

This table shows the effect of female name assignment on any purchase within 48 hours in various specifications. Female indicator determined in customer's first chat of the day. Fixed effects and customer controls, for past purchases, customer location, and customer chat history, are included based on the column notes. Name ethnicity FE indicates fixed effects assigned based on the assigned full name. The control group mean is reweighted by fixed effect cells to match the implied fixed effect-only OLS weights: columns (1-4) reweight by agent-month, column (5) by agent, column (6) by date, and column (7) does not reweight. Standard errors two-way clustered at the agent-day and customer-day level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A8: Name features and interactions on purchase outcomes

	Any purchase (48h)					
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-.038***				-.04***	-.04***
	(.013)				(.013)	(.013)
Name age (decades)		-.0015			-.00043	-.0005
		(.0044)			(.0042)	(.0082)
Name nbhd % black			-.0058		-.026	-.09
			(.04)		(.045)	(.07)
Name nbhd % poverty				-.019	-.015	-.064
				(.049)	(.055)	(.057)
Female X age						.000083
						(.0095)
Female X % black						.12
						(.089)
Female X % poverty						.12
						(.084)
Control Mean (wt)	.076					
N	2655	2418	2418	2418	2418	2418

This table shows the effect of name assignment measured in various ways on purchase outcomes. The outcome across all columns is any purchase made within 48 hours of the start of the chat. Female indicator is determined by the treatment assignment and associated first name presented. Name age (decades) is the average age of the people with the assigned first name based on Ohio voter rolls, measured in decades. Name nbhd % black is the average share of the neighborhood who is black for people with this first name based on Ohio voter rolls and ACS records. Name nbhd % poverty is the average neighborhood share below the poverty line for people with this first name based on Ohio voter rolls and ACS records. Variables of the form Female X Variable are interactions between the female indicator and the listed name variable. Controls include agent-month fixed effects. The control group mean is reweighted by fixed effect cells to match the implied agent-month fixed effect OLS weights. Standard errors two-way clustered at the agent-day and customer-day level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A9: Effect of female name assignment and customer gender

	Any cust. gender	Any purchase (48h)			
	(1)	(2)	(3)	(4)	(5)
Female	-.0073	-.031***	-.094	-.12	-.095
	(.008)	(.011)	(.06)	(.13)	(.078)
Sample	Full	Full	Any cust. gender	Female cust.	Male cust.
Control Mean (wt)	.047	.066	.23	.098	.28
N	2655	2655	135	29	102

This table investigates female name assignment on purchases, incorporating information on customer name reveal. The outcome in column (1) is a binary variable equaling 1 for chats in which a customer reveals their name, while the outcome in columns (2-5) is any purchase made within 48 hours of the start of the chat. The sample in columns (2-5) is the full sample, the sample of chats in which customers reveal their gender, the female customer sample, and the male customer sample, respectively. Female indicator determined in customer's first chat of the day. Controls include agent and month fixed effects. Standard errors in parentheses and clustered at agent-day level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A10: Effect of female assignment on any sales by agent (48 hours)

	Purchases (48h)
	(1)
	Any
Female * Agent 1	-.02
	(.013)
Female * Agent 2	-.022
	(.038)
Female * Agent 3	-.014
	(.033)
Female * Agent 4	-.039***
	(.011)
Female * Agent 5	-.048**
	(.02)
Control Mean (wt)	.076
Joint *p*-value	.69
Proportional *p*-value	.58
N	2636

This table shows the effect of female name assignment on any purchase within 48 hours of the start of the chat by agent. Female indicator determined in customer's first chat of the day. Joint *p*-value tests equality of all coefficients. Proportional *p*-value tests equality of all effects proportional to the agent-specific control group mean. Controls include agent-month fixed effects. The control group mean is reweighted by fixed effect cells to match the implied agent-month fixed effect OLS weights. Standard errors two-way clustered at the agent-day and customer-day level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A11: Effect of female assignment on chat tones

	Tone					
	(1) Any	(2) Angry	(3) Happy	(4) Ecstatic	(5) Impatient	(6) Sad
Female	-.031**	.00047	-.021***	-.0016	-.018**	.0095*
	(.014)	(.0055)	(.0076)	(.0011)	(.0075)	(.0053)
Control Mean (wt)	.09	.01	.04	.00	.03	.01
N	1745	1745	1745	1745	1745	1745

This table shows the effect of female name assignment on chat tone outcomes. Outcomes measure either any tone, or any of the specific types of chat tones. Female indicator determined in customer's first chat of the day. Controls include agent-month and hand coder fixed effects. The control group mean is reweighted by fixed effect cells to match the implied agent-month fixed effect OLS weights. Standard errors two-way clustered at the agent-day and customer-day level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.