She? The Role of Perceived Agent Gender in Social Media Customer Service

Junyuan Ke
University of Rochester, junyuan.ke@simon.rochester.edu

Yang Gao
Singapore Management University, ygao1@illinois.edu

Shujing Sun
University of Texas at Dallas, shujing.sun@utdallas.edu

Huaxia Rui
University of Rochester, huaxia.rui@simon.rochester.edu

She? The Role of Perceived Agent Gender in Social Media Customer Service

Completed Research Paper

Junyuan Ke
University of Rochester
Junyuan.Ke@Simon.Rochester.edu

Yang Gao
University of Illinois Urbana-Champaign
Ygao1@Illinois.edu

Huaxia Rui
University of Rochester
Huaxia.Rui@Simon.Rochester.edu

Shujing Sun
The University of Texas at Dallas
Shujing.Sun@UTDallas.edu

Abstract

This paper investigated the role of perceived agent gender in customer behavior using a unique dataset from Southwest Airlines' Twitter account. We inferred agent gender based on the first names provided by agents when responding to customers. We measured customer behavior using three outcomes: whether a customer decided to continue the service conversation upon receiving an agent's initial response as well as the valence and arousal levels in their second tweet if the customer chose to continue the interaction. Our identification strategy relied on the Backdoor Criterion and hinged on the assumption that customer service requests are assigned to the next available agent, independent of agent gender. The findings revealed that customers were more likely to continue interactions with female agents than male agents and they were more negative in valence but less intense in arousal with the former group than with the latter.

Keywords: gender, stereotype, customer service, social media

Introduction

Imagine that you complained to an airline's Twitter account about a recent flight experience and received a tweet in response from a customer service agent named *Andrew*. How would you carry on the conversation? Now imagine that, instead of *Andrew*, the response was signed with the name *Alice*. Would your reaction be somewhat different simply because of the genders implied by these names? Would you be more or less likely to continue the conversation? If you decided to continue communicating, would your next message be more negative or perhaps less intense?

These are not just intriguing questions for scientifically-minded researchers; they are also in the minds of real-life practitioners. For example, to understand how gender affects support interactions, Olark, a live chat software vendor, conducted a gender experiment[1] in which *Sarah Betts*, a customer service veteran, changed her name to *Samuel* and used an obviously male portrait in her profile during a one-month period of live chatting with customers. Even with this rudimentary design to directly compare various interaction scores, the company was able to identify different behavioral patterns when customers interacted with agents of different perceived genders. In particular, the experiment revealed that customers were more satisfied with yet more abusive towards *Sarah* than *Samuel*.

[1] See https://blog.olark.com/live-chat-gender-equality-experiment.

Of course, these analyses and findings need to be rigorously evaluated and carefully interpreted, but why should we expect gender perceptions to influence a customer's interaction with an agent? These influences can be explained by at least two mechanisms, which are rooted in theories and evidence from psychology. The first mechanism involves perceived gender differences in terms of empathy and helping behaviors. Women are generally perceived as more empathetic than men (Hoffman 1977) and often rated as more helpful, kind, and compassionate than men (Bem 1974). The second mechanism concerns occupational stereotype, the well-known tendency to view an occupation dominated by one gender as better suited to the characteristics of that gender (Basow 1992; Fiske and Linville 1980). In the customer service industry, women have historically dominated men in terms of number (Steiger and Wardell 1995). These two mechanisms may lead to a natural preference for female agents, which could result in a higher probability that customers continue interacting with such agents.

Moreover, subsequent messages from a customer may be subtly influenced by perceived gender differences. For example, some customers may strategically exaggerate or emphasize their negative experiences in messages to female agents to obtain better redress due to the expectation that women are more compliant and softhearted than men (Del Boca and Ashmore 1980; Eagly 1983). In addition, customers may be more willing to share their distress with women than with men due to the perception that the former is more likely to resonate with and provide emotional support to them. In either case, female agents may end up receiving messages filled with more negative valence. Meanwhile, the public nature[2] of social media customer service and cultural expectations regarding the manner toward and treatment of women may regulate the emotional intensity expressed in messages to female agents, such as through the use or avoidance of a ranting style or abusive words.

To test these theoretical predictions and shed further light on how gender affects customer support interactions, we utilized a public dataset consisting of all customer service interactions on Twitter between Southwest Airlines and its customers from March 2018 to September 2019. During this period, customer service agents added their first names to their response messages, which allowed us to infer the gender of these agents. We developed three outcome variables to capture whether each customer continued the interaction after the agent's gender could be inferred as well as the emotional state of customers who continued engaging with agents. To characterize a customer's emotional state, we drew upon the circumplex model of affect in emotion science in which each emotional state is decomposed into the valence dimension and the arousal (or intensity) dimension. Our identification strategy was based on the Backdoor Criterion (Pearl 2009) and hinged on two important observations. First, the assignment of customers—or, more accurately, customer service requests—to agents was largely random because a customer service request was usually assigned to the next available agent without considering agent gender. This assumption is supported both by anecdotal evidence and the well-balanced characteristics of customers handled by male and female agents. Second, since tweets are short, agents' initial responses were often quite standard, making it less challenging to control for content variation in those responses, which may have also influenced the outcome variables. Empirical analyses using various content control methods revealed that customers were more likely to continue interacting with female agents than with male agents and that their messages were more negative in valence but less intense in arousal. These results are consistent with our theoretical predictions and offer valuable insights both to practitioners and to academics.

Literature Review

Due to the dominant percentage of female agents in the customer service industry (Olivetti and Petrongolo 2016; Rendall 2018; Steiger and Wardell 1995), customers' agent gender preferences have been well recognized by academia and industry as playing an important role in offline customer service (Foster and Resnick 2013). Fischer et al. (1997)'s seminal work documented two possible causes of customers' gender preferences in offline customer service. First, according to *gender stereotype bias* or *gender congruence bias*, occupations dominated by a specific gender will be better suited to people with the characteristics and skills of that gender. Second, according to *in-group bias*, customers expect agents of the same gender as themselves to provide better service than agents of the opposite gender.

This stream of literature primarily focuses on the offline retailing context in which face-to-face interactions are most prevalent. The effect of gender congruence bias on customers' evaluation crucially depends on

[2] Please note that, unlike Olark's experiment, which was based on private chats, social media customer service is public.

service quality conditions (Luoh and Tsaur 2007). Under unfavorable service quality conditions (e.g., service failures), female agents receive lower evaluations from customers than male agents (Snipes et al. 2006). For example, using laboratory experiments, Hekman et al. (2010) demonstrated that, when handling service failures, agents who were nonwhite and female usually received lower aggregated customer satisfaction evaluations. However, under favorable service quality conditions, customers tend to perceive higher service quality from female agents than male ones (Luoh and Tsaur 2007). For example, Otterbring et al. (2021) found that female agents received more favorable customer evaluations and word-of-mouth (WOM) ratings when the consumption context was more feminine. Furthermore, based on a field experiment within the hotel industry, Kim et al. (2021) concluded that gender congruence bias is asymmetric. They found that customers were less satisfied and less willing to revisit the hotel when female agents failed in male-dominated jobs; however, their satisfaction and intention to revisit were unaffected when male agents failed in female-dominated jobs.

The findings on the in-group bias in offline customer service are mixed. Mohr and Henson (1996) showed that customers evaluated agents more positively if they were the same gender as the customers rather than the opposite one. More importantly, their results suggested that the effect of in-group bias was strongest for female and gender-aschematic customers. However, in a study of phone-based service encounters in which agent identity was inferred from verbal cues, Moshavi (2004) found that, although customers were equally satisfied with male and female service agents, they were more satisfied with agents of the opposite gender than those of the same gender.

Compared to the offline customer service literature, the role of agent gender has thus far not been studied in the online customer service literature. Unlike in offline settings (e.g., face-to-face interactions, phone calls) where customers have rich visual or audio cues to infer agent identity, agent identity cues are often limited in online settings, especially in text-based customer service. Very recently, firms started requiring agents to include identity cues (e.g., personal signatures, personal profiles) in their responses to service requests on social media, while emerging studies have been looking at the effect of such identity disclosure on customers' willingness to complain as well as outcome of customer service provisions (Cheng and Pan 2021; Gao et al. 2023). For example, Gao et al. (2023) demonstrated that including identity cues in social media customer service significantly increased customers' willingness to engage with an agent and the likelihood of reaching a resolution. Thanks to this timely new development, the present study closes the research gap by examining the role of agent gender in the delivery of customer service on Twitter. Note there is an interesting recent study (Proserpio et al. 2021) that demonstrated the effect of customer gender on management response on an online review platform. Specifically, they showed that firms' management responses to a self-identified female customer were more confrontational or more aggressive or tried to discredit the reviewer. Such discrimination was found to discourage self-identified female reviewers from posting negative reviews in the future.

Hypothesis Development

Agent Gender and Customer's Willingness to Engage

Customer service is an act of helping and emotionally supporting customers. The social psychological literature has long argued that gender differences in helping behavior may shape customers' expectations about the help they receive from male and female agents. Psychologists have argued that the norms regulating helping behavior differ significantly in female and male gender roles. Women are generally perceived as more empathetic and sympathetic than men (Hoffman 1977), and they are expected to facilitate the progress of others toward their goals by caring about their personal and emotional needs. Indeed, women are often rated as more helpful, kind, and compassionate than men (Bem 1974), and such attributes are considered more desirable in women than men. Meanwhile, the male gender role in helping behavior has traditionally emphasized saving others from harm at the risk of oneself, which, in most cultures, is an act of heroism. The expected gender differences in helping behavior are also consistent with the disproportionately large percentage of females in occupations that emphasize some form of assisting others, such as nurses and teachers. In our research context, women have historically held the majority of customer service positions (Steiger and Wardell 1995). Until recently, nearly 69.5% of the over 1.6 million customer service representatives in the United States (US) were female.[3]

Even in the absence of actual quality differences in the customer services delivered by agents of different genders, such a prevalence of female agents in the customer service industry may have led to the stereotype that females are better at delivering customer service. Indeed, an abundance of research (e.g., Basow (1992)) has suggested that an occupation dominated by one gender is stereotyped as being better suited to the characteristics of that gender; this conclusion is rooted in cognitive theories (e.g., Fiske and Linville (1980)) that have been used to explain stereotypes in general.

In summary, regardless of whether there is any difference in customer services delivered by agents of different genders, customers may prefer female agents due to perceived gender differences in helping behavior or occupational stereotypes. Hence, we propose the following hypothesis for empirical testing.

Hypothesis 1: *After the initial exchange of messages, customers are more likely to continue conversations with female agents than with male agents.*

Agent Gender and Customer Emotion

According to the classical theory of affect, human emotion is a discrete set of basic emotions that are activated through a distinct neural system. However, the lack of evidence in neuroscience for such a theory and the fact that people often recognize emotions as ambiguous and overlapping experiences, much like the color spectrum, have prompted researchers to consider dimensional theories of emotion. One prominent theory is the circumplex model of affect, which suggests that all emotions arise from two independent neurophysiological systems and can be mapped to points on a circle in a two-dimensional space. Although different conceptualizations of these dimensions exist, the valence and arousal dimensions (see Figure 1) fit our research context particularly well (Fragopanagos and Taylor 2005; Kuppens et al. 2013; Lang et al. 1995; Posner et al. 2005; Russell 1980).

The valence dimension, visualized in Figure 1 as the horizontal axis, is a pleasure-displeasure dimension. The arousal dimension, visualized in Figure 1 as the vertical axis, is an arousal-sleep or energy dimension. An individual experiencing an emotional state high in the arousal dimension is highly activated and reactive to stimuli. Within this framework, each emotion can be characterized as a distinct point on a circle in this two-dimensional space. For example, the emotional state of feeling serene can be characterized as high valence and low arousal, while the emotional state of being furious can be described as low valence and high arousal. Drawing upon this theoretical framework, we analyze how perceived agent gender affects the emotional state of a customer in each dimension. To understand the role of gender in a customer's emotional valence, we first recognize that a complaining customer is typically driven by two distinct types of motivations: goal-oriented and emotion-focused (Kowalski 1996). While goal-oriented motivation prioritizes redress seeking, emotion-focused motivation is more about venting dissatisfaction and getting emotional support.

[3] For details, please see https://www.zippia.com/customer-service-representative-jobs/demographics/.

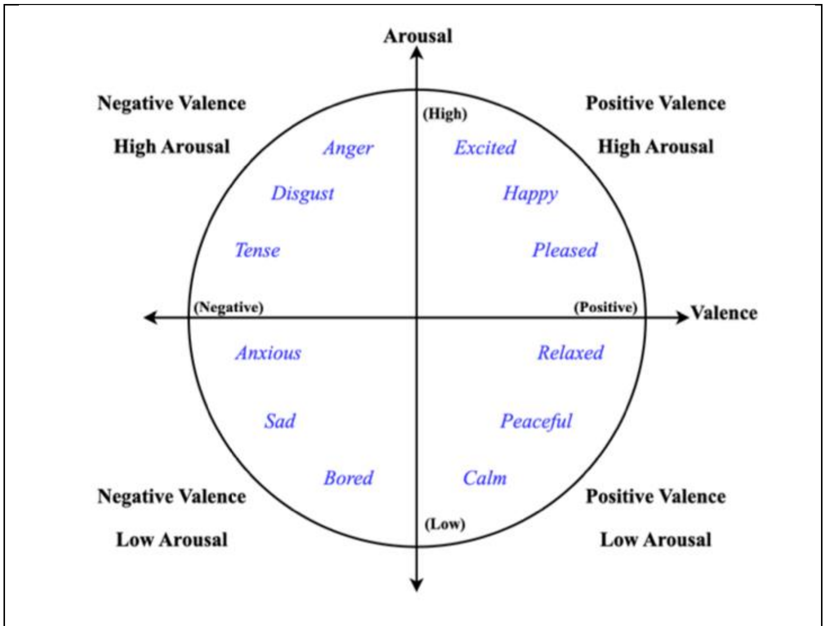

Figure 1. The Circumplex Model of Affects

For goal-oriented customers, the primary objective is to seek redress or economic compensation through complaining. The emotional valence embedded in their messages serves as an instrument for achieving this objective. These customers may strategically emphasize or exaggerate the negative valence when communicating with female agents because women are often perceived as more compliant and soft-hearted than men (Del Boca and Ashmore 1980; Eagly 1983). In other words, goal-oriented customers may subconsciously or even intentionally exploit the perceived soft-heartedness of female agents to increase the likelihood of achieving their goals. From an economic perspective, since it is emotionally costly for a customer to communicate negative sentiments in a conversation and female agents are perceived as more accommodating than their male counterparts, a goal-oriented customer would optimally exhibit a more negative emotional valence towards a female agent for the best return on investment.

Meanwhile, the desire to express emotional dissatisfaction is the primary driver of complaints from emotion-focused customers. These customers may consider female agents to be a better audience because women are often perceived as more empathetic than men (Christov-Moore et al. 2014; Eagly and Johnson 1990; Eagly and Wood 2012). In fact, a recent study[4] found that compared to men, women were more capable of understanding the thoughts and feelings of others, and this gender difference exists across cultures and is independent of family influences. Therefore, emotion-focused customers may expect female agents to better understand and even resonate with their emotions than male agents, making these customers more willing to share their distress with female agents and resulting in more negative valence in their messages.

In reality, customers are likely driven to complain by a mix of both factors. Regardless, the above arguments predict that a customer message addressed to a female agent should exhibit lower emotional valence than one sent to a male agent. Hence, we propose the following hypothesis for empirical testing.

Hypothesis 2: *Conditional on engagement, customers show lower emotional valence towards female agents than male agents.*

To hypothesize the effect of agent gender on a customer's emotional arousal, we propose four potential mechanisms. First, because customers complaint to seek redress and/or express dissatisfaction, the perceived ability of agents to understand their feelings is critical to their emotional reaction. Females are often perceived as more capable of understanding others' feelings than males; hence, the presence of a female agent, rather than a male agent, could make a customer feel less agitated and exhibit less emotional

[4] Please see https://www.cnn.com/2022/12/26/health/empathy-women-men/index.html.

arousal. Second, social desirability states that people conform to social norms during social interactions (Chung and Monroe 2003). Since, traditionally, masculinity is associated with power, and femininity is associated with vulnerability (Eagly and Crowley 1986), females typically receive greater chivalrous treatment than males (FeldmanHall et al. 2016). Chivalrous behaviors include courteous and protective acts towards women, and one of the chivalric vows taken by medieval knights was a vow "to respect the honor of women." Therefore, when customers interact with female agents, they are likely to be more respectful and refrained, resulting in a less intense conversation with lower emotional arousal. Third, because women have historically constituted a larger proportion of customer service agents than men (Steiger and Wardell 1995), the confirmation or disconfirmation of such a belief in the presence of agent gender information may alter the arousal level of a customer. For example, if a customer subconsciously expects a female agent but is served by a male agent, a sense of surprise may increase emotional arousal. Indeed, according to Russell (1980), two words that rank particularly high in the arousal dimension are "alarmed" and "astonished". In contrast, a customer served by a female agent may feel a sense of ease and familiarity (Vittengl and Holt 1998), thereby reducing emotional arousal. Finally, unlike the above mechanisms, some customers may exhibit more emotional arousal when interacting with female agents because women are perceived as more compliant and empathetic (Eagly 1983; Impett and Peplau 2003), especially when women are emotionally pressured.

Although it is unclear which of the above mechanism would dominate in our setting, we predict, simply based on the number of mechanisms, that the overall effect of female gender cues on emotional arousal would be negative. Hence, we propose the following hypothesis for empirical testing.

Hypothesis 3: *Conditional on engagement, customers exhibit lower emotional arousal towards female agents than male agents.*

Empirical Strategy

Data and Variables

On March 16, 2018, Southwest Airlines introduced a signature policy on Twitter that requires its customer service agents to include their first names at the end of their responses to customer inquiries. The presence of these names signals social presence (Gao et al. 2023). Most importantly for our study, a name provides an identity cue through which customers may consciously or subconsciously infer an agent's gender. Figure 2 illustrates two customer service tasks handled by agents with female- and male-dominant names. A Southwest Airlines customer service task typically involves the following steps. First, a customer initiates the conversation by mentioning the official Twitter account of Southwest Airlines (@SouthwestAir) in the tweet. A Southwest Airlines agent picks up this customer service request by responding to the customer with his/her name at the end of the message. To continue the conversation, the customer responds by tweeting within the same Twitter thread.

We collected all tweets received and posted by Southwest Airlines from March 2018 to September 2019. We leveraged Twitter metadata to reconstruct customer service conversations, starting with initial service requests by customers, followed by back-and-force interactions between agents and customers. In total, we obtained 126,429 conversations consisting of 332,693 back-and-forth tweets. We leveraged the support vector machine (SVM) classifier constructed by Gao et al. (2023) to identify customer service-related conversations in the data. To infer this information from an agent's signature, we hired two annotators to label the dominant gender of each name based on US Census statistics (Word et al. 2008). After excluding conversations with ambiguous or invalid names from the sample, there were 67,736 customer service-related conversations. Within this sample, 23,331 conversations continued past the agent's initial response to the initial service request. Consistent with popular perception, the majority (i.e., 64%) of these agents were female, at least based on gender inferred by names.

Table 1 lists the variables with their definitions and summary statistics. The explanatory variable, *Female*, is a binary variable equal to one if a customer service task is handled by a female agent. We measure a customer's willingness to engage through a binary variable, *SecondTweet*, which measures whether the customer follows up on the agent's initial response by sending a second tweet. This is a clean, reasonable measure of engagement because the entire conversation usually ends if a customer does not send a second tweet.

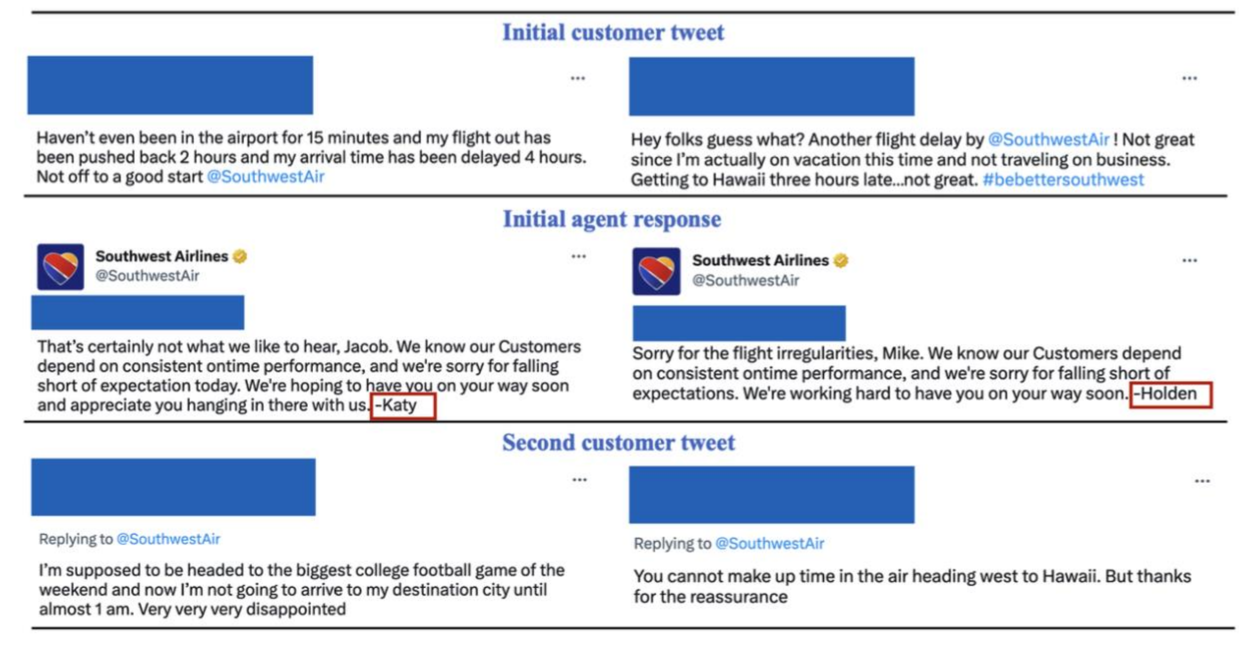

Figure 2. Sample Conversations

We then define two outcome variables, *SecondTweetValence* and *SecondTweetArousal*, as continuous measures that captures a customer's emotional state in the second tweet. We only use the second tweet instead of all tweets in the conversation for two reasons. First, the second tweet is the first observable customer reaction after an agent's gender could be inferred from his/her initial response. Therefore, the identification of the gender effect is less vulnerable to potential confounding factors introduced after the second tweet. Second, to measure valence and arousal, we use machine learning algorithms, whose accuracy depends on the text length. Since the number of customer tweets varies from conversation to conversation, using all tweets in a conversation could have led to measurement errors and introduced new confounding factors.

To further alleviate measurement error concerns, we use two algorithms to construct each outcome variable. For *SecondTweetValence*, we use classifiers based on two supervised machine learning algorithms to generate *SecondTweetValence_LR* and *SecondTweetValence_NB*. Specifically, we randomly selected 5000 customers' messages and hired two annotators to manually label the valence. Based on the labeled data, we built classifiers using four popular machine learning algorithms: logistic regression, Naïve Bayes, SVM, and bidirectional encoder representations from transformers (BERT). We chose the logistic regression and the Naïve Bayes classifiers to construct the measures because they outperformed the other two classifiers in terms of the F1 score. For *SecondTweetArousal*, we use supervised machine learning and lexical-based weighting approaches to construct *SecondTweetArousal_LSTM* and *SecondTweetArousal_Lexical*. Specifically, we built a long short-term memory (LSTM) classifier to predict arousal values based on data provided by Buechel and Hahn (2017) while using the lexicon and emotional weights provided by Warriner et al. 2013 to construct the lexical classifier. Due to the occurrence of terms within the emotional lexicon, the lexical classifier constructed arousal weights for 65,574 out of 67,736 tweets.

Variables	Mean	S. D.	Definition
Treatment			
Female	0.74	0.44	Binary variable indicating whether an agent is female
Outcome Measures			
SecondTweet	0.33	0.47	Binary variable indicating whether a customer engages with an agent's first reply

SecondTweetValence_LR	0.67	0.20	Continuous measure of customers' emotional valence in the second tweet based on the Logistic Regression classifier (within [0,1])
SecondTweetValence_NB	0.65	0.19	Continuous measure of customers' emotional valence in the second tweet based on the Naïve Bayes classifier (within [0,1])
SecondTweetArousal_LSTM	0.71	0.01	Continuous measure of customers' emotional arousal in the second tweet based on the LSTM classifier (within [0,1])
SecondTweetArousal_Lexical	0.66	0.10	Continuous measure of customers' emotional arousal in the second tweet based on lexical weights (within [0,1])
Customer Characteristics			
LogFollowers	5.07	1.97	Log-transformed number of followers a customer has
LogFollowings	5.62	1.47	Log-transformed number of followings of a customer
LogUpdates	7.30	2.44	Log-transformed number of tweets of a customer
InitialValence_LR	0.58	0.16	Continuous measure of customers' emotional valence in the initial tweet based on the Logistic Regression classifier (within [0,1])
InitialValence_NB	0.57	0.15	Continuous measure of customers' emotional valence in the initial tweet based on the Naïve Bayes classifier (within [0,1])
InitialArousal_LSTM	0.72	0.01	Continuous measure of customers' emotional valence in the initial tweet based on the LSTM classifier (within [0,1])
InitialArousal_Lexical	0.68	0.11	Continuous measure of customers' emotional valence in the initial tweet based on the lexical weights (within [0,1])
InitialAggression	0.10	0.12	Continuous measure of a customer's aggression in the initial tweet
InitialWords	26.40	13.04	Number of words in a customer's initial Tweet
InitialHello	0.11	0.31	Binary variable indicating whether a customer greets an agent
InitialGratitude	0.07	0.27	Binary variable indicating whether a customer is polite in their initial request using words such as "thank you", "I appreciate", etc.
InitialBareCommand	0.19	0.41	Binary variable indicating whether a customer expresses the service request in grammatically polite structure (i.e., starts a sentence with the unconjugated verbs)
InitialQuestions	0.16	0.40	Binary variable indicating the existence of question marks in the initiated tweet
Agent Reply Quality			
AgentWords	28.98	13.88	Number of words in an agent's first reply
ResponseTime	0.43	1.01	Number of hours it takes for an agent to respond
DM	0.29	0.45	Binary variable indicating whether an agent mentions direct messages
Please	0.28	0.45	Binary variable indicating whether an agent mentions "please"
Hello	0.33	0.47	Binary variable indicating whether an agent greets a customer
Apology	0.47	0.50	Binary variable indicating whether an agent apologizes to a customer
Hedges	0.36	0.48	Binary variable indicating whether an agent shows uncertainty

Note. This table reports the summary statistics of the key variables. S.D. stands for standard deviation. The number of observations is 67,736. InitialValence_LR is controlled for when SecondTweetValence_LR is the outcome variable, and InitialValence_NB is controlled for when SecondTweetValence_NB is the outcome variable. Similarly, InitialArousal_LSTM is included when SecondTweetArousal_LSTM is the outcome variable, InitialArousal_Lexical is included when SecondTweetArousal_Lexical is the outcome variable.

Table 1. Summary Statistics

Identification

To test our hypotheses, we estimate the following econometric model at the conversation level.

$$Y_i = \beta_0 + \beta_1 Female_i + \beta_2 X_i + \beta_3 Z_t + \epsilon_i$$

The outcome variable Y_i was either *SecondTweet*, *SecondTweetValence*, or *SecondTweetArousal* for conversation i, depending on which hypothesis was tested. The key coefficient of interest, β_1, captured the main treatment effect of an agent's gender on a customer's engagement or emotional state. X_i included the characteristics of conversation *i*, such as the profile of the customer, the content of the first customer tweet, and the content of the first agent response. Z_t included seasonality variables, such as the day of the week and the hour of the day.

An ideal research design would involve the random assignment of customer service requests to male or female agents as well as an identical initial response, other than agent names, from the agents. This would ensure that any statistically significant difference in an outcome variable derived from the second tweet between requests handled by male and female agents is caused by the name-induced gender cue and customers' resulting perceptions.

Although such a research design is not possible with observational data, it can be closely approximated thanks to two unique features of our research context. To facilitate the discussion, we illustrate the causal structure among key constructs using the directed acyclic graph (DAG) in Figure 3.

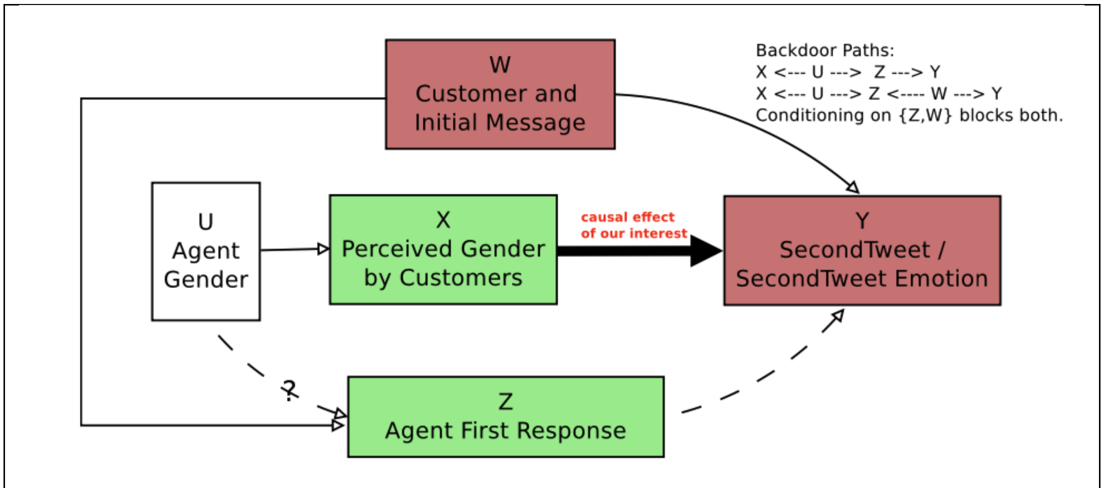

Figure 3. Causal graph and identification strategy

In this causal graph, the thick solid line connecting the explanatory variable X (i.e., *Perceived Gender by Customers*) and the outcome variables Y (i.e., *SecondTweet*, *SecondTweetValence*, and *SecondTweetArousal*) represents the main causal effect of interest (i.e., X → Y in the causal graph).

We first argue that there is no arc connecting node W and node U because, in large organizations such as Southwest Airlines, a customer complaint is usually assigned to the next available agent, which should be largely independent of the agent's gender, especially after controlling for seasonality (e.g., the day of the week, the hour of the day). This makes sense for two reasons. First, customers care about the speed of response and firms care about workforce efficiency, which makes it beneficial to assign a service request to the next available agent, regardless of the agent's gender. Indeed, the Next Available Agent (NAA) strategy is commonly used by call centers to reduce the average wait time for customers and to optimize agent productivity. Second, the practice of assigning service requests to agents based on their gender is at best controversial, and at worst legally problematic. Therefore, we can think of the assignment of customers to the two groups of agents (i.e., male and female) as independent of the group ID, much like in a randomized experiment. Similar independence assumptions have been made in the literature, for example, to identify the long-run consequences of living in a poor neighborhood (Oreopoulos 2003) where a public housing program in Toronto assigns each eligible family to the next available residence with the correct number of bedrooms. To empirically evaluate the plausibility of this assumption, we check the balance of customer characteristics and the content of the customer's initial tweet across the two groups. For customer characteristics, we consider the log-transformed numbers of a customer's followers and followings (*LogFollowers* and *LogFollowings*), and the log-transformed number of tweets a customer has posted (*LogUpdates*) at the time of a service request. For the characteristics of the initial tweet, we consider the number of words (*InitialWords*); the level of aggression derived from the Google Perspective API

(*InitialAggression*)[5]; a set of binary variables for the writing style, such as (*InitialSwearing*), which is one if the customer swears in the initial tweet; and the valence and arousal scores, which are similarly constructed as our outcome variables. We find that the standardized differences are all below the threshold of 0.10 suggested by Austin (2009) to assess covariate balance for samples generated from matching or from randomized experiments. Therefore, we believe that the assignment process is largely random and that any unobserved customer characteristic is likely balanced across the two groups.

Recall blocking means stopping the flow of dependency between nodes connected by paths. Formally, given three disjoint subsets of nodes X, Y, and Z, a path p is said to be **blocked** by Z if and only p contains a **chain** ($i \rightarrow z \rightarrow j$) or a **fork** ($i \leftarrow z \rightarrow j$) such that z ∈ Z, or if p contains a **collider** ($i \rightarrow m \leftarrow j$) such that m ∉ Z and no descendants of *m* is in Z.

In our causal graph, there are two potential backdoor paths, where a backdoor path between a causal node and an outcome node is defined as any path with an arrow into the causal node. The first backdoor path (i.e., *Perceived Gender by Customers ← Agent Gender → Agent First Response → Second Tweet / Second Tweet Emotion*), connected through the dashed line at the bottom of Figure 3, may create a spurious correlation between perceived agent gender and the outcome variables through the mechanism of different initial responses by agents of different genders. To block this backdoor path, we need to control for an agent's initial response. While it is straightforward to include the response time (i.e., the elapsed time between a customer's initial tweet and an agent's initial response) as a control, it is less clear how to control for the content of an agent's initial response due to the complexity of natural language. We believe this is achievable thanks to the second feature of our research context: the lack of significant, meaningful variation in the first response by customer service agents. Moreover, the short text length of the first response allows us to control for its content sufficiently well so that the potential bias caused by any residue is likely small. We use several approaches to control the content of an agent's first response. The first is the traditional approach of feature extraction in which we include features explicitly constructed to characterize agent response. For example, we control for the number of words in the reply (*AgentWords*) and the time it takes for an agent to respond to a customer's initial request (*ResponseTime*). In addition, we include a set of binary control variables to account for the agent's response style, such as whether the response mentions direct messages (*DM*) and includes the word "please" (*Please*) as well as whether the agent greets the customer (*Hello*), apologizes to the customer (*Apology*) or shows hedging in the response (*Hedges*). Our second approach involves creating an embedding vector for each agent response. Specifically, we employ latent semantic analysis (LSA) to create a 20-dimensional dense vector for each initial response in which each element of the vector is a continuous, rather than discrete, measure of the response in some latent semantic dimension. The idea of embedding has been widely used in natural language processing over the past two decades, and its value for causal inference has been recognized in recent years. In our third approach, we employ clustering to categorize content types, which have been recently exploited for identification. Specifically, we generate the document-term matrix for all agents' responses and perform singular value decomposition (SVD) to create word embedding. Next, we apply k-means clustering on word embedding to create agent reply clusters. We use the silhouette score to determine the optimal number of clusters, with 15 clusters generating the highest silhouette score. Finally, we include cluster dummies in the model as the control for content.

The second potential backdoor path (i.e., *Perceived Gender by Customers ← Agent Gender → Agent First Response ← Customer and Initial Complaint → Second Tweet / Second Tweet Emotion*) is blocked if we control for *Customer and Initial Complaint*, which is a fork. Note that *Agent First Response* is a collider here. Since we want to control for this node to block the first potential backdoor path, we have to block the second potential backdoor path by controlling for the fork node. Similar to our control for *Agent First Response*, we can use various strategies to control for characteristics of the customer and the initial complaint. Note that if agent gender only affects agent first response in a way that does not affect the outcome node (i.e., the part of the node *Agent First Response* affecting the outcome node is not affected by *Agent Gender*), then we could identify the causal effect simply by not controlling for *Agent First Response* because it is a collider in the second potential backdoor path. In fact, the results of this simplified identification strategy are precisely what the model-free evidence shows.

[5] For details on the Google Perspective API, please see https://perspectiveapi.com/.

In summary, if we condition on the nodes *Agent First Response* and *Customer and Initial Complaint*, both potential backdoor paths are blocked. Finally, recall the Backdoor Criterion (Pearl 2009) that for three disjoint sets of nodes, $\mathcal{D}$, $\mathcal{Y}$, and X, causal effect of $\mathcal{D}$ on $\mathcal{Y}$ can be identified if, for any $D \in \mathcal{D}$, $Y \in \mathcal{Y}$, the set of controls X contains no descendant of D and blocks every backdoor path between D and Y. Therefore, we can use observable data to identify our causal effects of interest.

Empirical Results

Main Results

Since we may improve statistical inference with a matched sample in the absence of a completely balanced sample, we also perform propensity score matching (PSM) with one-to-one matching using customer and agent characteristics before the second tweet as matching variables and a caliper of 0.001 following Austin (2011). Table 2 reports the estimation results with *SecondTweet* as the outcome variable. The first three columns correspond to the full sample, and the last one corresponds to the matched sample. The differences among the first three columns reflect the three strategies (i.e., feature-based control, embedding, and cluster dummies) used to control for the content of an agent's first response. Despite the differences in sampling and content control strategies, the estimated coefficient of the *Female* variable is consistently negative and statistically significant (p < 0.01). We replicated this analysis combining feature-based controls, agent response word embedding, and agent response cluster dummies on both the full sample and the matched sample. The results combining all three control strategies on both samples are consistent and are available upon request. Hence, as our model-free evidence has previously suggested, customers are more likely to continue a service conversation with female agents than with male agents, thereby supporting **Hypothesis 1**.

	SecondTweet			
	(1)	(2)	(3)	(4)
	Full Sample	Full Sample	Full Sample	Matched Sample
Female	0.0393*** (0.0040)	0.0364*** (0.0042)	0.0285*** (0.0040)	0.0410*** (0.0043)
Customer Controls	Yes	Yes	Yes	Yes
Agent Controls	Yes			Yes
Word Embedding		Yes		
Response Cluster FE			Yes	
Observations	67,736	67,736	67,736	63,993
R^2	0.0424	0.0500	0.0411	0.0425

Note. *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors are reported in parentheses. Agent response time and seasonality FE are included in all specifications.

Table 2. Agent Gender on Customer Engagement

Table 3 reports the estimation results with *SecondTweetValence* as the outcome variable, which is operationalized using two algorithms. The first four columns report the estimation results when the second tweet valence is measured by *SecondTweetValence_LR*, while the last four columns report the estimation results when the valence is measured by *SecondTweetValence_NB*. The last column of each group (i.e., Columns 4 and 8) corresponds to the matched sample, while the other three correspond to the full sample with different strategies for content control of an agent's first response. As evident in the table, regardless of the differences in measure construction, sampling strategy, and content control, the estimated coefficients of *Female* are consistently negative and statistically significant (p < 0.01), indicating that customers' emotional valence is lower when their service requests are handled by a female agent than by a male agent. We replicated this analysis combining feature-based controls, agent response word embedding, and agent response cluster dummies on both the full sample and the matched sample. The results combining all three control strategies on both samples are consistent and are available upon request. Therefore, **Hypothesis 2** is supported.

	SecondTweetValence_LR				*SecondTweetValence_NB*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Full Sample	Full Sample	Full Sample	Matched Sample	Full Sample	Full Sample	Full Sample	Matched Sample
Female	-0.0121*** (0.0029)	-0.0119*** (0.0029)	-0.0114*** (0.0029)	-0.0130*** (0.0029)	-0.0136*** (0.0028)	-0.0131*** (0.0028)	-0.0123*** (0.0028)	-0.0144*** (0.0028)
Customer Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Agent Controls	Yes			Yes	Yes			Yes
Word Embedding		Yes				Yes		
Response Cluster FE			Yes				Yes	
Observations	22331	22331	22331	21308	22331	22331	22331	21308
R^2	0.1445	0.1953	0.1455	0.1466	0.1451	0.1902	0.1450	0.1465

Note. *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors are reported in parentheses. Agent response time and seasonality FE are included in all specifications.

Table 3. Agent Gender on Customer Second Tweet Valence

Table 4 reports the estimation results with *SecondTweetArousal* as the outcome variable. Similar to Table 3, Table 4 contains two constructions of the outcome variables, two samples, and three strategies to control for the content of an agent's first response. The negative estimated coefficients ($p < 0.01$) of the *Female* variable across all specifications suggest that a customer's emotional arousal is reduced in the presence of a female agent, thereby supporting **Hypothesis 3**. We replicated this analysis combining feature-based controls, agent response word embedding, and agent response cluster dummies on both the full sample and the matched sample. The results combining all three control strategies on both samples are consistent and are available upon request.

	SecondTweetValence_LR				*SecondTweetValence_NB*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Full Sample	Full Sample	Full Sample	Matched Sample	Full Sample	Full Sample	Full Sample	Matched Sample
Female	-0.0006*** (0.0001)	-0.0006*** (0.0002)	-0.0005*** (0.0002)	-0.0005*** (0.0002)	-0.0024** (0.0010)	-0.0019* (0.0011)	-0.0020** (0.0010)	-0.028** (0.0011)
Customer Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Agent Controls	Yes			Yes	Yes			Yes
Word Embedding		Yes				Yes		
Response Cluster FE			Yes				Yes	
Observations	22331	22331	22331	21308	20280	20280	20280	19348
R^2	0.0099	0.0123	0.0082	0.0101	0.0108	0.0196	0.0130	0.0110

Note. *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors are reported in parentheses. Agent response time and seasonality FE are included in all specifications.

Table 4. Agent Gender on Customer Second Tweet Arousal

Additional Analyses

The Role of Agent Race

Besides gender identity, racial identity is another important piece of information that customers may infer from agent signatures.[6] Racial identity has been shown to affect agents' treatment of customers (Gunarathne et al. 2022). Likewise, customers may have race-based preferences regarding agents. Even though gender-role is profound due to the female-dominant customer service industry, if race indeed affects customers' preferences, our identification of the gender effect would be compromised without controlling for agent race. To alleviate this concern, we infer the race information from agent signatures. Specifically, we construct a binary variable, *White*, by matching each signature with US demographics and assigning it to the most likely race (Tzioumis 2018). In our sample, 75.5% of agents are inferred as white, and 24.5% are non-white. Using this information, we control for both agent gender and race in our analysis. As shown in Table 5, for all three outcome variables, we find insignificant coefficient estimates of *White*, suggesting that agent race plays a minimal role in customer behavior. In the meantime, the coefficient estimates of *Female* remain statistically significant, suggesting the robustness of our findings on the gender effect. In other words, customer preferences in customer service are driven by agent gender rather than agent race. We replicated this analysis with the other three specifications (i.e., embedding, cluster dummies, PSM with feature-based control). The results are consistent and are available upon request.

	(1)	(2)	(3)	(4)	(5)
	SecondTweet	*SecondTweetValence*		*SecondTweetArousal*	
		LR	NB	LSTM	Lexical
Female	0.0311***	-0.0135***	-0.0151***	-0.0083**	-0.0021**
	(0.0042)	(0.0029)	(0.0028)	(0.0034)	(0.0010)
White	-0.0048	0.0036	0.0038	0.0048	-0.0000
	(0.0039)	(0.0026)	(0.0026)	(0.0031)	(0.0009)
Customer Controls	Yes	Yes	Yes	Yes	Yes
Agent Controls	Yes	Yes	Yes	Yes	Yes
Observations	62946	21874	21874	21874	19874
R^2	0.0075	0.1013	0.0969	0.0095	0.0086

Note. *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors are reported in parentheses. Agent response time and seasonality FE are included in all specifications. Due to not all names being recognized, the number of observations is 62,946 for the full sample and 21,874 conditional on engagement. In the case of SecondTweetArousal_Lexical as the outcome variable, the number of observations is 19,874 conditional on engagement.

Table 5. The Role of Agent Race

Quadrant Analyses

The central tenet of the circumplex model of affect is the circular ordering of eight human emotions organized into four bipolar pairs, with the horizontal and vertical dimensions representing the misery-pleasure pair and the arousal-sleep pair, respectively. The other two bipolar pairs do not form independent dimensions: the distress-contentment pair occupies the northwest-southeast axis, and the excitement-depression pair occupies the northeast-southwest axis. This structure suggests alternative ways of investigating the effect of agent gender on customer emotion by focusing on different quadrants. Therefore, we conduct two sets of analyses given the following labeling of quadrants: quadrant 1 (negative valence, high arousal), quadrant 2 (negative valence, low arousal), quadrant 3 (positive valence, low arousal), and quadrant 4 (positive valence, high arousal).

In the first quadrant analysis, we focus on quadrants 1 and 2, which are particularly suitable for our study for two reasons. First, these two quadrants correspond to negative valence, which is of particular interest

[6] Other identity information, such as age, is rather difficult to be inferred from a signature consisting of a first name.

in the context of customer service. Second, the linguistic inquiry and word count (LIWC) lexicon (Tausczik and Pennebaker 2010) has a word class for anger, which belong to quadrant 1, and word classes for anxiety and sadness, which belong to quadrant 2. We use these word classes to construct two additional outcome variables—*quadrant1* and *quadrant2*—to measure the percentage of anger words and anxiety and sadness words, respectively, in a given tweet. Table 6 reports the estimation results of our econometric model with *quadrant1* and *quadrant2* as the outcome variables. We find a negative coefficient ($p < 0.01$) of *Female* when the outcome variable is *quadrant1* and a positive coefficient ($p < 0.01$) of *Female* when the outcome variable is *quadrant2*. Therefore, consistent with our main results, we find reduced emotional arousal in the presence of female agents. In particular, customers tend to use fewer high-arousal words (e.g., LIWC words related to anger) and more low-arousal words (e.g., LIWC words related to anxiety and sadness) when communicating with female agents when compared to male agents.

	quadrant1				*quadrant2*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Full Sample	Full Sample	Full Sample	Matched Sample	Full Sample	Full Sample	Full Sample	Matched Sample
Female	-0.0934^{***} (0.0185)	-0.1892^{***} (0.0171)	-0.0743^{***} (0.0181)	-0.0972^{***} (0.0191)	0.1656^{***} (0.0324)	0.01517^{***} (0.0334)	0.0727^{***} (0.0376)	0.1608^{***} (0.0333)
Customer Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Agent Controls	Yes			Yes	Yes			Yes
Word Embedding		Yes				Yes		
Response Cluster FE			Yes				Yes	
Observations	22331	22331	22331	21308	22331	22331	22331	21308
R^2	0.0233	0.2384	0.0752	0.0241	0.3064	0.3221	0.0813	0.3099
Note. *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors are reported in parentheses. Agent response time and seasonality FE are included in all specifications.								

Table 6. Quadrant Analysis I

In the second quadrant analysis, we construct a categorical outcome variable, *Quadrant*, based on the aforementioned labeling of the four quadrants. By definition, the values of the variable *Quadrant* are ordered by how favorable a quadrant is to an agent regarding customers' sentiments. *Quadrant = 1* (i.e., negative valence, high arousal) is the least favorable to an agent, while *Quadrant = 4* (i.e., positive valence, high arousal) is the most favorable. Accordingly, we implement an ordered logit model to estimate the effect of agent gender on the chances of falling into a more favorable quadrant. Table 7 reports the estimation results. Across different specifications, we find significantly negative coefficients of *Female*, suggesting that female agents are less likely to be in a favorable quadrant when compared to male agents. In other words, customers tend to be more emotionally draining on female agents.

	Quadrant			
	(1)	(2)	(3)	(4)
	Full Sample	Full Sample	Full Sample	Matched Sample
Female	-0.1274^{***} (0.0395)	-0.1229^{***} (0.0421)	-0.1455^{***} (0.0394)	-1.8511^{***} (0.1161)
Customer Controls	Yes	Yes	Yes	Yes
Agent Controls	Yes			Yes
Word Embedding		Yes		
Response Cluster FE			Yes	

Observations	22331	22331	22331	21308
R^2	-11014.950	-10635.352	-11199.817	-10309.991
Note. *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors are reported in parentheses. Agent response time and seasonality FE are included in all specifications.				

Table 7. Quadrant Analysis II

Conclusions

This paper studies the effect of perceived agent gender on customer behavior in the context of social media customer service. Using a unique data set with rich conversation details, we found that customers are more willing to engage with female agents than male agents in customer service conversations. Based on the theoretical decomposition of emotion into valence and arousal, we found empirical evidence that, upon engagement, customers show lower emotional valence and arousal towards female agents than male agents.

The main contributions of our paper are twofold. First, and most importantly, this is the first academic study to empirically examine how the perceived gender of customer service agents affects customer behavior in an online environment. Second, this study is among the first in the IS field to evaluate customer emotion along the valence and arousal dimensions in online customer service, especially by leveraging machine learning techniques. Even though the circumplex model of affect (Posner et al. 2005; Russell 1980) suggests that arousal is as critical as valence, the former is much less studied by researchers, partly due to the lack of methodologies to accurately measure this dimension. The significant effects of perceived agent gender on these indispensable dimensions in our paper echo the theoretical framework and corroborate the importance of incorporating arousal when analyzing customer emotion.

Our paper also provides valuable insights to practitioners, especially managers who strive to improve the equity and equality of the service industry workforce. Since customer preferences in online customer service are often implicit and not readily detectable, issues are often left unnoticed by firms and researchers. Our findings demonstrate that customers' preferences concerning agent gender indeed exist and the resulting behavioral differences are real. As a result, it is imperative to take action to alleviate the potential negative consequences of gender inequality. Compared to male agents, female agents are in a much less favorable position (see Table 7). In addition to customers' emotional quantiles, we also conducted an analysis on customers' satisfaction at the end of the conversation. We found consistent results that customers are less likely to express appreciation (or feel satisfied) towards female agents compared to male agents, conditional on customer characteristics and agents' service quality. These results are available upon request. Therefore, managers should consider customers' gender preferences and the toll on female agents when evaluating service performance. For instance, besides the common assessment metrics, such as the number of requests handled, average response time, and customer sentiment, firms should also account for the emotional overload imposed on female agents by customers. Moreover, providing emotional support and training to female agents could help retain the best talent and, more importantly, create a more equitable and inclusive working environment.

References

Austin, P. C. (2009). "Using the standardized difference to compare the prevalence of a binary variable between two groups in observational research," *Communications in statistics-simulation and computation* (38:6), pp. 1228–1234.

Austin, P. C. (2011). "Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies," *Pharmaceutical statistics* (10:2), pp. 150– 161.

Basow, S. A. (1992). *Gender: Stereotypes and roles,* Thomson Brooks/Cole Publishing Co, Belmont, CA.

Bem, S. (1974). "The measurement of psychological androgyny," *Journal of Consulting and Clinical Psychology* (42:2), pp. 155–162.

Buechel, S. and Hahn, U. (2017). "EmoBank: Studying the Impact of Annotation Perspective and Representation Format on Dimensional Emotion Analysis," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers,* Valencia, Spain: Association for Computational Linguistics, 2017, pp. 578–585.

Cheng, H.T. and Pan, Y. (2021). "I'm Not a Chatbot: An Empirical Investigation of Humanized Profiles of Social Media Customer Service Representatives," in *Proceedings of the 54th Hawaii International Conference on System Sciences,* pp. 4167–4176.

Christov-Moore, L., Simpson, E. A., Coudé, G., Grigaityte, K., Iacoboni, M., and Ferrari, P. F. (2014). "Empathy: Gender effects in brain and behavior," *Neuroscience & biobehavioral reviews* (46:4), pp. 604– 627.

Chung, J. and Monroe, G. S. (2003). "Exploring social desirability bias," *Journal of Business Ethics* (44:4), pp. 291–302.

Del Boca, F. K. and Ashmore, R. D. (1980). "Sex stereotypes and implicit personality theory. II. A trait-inference approach to the assessment of sex stereotypes," *Sex Roles* (6:4), pp. 519–535.

Eagly, A. H. (1983). "Gender and social influence: A social psychological analysis." *American Psychologist* (38:9), pp. 971–981.

Eagly, A. H. and Crowley, M. (1986). "Gender and helping behavior: A meta-analytic review of the social psychological literature." *Psychological bulletin* (100:3), pp. 283–308.

Eagly, A. H. and Johnson, B. T. (1990). "Gender and leadership style: A meta-analysis." *Psychological bulletin* (108:2), pp. 233–256.

Eagly, A. H. and Wood, W. (2012). "Social role theory. In P. van Lange, A. Kruglanski, & E. T. Higgins (Eds.)" in *Handbook of theories in social psychology,* vol. 2. Sage Thousand Oaks, CA, pp. 458–476.

FeldmanHall, O., Dalgleish, T., Evans, D., Navrady, L., Tedeschi, E., and Mobbs, D. (2016). "Moral chivalry: Gender and harm sensitivity predict costly altruism," *Social psychological and personality science* (7:6), pp. 542–551.

Fischer, E., Gainer, B., and Bristor, J. (1997). "The sex of the service provider: does it influence perceptions of service quality?," *Journal of Retailing* (73:3), pp. 361–382.

Fiske, S. and Linville, P. (1980). "What does the schema concept buy us?," *Personality and Social Psychology Bulletin* (6), pp. 543–557.

Foster, C. and Resnick, S. (2013). "Service worker appearance and the retail service encounter: the influence of gender and age," *The Service Industries Journal* (33:2), pp. 236–247.

Fragopanagos, N. and Taylor, J. G. (2005). "Emotion recognition in human–computer interaction," *Neural Networks* (18:4), pp. 389–405.

Gao, Y., Rui, H., and Sun, S. (2023). "The Power of Identity Cues in Text-Based Customer Service: Evidence from Twitter," *MIS Quarterly* (47:3), pp. 983-1014.

Gunarathne, P., Rui, H., and Seidmann, A. (2022). "Racial Bias in Customer Service: Evidence from Twitter," *Information Systems Research* (33:1), pp. 43–54.

Hekman, D. R., Aquino, K., Owens, B. P., Mitchell, T. R., Schilpzand, P., and Leavitt, K. (2010). "An examination of whether and how racial and gender biases influence customer satisfaction," *Academy of Management Journal* (53:2), pp. 238–264.

Hoffman, M. (1977). "Sex differences in empathy and related behaviors," *Psychological Bulletin* (84:4), pp. 712–722.

Impett, E. A. and Peplau, L. A. (2003). "Sexual compliance: Gender, motivational, and relationship perspectives," *Journal of sex research* (40:1), pp. 87–100.

Kim, J.H., Youn, H., and Phau, I. (2021). "A triple whammy effect of employees' gender, job type, and service outcomes on consumer behavior," *Tourism Management Perspectives* (40), p. 100878.

Kowalski, R. M. (1996). "Complaints and complaining: Functions, antecedents, and consequences." *Psychological bulletin* (119:2), pp. 179–196.

Kuppens, P., Tuerlinckx, F., Russell, J. A., and Barrett, L. F. (2013). "The relation between valence and arousal in subjective experience." *Psychological bulletin* (139:4), pp. 917–940.

Lang, A., Dhillon, K., and Dong, Q. (1995). "The effects of emotional arousal and valence on television viewers' cognitive capacity and memory," *Journal of Broadcasting & Electronic Media* (39:3), pp. 313–327.

Luoh, H.F. and Tsaur, S.H. (2007). "Gender stereotypes and service quality in customer–waitperson encounters," *Total Quality Management* (18:9), pp. 1035–1054.

Mohr, L. A. and Henson, S. W. (1996). "Impact of employee gender and job congruency on customer satisfaction," *Journal of Consumer Psychology* (5:2), pp. 161–187.

Moshavi, D. (2004). "He Said, She Said: Gender Bias and Customer Satisfaction With Phone-Based Service Encounters," *Journal of Applied Social Psychology* (34:1), pp. 162–176.

Olivetti, C. and Petrongolo, B. (2016). "The evolution of gender gaps in industrialized countries," *Annual review of Economics* (8:1), pp. 405–434.

Oreopoulos, P. (2003). "The long-run consequences of living in a poor neighborhood," *The Quarterly Journal of Economics* (118:4), pp. 1533–1575.

Otterbring, T., Bhatnagar, R., Samuelsson, P., and Borau, S. (2021). "Positive gender congruency effects on shopper responses: Field evidence from a gender egalitarian culture," *Journal of Retailing and Consumer Services* (63), p. 102738.

Pearl, J. (2009). *Causality: Models, Reasoning and Inference, 2nd ed.* Cambridge University Press. New York, NY, USA.

Posner, J., Russell, J. A., and Peterson, B. S. (2005). "The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology," *Development and psychopathology* (17:3), pp. 715–734.

Proserpio, D., Troncoso, I., and Valsesia, F. (2021). "Does gender matter? The effect of management responses on reviewing behavior," *Marketing Science* (40:6), pp. 1199–1213.

Rendall, M. (2018). "Female market work, tax regimes, and the rise of the service sector," *Review of Economic Dynamics* (28), pp. 269–289.

Russell, J. A. (1980). "A circumplex model of affect." *Journal of personality and social psychology* (39:6), pp. 1161–1178.

Snipes, R. L., Thomson, N. F., and Oswald, S. L. (2006). "Gender bias in customer evaluations of service quality: an empirical investigation," *Journal of Services Marketing* (20:4), pp. 274–284.

Steiger, T. L. and Wardell, M. (1995). "Gender and employment in the service sector," *Social problems* (42:1), pp. 91–123.

Tausczik, Y. R. and Pennebaker, J. W. (2010). "The psychological meaning of words: LIWC and computerized text analysis methods," *Journal of language and social psychology* (29:1), pp. 24–54.

Tzioumis, K. (2018). "Demographic aspects of first names," *Scientific data* (5:1), pp. 1–9.

Vittengl, J. R. and Holt, C. S. (1998). "Positive and negative affect in social interactions as a function of partner familiarity, quality of communication, and social anxiety," *Journal of Social and Clinical Psychology* (17:2), pp. 196–208.

Warriner, A. B., Kuperman, V., and Brysbaert, M. (2013). "Norms of valence, arousal, and dominance for 13,915 English lemmas," *Behavior research methods* (45), pp. 1191–1207.

Word, D. L., Coleman, C. D., Nunziata, R., and Kominski, R. (2008). *Demographic aspects of surnames from census 2000*. United States Census Bureau. Retrieved April 28, 2023, from https://www2.census.gov/ topics/genealogy/2000surnames/surnames.pdf.