Customer Discrimination in the Workplace

Evidence from Online Sales

Erin Kelley
Gregory Lane
Matthew Pecenco
Edward Rubin

Abstract

Many workers are evaluated on their ability to engage with customers. This paper measures the impact of gender-based customer discrimination on the productivity of online sales agents working across Sub-Saharan Africa. Using a novel framework that randomly varies the gender of names presented to customers without changing worker behavior, we find that the assignment of a female-sounding name leads to 50 percent fewer purchases by customers. The results appear to be driven by relatively lower interest in engaging with female workers. Since worker productivity informs firm hiring, pay, and promotion decisions, these results are important for understanding the persistence of identity-based discrimination in the labor market.

Customer Discrimination in the Workplace: Evidence from Online Sales[*]

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

JEL Classification: J16; O12

Keywords: Labor, Discrimination, Gender

[†]Development Impact Evaluation Department (DIME), World Bank erinmkelley@worldbank.org

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

[§]Department of Economics, Brown University matthew_pecenco@brown.edu

[¶]Department of Economics, University of Oregon edwardr@uoregon.edu

1 Introduction

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. Since worker productivity informs firm hiring, pay, and promotion decisions, quantifying the magnitude of discriminatory customer behavior is important for understanding the persistence of identity-based discrimination in the labor market. It is also a relevant input for firms that may want to identify and promote the most talented workers rather than those least subject to bias, and for regulators who may wish to identify whether facially race-neutral policies like performance-based pay may actually promote discriminatory outcomes. Yet, we know very little about the magnitude of customer discrimination in the workplace.

Identifying the impact and scale of customer discrimination is challenging. First, many factors affect worker productivity—e.g., the worker's own skills and behavior, customers' behaviors and preferences, and the workplace environment—which makes it difficult to isolate the impact of any single determinant. Second, identifying discrimination typically involves testing whether wages or hiring outcomes are different across individuals *who are equally productive*. In doing so, these tests fail to account for discrimination that directly impacts productivity, such as consumer preferences. Consequently, they may erroneously conclude there is no discrimination at all.

This paper directly addresses these challenges to provide new evidence on the magnitude of gender-based discriminatory behavior by customers in the workplace. We do so by running a randomized experiment 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 customers who are mainly from other parts of the continent. We study workers who chat with customers online to answer questions and increase sales. This context allows us to measure worker productivity through customer purchases and rich patterns of engagement, including bargaining and harassment, through chat transcripts.

We apply a novel 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 they were chatting with.[1] Customers could only infer agents' gender from their names, as they 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 is unique and overcomes challenges with two common methods to study discrimination: audit and correspondence studies. Audit studies—where actors, who are as similar as possible except on one dimension, engage in a task like applying to the

---

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

same job—struggle to control for all other differences between the actors. They may also be subject to "demand effects" because actors are aware of their treatment status and do not have real incentives to perform well. Correspondence studies—where fictitious applications with different sounding names are sent to a possible discriminator like employers—can only measure coarse 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, the name masking alleviates concerns about demand effects, and by studying real workplace interactions we can collect detailed outcome measures—e.g., the likelihood of purchase—and specifics of the interaction—e.g., bargaining behavior.

While customer discrimination may be a global phenomenon, it may be particularly important in the context of this study for a number of reasons. First, the share of women employed in the service industry in Sub-Saharan Africa has increased exponentially in the last decade. Identifying the scale of customer discrimination in these occupations will be increasingly relevant for female labor force participation in the long run. Second, social norms in Sub-Saharan Africa favor men over women as economic agents and in business, which may contribute to higher rates of customer discrimination. For example, data from the World Bank World Development Indicators (WDI) show that households in Sub-Saharan Africa are more likely to agree that men make better business leaders and that women have no say in decisions on large household purchases (Jayachandran, 2015). Finally, it is particularly timely to study these issues as gender inequity in the workplace has become a central policy goal for governments across the continent and international institutions alike (Bank, 2011; O'Donnell et al., 2020).[2]

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 3.9 percentage points, or 51% relative to the baseline purchase rate (7.6%). We observe similarly large reductions in the total number of purchases and the total value of goods purchased.

These results suggest that productivity differences between workers in our context are not just a reflection of worker attributes but also of differential customer responses. While the literature has identified productivity differences as a key determinant of the gender wage gap (Sin et al., 2020; Gallen et al., 2017; Blau and Kahn, 2017; Caliendo et al., 2017), our results demonstrate that underlying customer-based discrimination can contribute to these productivity differences. This could lead decompositions of the gender wage gap to underappreciate the importance of discrimination and overestimate the role of productivity. This finding has important implications for policies such as "equal pay for equal work," which if assessed by productivity at work, may not account for the difficulties women face while doing the same job. While governments cannot regulate customers' discriminatory behaviors, they may regulate hiring and wage practices. For example, incentive-based pay schemes are commonly used in customer-facing roles, despite potentially exacerbating gender (or other

---

[2] For example, The African Union Strategy for Gender Equality includes the promotion of laws to achieve pay equality—as the gender wage gap is persistent across most industries where women work (ILO, 2019).

identity) differences in wages and promotions.

To confirm that our treatment effects result from the implied gender of agents' assigned names, we investigate whether customers are aware of the agents' names. In 11% of chats, customers mention the agent's assigned name—indicating names are a salient feature of the interaction.[3] Helpfully, we can also rule out a potential confound—customers 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 any mismatch comes from being assigned a male-sounding name.[4] Consequently, this "mismatch effect" would likely attenuate our estimated effect of discrimination against female agents. In sum, receiving a female name reduces productivity via consumers' gender preferences.

These reductions in worker productivity may be explained by several mechanisms: general customer disinterest in working with female agents, differential bargaining, or overtly negative interactions.[5] Data from agent-customer chat interactions suggest that customer disinterest is the most likely channel. We find that consumers respond more slowly to female agents, only responding after receiving additional messages from the agent.[6] This result suggests some consumers are hesitant to engage with female agents unless the agents persist through additional messaging. We also find that consumers are less likely to express any non-neutral conversational tone, which we interpret as another measure of engagement with the agent.

The data do not support other possible mechanisms. We find no evidence that consumers differentially bargain when agents receive female-sounding names. This result is interesting as bargaining (e.g., asking for discounts) is common in these interactions (occurring in 15% of conversations), and differences in bargaining by gender feature prominently in studies of wage gaps and job-application behavior (e.g., Card et al. (2016); Rousille (2021); Castillo et al. (2013)). We also find no differences in hostile or harassing behavior, although any form of harassment is rare in this context.

Our finding of customer bias does not directly indicate whether female workers would be more or less productive than their male co-workers after correcting for this bias. Discerning between these cases is informative about how the labor market functions. To investigate this, we compare other workers at the company—both female and male agents—outside of the experimental sample to compare how observable productivity differs by worker gender, inclusive of customer bias. We find average productivity is both economically and statistically indistinguishable across worker gender. This implies the female workers are actually *more* productive when customer bias is factored in. That women work in this industry despite the discrimination they face suggests they have a relative comparative advantage or derive non-pecuniary benefits in this occupation, or they face greater discrimination in the

---

[3] This information is not typically observed in studies of discrimination.

[4] More broadly, roughly two-thirds of the employees working in these roles at the company are female.

[5] These behaviors are consistent with various theories of discrimination, including statistical and taste-based discrimination. Differentiating between these theories is not the intent of this study.

[6] Agents always send the first message. The assigned name of the agent is revealed with this first message.

other jobs they might consider.

This paper makes three contributions. First, we provide causal evidence that customer discrimination lowers the measured productivity of female employees in the workplace by a meaningful margin. Studies on the impact of customer discrimination are scarce and find mixed evidence (Bar and Zussman, 2017; Combes et al., 2016; Holzer and Ihlanfeldt, 1998; Kahn and Sherer, 1988; Leonard et al., 2010; Nardinelli and Simon, 1990). Our study differs in two key ways. First, we use an experimental design to isolate customer discrimination rather than methodologies that may not fully control for differences in unobservable characteristics of workers and customers. Second, we can measure worker productivity rather than relying on wages or hiring outcomes. This provides new insights in how customer discrimination affects workers and for how tests of discrimination may be biased by assuming fixed worker productivity.[7] Our results also contribute to a growing literature that shows how upstream discriminatory processes can impact the final productivity measures often used for academic and policy analysis (Glover et al., 2017; Hengel, 2022; Parsons et al., 2011).

This paper is also related to a literature on 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).[8] While this literature has found more evidence of discrimination, the implications for *workers* in the labor market remain unclear for a number of reasons. First, List (2004) and Doleac and Stein (2013) find that customers statistically discriminate against minorities in markets for second-hand baseball cards and iPod sellers because they make assumptions about the product or seller's quality from the identity of the seller. We would not expect customers to statistically discriminate among workers who are selling the same products at the same company. This could explain why we do not find evidence of differential customer bargaining.[9] Second, engaging real workers at an established company means we can study customer discrimination in a real labor market setting, and we can test whether this discrimination exists among individuals who have selected into this job based on their comparative advantage.

Third, this paper documents how consumers' preferences can create meaningful barriers to women's success in labor markets in developing countries. Recent work documents a variety of other possible constraints: norms and bargaining dynamics within the household (Lowe and McKelway, 2021; Bursztyn et al., 2020; Dean and Jayachandran, 2019; Heath and Tan, 2020; Field et al., 2021; McKelway, 2021a,b); workplace attributes (Subramanian, 2021), safety during commutes (Borker, 2021), market demand (Hardy and Kagy, 2020), and employer discrimination (Jayachandran, 2015; Duflo, 2012; Sin et al., 2020). Most relatedly, 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. While similar in ob-

---

[7] Nardinelli and Simon (1990) is an exception as they define productivity of baseball players to be their entertainment value which they observe through the price of baseball cards.

[8] This paper also contributes to other literatures on bias in student evaluations of teachers (MacNell et al., 2015; Mengel et al., 2019) or by audience members for academic economists (Dupas et al., 2021).

[9] It is also hard to compare the magnitude of our estimates to this literature. For example, List (2004) and Doleac and Stein (2013) find reductions in 3-30% and 11% in 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.

jective, our study differs in two 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 (a disadvantage of this approach is that we did not choose the sales agent pool we study).

2 Context

2.1 Service sector in Sub-Saharan Africa

We study consumers' discriminatory behavior when engaging with online sales agents in Sub-Saharan Africa. As service-sector jobs increase, customer-facing roles are increasingly common across the continent. For example, the share of working-age individuals employed in services throughout Sub-Saharan Africa rose 12% from 2011 to 2019 (WDI). 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%. While one might assume that discrimination is lower in sectors where female labor force participation rates are high, women may choose to work in these sectors precisely because they have a comparative advantage in this space.

This trend will likely persist as internet connectivity spreads across the continent and service-sector jobs increasingly interface with clients online. In 2010, only 8.3% of the population in Africa had internet access. By 2017, internet access had increased to 22.3% (WDI). 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

We evaluate an experiment at an online travel agency with offices located across Sub-Saharan Africa—we work primarily with an East African field office. For confidentiality reasons, we cannot provide any identifiable information about the company. The company sells flights and hotels to customers that come from different parts of the continent (the average price of a flight/hotel conditional on making a sale is 110 euros). The company employs local sales agents who answer customer questions, field complaints, and encourage purchases (using phone calls and online chat interfaces).[10] In the company overall, approximately two-thirds of the sales agents are women, and these women account for 83% of chats. The experiment included six agents, all of whom were female. These sales agents are full-time company employees, and receive an annual wage.[11]

---

[10] Approximately 40-50% of chats with any customer response relate to purchases.

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

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 a short greeting message. 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 pulled one name per sales agent per day from an existing list. The agent then received the drawn name within the chat/sales 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 and ethnicity.[12] 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. Agents did not inquire about the plugin throughout the duration of the experiment. The plugin worked in the following way. Consider a day when agent *James* (real name) was assigned the name *Steve*. Whenever the customer typed "Steve" into the chat, James would only see "Agent" in his chat window. In contrast, the client would still see "Steve." This name masking included any references to the agent's assigned name in the chat transcript.[13]

The experiment was launched in January 2019 and the name assignment continued until October 2019. The company had full discretion over which agents participated in the study. They included all five agents based in the East African office, and one agent from the West Africa branch. The company informed agents that it was interested in learning more about how customers respond to different agents, and may change agents' display name in the chat. No further information about the nature or the objectives of the experiment was provided by the company, 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

---

[12] Of the 1,198 names, there are 1,196 unique full names, and 579 unique first names. Table A1 lists 20 example names, by gender. We do not have access to data sources like birth certificate records or a census to validate how common these names are for men or women.

[13] 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 section 4 shows the results are unaffected by excluding any days when this occurs.

any subsequent team meetings, and the focus on gender was never broached. 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 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.[14] The sales data were matched to the chats using date and customers' IP address.[15]

To measure overall purchases, we include purchases directly made by customers, which are observed in administrative sales records, and also purchases made by agents on behalf of customers. Agents may input customer details and purchase products on their behalf, which we capture by reading through the chat records and flagging instances where 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.

From the chat transcripts, we create objective and subjective outcome measures. 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—e.g., the overall tone, whether customers bargained with agents (e.g., asking for a discount), or whether customers harassed agents. Enumerators familiar with the cultural context hand-coded these subjective outcomes; 20% of the observations were double coded to ensure consistent measurement.

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. On days when agents responded to chats, they spent 2.9 hours on the online sales interface with customers, engaging in approximately 8 unique chat conversations per day. 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] 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).

[15] We restrict our sample to observations with five or fewer previous purchases as some users may have accessed the site using non-unique locations—e.g., public areas or businesses—and hence their purchase records were not well-linked. This restriction retains 98% of observations.

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.[16] 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 replacement). Given the selected gender, the randomization then chose a specific name from the name database. 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. We restrict our sample to weekdays when agents typically work regular schedules. Customers were allocated programmatically to agents; neither agents nor customers had any choice over who they are matched with.

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

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

where y_{iam} is the outcome of interest for customer i, working with agent a, in month m. The indicator $\mathbb{1}[\text{Assigned female}]_{ia}$ is 1 if agent a (matched to customer i) is assigned female on that date. The term γ_{am} represents agent by month-of-sample fixed effects—restricting comparisons within this grouping. We further control for customer characteristics, X_i, including country, past purchase history, and past chat history for precision. We can augment this regression specification to estimate individual-agent treatment effects, but we cannot estimate heterogeneity by customer gender or other demographics, as we do not observe them.[17]

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 in section 4.

---

[16] See section 2.2 for details.

[17] Note that the regression can analogously be run at the (grouped) agent-day level after employing a two-step regression procedure that matches 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.

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.

This approach contains two potential concerns for external validity. First, the name-masking procedure could affect agents' productivity if 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. Second, agents 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 will 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. All of the agents in our sample are women, and could only potentially 'confuse' a customer with their language when they are assigned a male-sounding name. However, because we find that being assigned a female name reduces the likelihood of a sale, any 'confusing' behavior from a male-sounding name may attenuate our estimates.

We provide a validation of the randomization procedure in Table A2. In column (3) of this table, we regress observable customer characteristics (e.g., number of past purchases) and agent characteristics (e.g., number of daily chats) on an indicator for whether the agent received a female name, controlling for agent-month fixed effects.[18] Female assignment does not correlate with any customer or agent characteristics at the 5% level. We fail to reject the joint null hypothesis that each of these effects is zero ($p = 0.65$). The table 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 a topic in the chat.

4 Results

4.1 Effect of name assignment

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 used agents' assigned names in 7% of all chats and 11% of chats in which consumers ever responded to agents'

---

[18] 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.

initial messages. We interpret this as a relatively high share of customer awareness as many chats are brief. Thus, agent names are indeed salient in chat interactions and could affect customers' behavior.

Table 1 presents the estimated effects of female-name assignment on outcomes related to customer purchases. 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.[19] We measure purchases in three ways: the probability of making any purchase, the number of distinct purchases, and the total price of purchases. As discussed in subsection 2.3, our measures of *any purchase* and *total number of purchases* include purchases made by customers and by agents on customers' behalf. In contrast, the total-price measure only includes purchases by customers.[20]

We find that consumers assigned to agents with female names are less likely to purchase products on the website. Column (1) shows that female-agent assignment decreases the probability that any purchase occurs (within 48 hours) by 3.9 percentage points ($p = 0.002$). The likelihood that a chat results in any purchase in the control group (male-sounding names) is only 7.6%.[21] Thus, the point estimate implies a 51% reduction in the likelihood of making a sale. Column (2) shows that consumers also purchase 0.039 fewer total products ($p = 0.004$) when interacting with female-sounding names; column (3) shows that the total value of purchases falls by 3.5 euros. Columns (4-6) repeat the same outcomes but use a 24-hour window after the chat. The results are very similar.[22]

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 (Sin et al., 2020; Gallen et al., 2017; Blau and Kahn, 2017; Caliendo et al., 2017). 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. While we cannot speak to optimal policy responses, regulation that prevents employer discrimination could limit some consequences of customer discrimination.

An advantage of our study design, which assigns each agent to both treatment sta-

---

[19] Statistical power is also likely to be higher in the period directly after these focal events.

[20] Interpretation of the total-price measures, based on only administrative customer-purchase records, is more challenging due to the mode of purchase potentially being endogenous to treatment status. 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 A5). This means we very likely *underestimate* the coefficient on total-price, although the effect relative to the baseline mean is not clear.

[21] 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, $\beta^{OLS} = \sum_{c \in \mathscr{C}} (E[Y|C = c, D = 1] - E[Y|C = c, D = 0])w_c$. The reweighting does not account for additional customer controls.

[22] 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 ($p = .377$).

tuses (female and male) over time, is that we can test whether treatment effects vary across agents. This allows us to check (1) if our treatment effects are being driven by only one or two agents, and (2) whether the magnitude of negative treatment effects are meaningfully different across agents. To estimate agent-specific treatment effects, we augment the baseline model by interacting treatment with agent-specific indicators. Agent-specific estimates may reflect differences in agent characteristics and/or the types of consumers that agents encounter—since the study's design does not randomize customers across agents. While the estimated effect of female-name assignment is negative for all agents (except one, whose positive coefficient is not statistically different from zero), we can reject that the treatment effects are the same across all agents ($p = 0.018$).[23] This result suggests that the impacts of consumer-based gender discrimination on productivity (sales) likely differ across agent and/or consumer types. Therefore, any policies that attempt to compensate employees for customer discrimination would likely need to take this heterogeneity into account.

Our results are robust to various analysis choices. In particular, they are almost identical when we aggregate to the customer-day level (Table A3) or agent-day level (Table A4)—and are statistically significant and of similar magnitude when looking at either purchases from chat-based records (via agents) or from administrative records (via customers) (Table A5). Table A6 shows our main results are robust to including day-of-week or week fixed effects, removing customer controls, changing fixed effects to agent and month, only including date fixed effects, or removing all fixed effects and controls—the effect of female status leads to proportional reductions of 40-52% across all specifications. Table A7 shows our results are unaffected by excluding any of the relatively few days when customers use a gendered identifier (e.g. Sir or Miss).

One potential concern with varying name assignment is that we are actually measuring a factor correlated with gender, but it is not actually gender. The most salient other feature in this context is ethnicity, which a priori is unlikely to be important since only first names are shown. To further support this, Table A8 shows that our results are unaffected by directly controlling for name ethnicity as fixed effects, alleviating this concern.[24] Even more, the name ethnicity is not correlated to the likelihood of any sale within 48 hours ($F = 1.37, p = 0.16$).

4.2 Mechanisms

There are many reasons why purchases may fall when consumers chat with female agents. Our data allow us to explore three potential mechanisms. First, customers may be hesitant to engage with female sales agents because of taste-based or statistical discrimination. 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—likely harming their productivity (Georgieva, 2018; Du-

---

[23] Table A9 shows these agent-specific treatment effects.

[24] 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.

pas et al., 2021; Folke and Rickne, 2020). Finally, an extensive literature documents women and men may face different bargaining processes (Ashraf, 2009; Rousille, 2021; Vesterlund, 2018; Castillo et al., 2013; Card et al., 2016)—a fact customers may attempt to exploit by bargaining more with female sales agents.

We first explore whether customers are hesitant to engage with female agents. We investigate this along two dimensions. On the extensive margin, whether the customer engages with an agent at all, some consumers may be hesitant to chat with female agents or may entirely avoid female agents. On the intensive margin, consumers may engage differently by using different tones when they chat with female agents. 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 many 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 (1-3) of Table 2 show the effect of female-name assignment on *extensive margin* consumer interactions. Mechanically, agents always send the first message; the conversations begin there. In column (1), female assignment leads to a negative but statistically insignificant effect on the likelihood the customer ever responds ($p = 0.184$).

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 ($p = 0.022$), suggesting lower customer engagement (higher hesitance). Finally, we test engagement using the number of messages the customer sends in their response to the agent. Column (3) shows that consumers send fewer messages when initiating a conversation with an agent with a female-sounding name ($p = 0.059$). Together, these three measures suggest that some consumers may hesitate to engage with female agents.

We investigate customer hesitancy along the *intensive margin* by analyzing the conversations' tones. While specific tones are likely imperfect proxies for genuine emotions, whether a tone exists may reflect a customer's level of engagement with the agent.[25] To this end, we construct a measure for any non-neutral tone detected in the conversation. Column (1) of Table 3 demonstrates a 2.7 percentage point reduction in the probability of any tone when customers engage with female-assigned agents, a 31% reduction relative to the control-group mean ($p = 0.053$). This result again suggests that customers exhibit weaker levels of engagement with female-assigned agents, echoing our extensive-margin findings.

The second possible mechanism—customers are more abusive toward women—is motivated by a growing literature that documents high rates of harassment for women in the workplace. However, the results in column (2) of Table 3 suggest this mechanism does not explain the differences in sales in our setting. The outcome measures whether any language is classified as harassment within the chat. The data contain few instances of harassment interactions: 0.3% of conversations for the male-assigned (control) sample indicate harass-

---

[25] Table A10 also presents results for each tone separately.

ment. The rate in the female-assigned sample is practically identical to the male-assigned sample and does not differ statistically.

Finally, column (3) of Table 3 tests whether customers bargain more often with female sales agents. While 15% 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 is fairly precise relative to the baseline level and rules out changes to bargaining that exceed 20%. Therefore, in this context, differential bargaining does not appear to drive the observed productivity differences.

Together, our results suggest that customers interact differently with women and men in ways that can meaningfully reduce productivity. This result is especially consequential for the service industry, where customer-facing roles abound. Our investigation of the mechanisms behind this behavior suggests that consumers engage less with female agents—along extensive (any engagement) and intensive (tone used) margins. We find no evidence that customers differentially harass or bargain with women in our setting.[26]

4.3   Comparison to non-experimental estimates

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 *outside* the experimental sample.[27] We include similar controls in both specifications but cannot include agent fixed effects in the non-experimental comparison as they are collinear with gender.

Table 4 shows correlations between agents' actual gender and sales in Panel A and experimental estimates of female name assignment in Panel B. Only administrative sales records are available for the non-experimental sample, so we compare across samples using these outcomes. In Panel A, we find no statistically significant differences between male and female agents across any of the three purchase outcomes. In Panel B, we find female name assignment leads to statistically significant reductions in sales across all of the administrative-record purchase outcomes.[28] 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.

The difference between the experiment and correlational estimates sheds light on gender-based selection into this occupation. In particular, it suggests women in these jobs may be

---

[26] We are unable to test for other mechanisms—e.g., homophily (a customer's preference to interact with an agent of the same gender)—because we lack sufficient customer information. That means we can identify the absolute customer bias workers face in their actual job, which is the key parameter of interest when considering the impact of customer discrimination. We cannot test for relative customer bias depending on customer characteristics which has been a focus of other papers (Combes et al., 2016; Leonard et al., 2010; Bar and Zussman, 2017).

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

[28] The coefficients are smaller than in Table 1 since they do not also include agent-initiated purchases.

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. The fact that workers are sorting into jobs where they face this important source of discrimination suggests they derive other non-pecuniary benefits from this job, or they face greater discrimination in other jobs.

5 Conclusion

This paper demonstrates that customer-based discrimination can negatively impact 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 falls. An exploration of potential mechanisms suggests these results are most consistent with customer disinterest in working with female agents—rather than differential bargaining or openly hostile behavior. The magnitude of these results is especially consequential because customer-based discrimination is both unregulated (Bartlett and Gulati, 2016), and is unlikely to be competed away by market competition (Becker et al., 1971).

Our results have several implications. First, they suggest that female sales agents in our context may be more productive than their male counterparts when holding customer behavior constant. This result speaks to the "twice as hard" phenomenon whereby members of a discriminated group need to perform better than their counterparts in order to maintain their position in the workplace (Sofoluke and Sofoluke, 2021).

Second, these results indicate that *equal-pay-for-equal-work* policies may not fully resolve discrimination's effects when workers face discrimination from customers. This outcome is particularly relevant to the service industry, which often ties employee pay to productivity/output (e.g., number of sales) through piece-rate wages. Our results suggest that individual-based incentivized pay schemes may increase the impact of customer discrimination on worker pay.[29] 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. Alternatively, companies with pro-social intentions could endeavor to sensitize customers to female workers with the exclusive use of female-sounding names.

Finally, the results speak to specific barriers women face in the labor market in Sub-Saharan Africa, and contribute to an ongoing policy dialogue about workplace gender equality in low-income countries. Like many regions around the world, 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,

---

[29] For example, employers could 'pool' performance-based bonuses—a common practice for sharing tips in the restaurant industry.

2016). Numerous economic models and empirical studies suggest that improvements in gender parity can drive substantial economic growth (World Economic, 2017). Recognizing this potential, governments throughout the continent are leading initiatives to address equity issues—including powerful provisions that support gender equality (e.g., USAID). While companies and institutions have attempted to equalize opportunities for women and men (e.g., better wages and flexible hours (World Bank, 2013)), customer discrimination can still significantly affect women's productivity in the workplace. Our results provide an opportunity to design new solutions that mitigate the effect of customer preferences.

References

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.

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

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

BARTLETT, K. T. AND M. GULATI (2016): "Discrimination by Customers," https://ilr.law.uiowa.edu/print/volume-102-issue-1/discrimination-by-customers/.

BECKER, G. S. ET AL. (1971): "The Economics of Discrimination," *University of Chicago Press Economics Books*.

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

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.

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*.

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.

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," .

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.

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. K. RICKNE (2020): "Sexual harassment and gender inequality in the labor market," .

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.

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.

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.

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

ILO (2019): "Wages in Africa," .

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.

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.

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*.

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.

SOFOLUKE, R. AND O. SOFOLUKE (2021): *Twice As Hard: Navigating Black Stereotypes and Creating Space For Success*, New York, NY: DK.

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.

USAID (????): "Gender Equality and Women's Empowerment in Kenya | Kenya | Archive - U.S. Agency for International Development," https://2012-2017.usaid.gov/kenya/gender-equality-and-womens-empowerment-kenya.

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

WORLD BANK, G. (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.

WORLD ECONOMIC, F. (2017): "The Global Gender Gap Report 2016 - World," https://reliefweb.int/report/world/global-gender-gap-report-2016.

6 Tables

	Purchases (48h)			Purchases (24h)		
	(1) Any	(2) Total	(3) Total price	(4) Any	(5) Total	(6) Total price
Female	-.039***	-.039***	-3.5***	-.037***	-.036***	-3.4***
	(.012)	(.013)	(1.2)	(.011)	(.012)	(1.2)
Control Mean (wt)	.076	.081	5.307	.070	.073	4.891
N	2653	2653	2653	2653	2653	2653

This table shows the effect of female name assignment on purchase outcomes. Any represents any purchase, Total represents number of purchases, and Total price is the cumulative price of all purchases in EUR. Any purchases and total purchases combine purchases by customer and by agent, while total price is based only on customer purchases. 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, customer location, customer purchase history, and customer chat history fixed effects. The control group mean is reweighted by fixed effects cells to match the implied agent-month fixed effects-only 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 2: Effect of female assignment on conversation response

	Initial messages		
	(1) Ever respond (C)	(2) Msgs to response (A)	(3) Msgs to response (C)
Female	-.027	.0094**	-.11*
	(.02)	(.0041)	(.059)
Control Mean (wt)	.676	1.009	1.332
N	2653	2653	2653

This table shows the effect of female name assignment on customer and agent responses. Ever respond (C) is a 1 if the customer ever responded. Msgs to response (A) is the number of messages sent by agent before customer first response. Msgs to response (C) is the number of messages by customer in intial response. Female indicator determined in customer's first chat of the day. Controls include agent-month, customer location, customer purchase history, and customer chat history fixed effects. The control group mean is reweighted by fixed effects cells to match the implied agent-month fixed effects-only 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 assignment on chat outcomes

	Tone	Negativity		Bargaining
	(1)	(2)	(3)	(4)
	Any	Harass	Any neg.	Any
Female	-.027*	.00018	.0064	.0081
	(.014)	(.0028)	(.0041)	(.018)
Control Mean (wt)	.087	.003	.008	.141
N	1742	1742	1742	1742

This table shows the effect of female name assignment on chat outcomes. Column (1) measures any non-neutral chat tone, column (2) measures any harassment of the agent, column (3) measures any negative words or phrases, and column (4) measures any bargaining. The sample include only chats with any consumer response. Female indicator determined in customer's first chat of the day. Controls include agent-month, customer location, customer purchase history, customer chat history, and hand coder fixed effects. The control group mean is reweighted by fixed effects cells to match the implied agent-month fixed effects-only 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

	Sales (48h)			Sales (24h)		
	Any (1)	Total (2)	Total price (3)	Any (4)	Total (5)	Total price (6)
Panel A: Correlational estimates						
Female	0.003	0.005	-1.577	0.002	0.001	-2.110
	(0.008)	(0.008)	(1.542)	(0.007)	(0.008)	(1.508)
Control Mean (wt)	0.024	0.026	4.644	0.021	0.022	4.416
Observations	8,863	8,863	8,863	8,863	8,863	8,863
Panel B: Experimental estimates						
Female	-0.020**	-0.022**	-3.495***	-0.018**	-0.018**	-3.432***
	(0.010)	(0.011)	(1.240)	(0.009)	(0.009)	(1.166)
Control Mean (wt)	0.030	0.030	4.898	0.030	0.030	4.639
Observations	2,653	2,653	2,653	2,653	2,653	2,653
Equality b/w Exp/Non-exp (p)	0.047	0.044	0.324	0.067	0.094	0.481

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. All outcomes based on customer purchases only. Sales are measured within 24 or 48 hours of the start of the chat. Female indicator in Panel B determined in customer's first chat of the day. Controls include month, customer location, customer purchase history, and customer chat history fixed effects. The experimental estimates additionally include agent fixed effects. The control group means are reweighted to match the OLS weights. For the correlational estimates, the control mean is reweighted by month fixed effects, while in the experimental estimates it is reweighted by agent-month fixed effects. Standard errors two-way clustered at the agent-day and customer-day level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Online Appendix

A1 Tables

Table A1: Example names used in name assignment, by gender

Female	Male
Nasiru	Bakare
Habu	Jimada
Teresia	Ebiason
Margaret	Benjamin
Kilel	Stephen
Adaeze	Yahaya
Caroline	Elias
Evaline	Mudassir
Pamela	Amir
Lydia	Taofeek
Ariel	Tom
Esther	Ombache
Annah	Desmond
Catherine	Immanuel
Sekinat	Lawrence
Annmarie	Edwin
Mwanamisi	Fredrick
Salihu	Aliyu
Justina	Kingsley
Monica	Evans

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.

Table A2: Placebo tests for female assignment

	N	Var. Mean	Female
Customer mention agent true name	2655	.00	-.00128
			(.00234)
Customer amount of past chats	2655	.11	-.046*
			(.0251)
Customer amount of past purchases	2655	.29	-.0166
			(.0458)
Agent first message length	2655	5.47	-.00066
			(.0041)
Agent chats (daily)	337	7.76	-.106
			(.596)
Agent hours worked (daily)	337	2.57	-.0291
			(.172)
Joint *p*-value			.64

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 and hours worked by agents are at the day level, while the other variables are at the chat level. Controls include agent-month fixed effects. Female indicator determined in customer's first chat of the day. 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 A3: Effect of female assignment on purchase outcomes (customer-day level)

	Purchases (48h)			Purchases (24h)		
	(1) Any	(2) Total	(3) Total price	(4) Any	(5) Total	(6) Total price
Female	-.04***	-.042***	-3.2***	-.039***	-.038***	-3.1***
	(.012)	(.014)	(1.2)	(.011)	(.013)	(1.2)
Control Mean (wt)	.077	.083	5.263	.072	.074	4.808
N	2169	2169	2169	2169	2169	2169

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, Total represents number of purchases, and Total price is the cumulative price of all purchases in EUR. Any purchases and total purchases combine purchases by customer and by agent, while total price is based only on customer purchases. 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, customer location, customer purchase history, and customer chat history fixed effects. The control group mean is reweighted by fixed effects cells to match the implied agent-month fixed effects-only OLS weights. Standard errors clustered at the agent-day level. $^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$.

Table A4: Effect of female assignment on purchase outcomes (agent-day level)

	Purchases (48h)			Purchases (24h)		
	(1) Any	(2) Total	(3) Total price	(4) Any	(5) Total	(6) Total price
Female	-.04***	-.042***	-3.2**	-.038***	-.038***	-3.1***
	(.013)	(.014)	(1.3)	(.012)	(.013)	(1.2)
Control Mean (wt)	.059	.064	4.286	.057	.060	4.094
N	335	335	335	335	335	335

This table shows the effect of female name assignment on purchase outcomes. Data is at the agent-day level. Any represents any purchase, Total represents number of purchases, and Total price is the cumulative price of all purchases in EUR. Any purchases and total purchases combine purchases by customer and by agent, while total price is based only on customer purchases. 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 while a first-step regression included these and additionally customer location, customer purchase history, and customer chat history fixed effects. The outcome is the coefficient on agent-day from the first step-regression. The control group mean is reweighted by fixed effects cells to match the implied agent-month fixed effects-only 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 chat-based and admin-based purchase outcomes

	Any chat-based	Any admin-based		Combined purchases	
	(1)	(2)	(3)	(4)	(5)
	All	(48h)	(24h)	(48h)	(24h)
Female	-.018***	-.02**	-.018**	-.039***	-.037***
	(.0067)	(.0097)	(.0087)	(.012)	(.011)
Control Mean (wt)	.033	.043	.037	.076	.070
N	2653	2653	2653	2653	2653

This table shows the effect of female name assignment on purchase outcomes for administrative, non-administrative, and combined purchases. Non-administrative purchase records are recovered from chats, as they are purchases made by agents for customers. Administrative purchases are purchases made by customers observed through administrative records. 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, customer location, customer purchase history, and customer chat history fixed effects. The control group mean is reweighted by fixed effects cells to match the implied agent-month fixed effects-only 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 A6: Effect of female assignment on purchase outcomes, robustness to alternative specifications

	Any purchases (48h)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	-.039***	-.038***	-.036***	-.034***	-.031***	-.029**	-.025**
	(.012)	(.013)	(.012)	(.011)	(.011)	(.012)	(.012)
Control Mean (wt)	.076	.076	.076	.076	.066	.056	.062
Proportional effect	-.51	-.50	-.48	-.45	-.47	-.52	-.41
Agent-month FE	X	X	X	X			
Agent FE						X	
Month FE						X	
Controls	X		X	X			
DOW FE			X				
Week FE				X			
Date FE							X
N	2653	2655	2653	2652	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 controls, for past purchases, customer location, and customer chat history, are included based on the column notes. The control group mean is reweighted by fixed effects cells to match the implied fixed effects-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 A7: Effect of female assignment on purchase outcomes, overall and without days with gendered language

	Purchases (48h)	
	(1)	(2)
	Full sample	Non-gender ID sample
Female	-.039***	-.041***
	(.012)	(.013)
Control Mean (wt)	.076	.078
N	2653	2273

This table shows the effect of female name assignment on any purchase within 48 hours. Column (1) uses the full sample, while column (2) only includes observations from days when a customer did not use a gendered identifier in any chat. Gendered identifiers include: sir, maam, ma'am, brother, sister, miss. Female indicator determined in customer's first chat of the day. Controls include agent-month, customer location, customer purchase history, and customer chat history fixed effects. The control group mean is reweighted by fixed effects cells to match the implied agent-month fixed effects-only 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 A8: Effect of female assignment on purchase outcomes, controlling for name ethnicity

	Purchases (48h)			Purchases (24h)		
	(1) Any	(2) Total	(3) Total price	(4) Any	(5) Total	(6) Total price
Female	-.045***	-.047***	-4***	-.042***	-.042***	-3.9***
	(.012)	(.013)	(1.3)	(.011)	(.012)	(1.2)
Control Mean (wt)	.076	.081	5.307	.070	.073	4.891
N	2652	2652	2652	2652	2652	2652

This table shows the effect of female name assignment on purchase outcomes, controlling for the ethnicity of assigned names. Ethnicity is assigned based on the full name, although only the first name is shown. Any represents any purchase, Total represents number of purchases, and Total price is the cumulative price of all purchases in EUR. Any purchases and total purchases combine purchases by customer and by agent, while total price is based only on customer purchases. 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 name-implied ethnicity, agent-month, customer location, customer purchase history, and customer chat history fixed effects. The control group mean is reweighted by fixed effects cells to match the implied agent-month fixed effects-only 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 assignment on any sales by agent (48 hours)

	Purchases (48h)
	(1)
	Any
Female * Agent 1	-.014
	(.014)
Female * Agent 2	-.023
	(.039)
Female * Agent 3	-.017
	(.033)
Female * Agent 4	-.04***
	(.011)
Female * Agent 5	.0058
	(.0099)
Female * Agent 6	-.05**
	(.019)
Control Mean (wt)	.076
Joint *p*-value	.02
N	2653

This table shows the effect of female name assignment on purchase outcomes by agent. Any represents any purchase. Purchases are measured within 48 hours of the start of the chat. Female indicator determined in customer's first chat of the day. Joint *p*-value tests equality of all coefficients. Controls include agent-month, customer location, customer purchase history, and customer chat history fixed effects. The control group mean is reweighted by fixed effects cells to match the implied agent-month fixed effects-only 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 A10: Effect of female assignment on chat tones

	Tone					
	(1) Any	(2) Angry	(3) Happy	(4) Ecstatic	(5) Impatient	(6) Sad
Female	-.027*	.002	-.02***	-.0017	-.017**	.01**
	(.014)	(.0053)	(.0076)	(.0011)	(.0074)	(.0051)
Control Mean (wt)	.09	.01	.04	.00	.03	.01
N	1742	1742	1742	1742	1742	1742

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, customer location, customer purchase history, customer chat history, and hand coder fixed effects. The control group mean is reweighted by fixed effects cells to match the implied agent-month fixed effects-only 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: 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 (A)	Number of messages sent by agent before customer first response.
Messages to Response (C)	Number of messages sent by customers in their initial response. Customers that never respond are coded as 0.

Tone

	We employed two research assistants (RA) based in Sub-Saharan Africa to read through all of the chats. We used a double-blind process so that 20% of all chats were reviewed by both assistants. Any discrepancies in how questions were being coded were flagged early in the process to streamline coding styles. We chose to hand code these outcomes as opposed to using natural language processing (which we attempted) for three reasons. First, the vast majority of interactions include many "polite" words such as "Thanks" or "Please", which meant many conversations were coded as friendly by the machine learning algorithm even if they were acrimonious. Second, the chats contain a large number of misspellings and chat shorthand, which are not included in language databases. Finally, we thought it best to have individuals who are familiar with the cultural context interpreting the tone of the conversation.
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 negative	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.