The Power of Identity Cues in Text-Based Customer Service: Evidence from Twitter

Yang Gao
School of Computing and Information Systems, Singapore Management University, Singapore.
{ygao@smu.edu.sg}

Huaxia Rui
Simon Business School, University of Rochester, Rochester, NY U.S.A.
{huaxia.rui@simon.rochester.edu}

Shujing Sun
Naveen Jindal School of Management, University of Texas at Dallas, Dallas, TX U.S.A.
{shujing.sun@utdallas.edu}

Abstract. Text-based customer service is emerging as an important channel through which companies can assist customers. However, the use of few identity cues may cause customers to feel limited social presence and even suspect the human identity of agents, especially in the current age of advanced algorithms. Does such a lack of social presence affect service interactions? We studied this timely question by evaluating the impact of customers' perceived social presence on service outcomes and customers' attitudes toward agents. Our identification strategy hinged on Southwest Airlines' sudden requirement to include a first name in response to service requests on Twitter, which enhanced customers' perceived level of social presence. This change led customers to become more willing to engage and more likely to reach a resolution upon engagement. We further conducted a randomized experiment to understand the underlying mechanisms. We found that the effects were mainly driven by customers who were ex ante uncertain or suspicious about the human identity of agents, and the presence of identity cues improved service outcomes by enhancing customers' perceived levels of trust and empathy. Additionally, we found no evidence of elevated verbal aggression from customers toward agents with identity cues, although a mechanism test revealed the moderating role of customers' emotional states. Our study highlights the importance of social presence in text-based customer service and suggests a readily available and almost costless strategy for firms: signal humanization through identity cues.

Keywords: social media, customer service, social presence, humanization, identity cue

1. Introduction

Text-based customer service channels, such as in-app messaging, live website chats, short message services, and social media, are becoming prevalent. Customers across all age demographics are moving toward text-based channels for customer support due to their convenience and asynchronous features. For instance, in a survey conducted by UJET Inc. (n.d.), 72% of customers across the United States (US) between the ages of 18 and 64 years old stated that texting with an agent improved their support experience. Despite various advantages, such as convenience for customers and low cost to brands, textual communications generally lack social context cues, which reduce the level of social presence felt by customers (Oh et al., 2018) and lead customers to the perception that they are communicating with dehumanized subjects (Mesch & Beker, 2010).

Indeed, customers nowadays wonder about the human identity of agents in text-based service provisions, especially when agents' responses seem monotone or out of context. The epic failure of Bank of America (BoA)'s customer service on Twitter provides a classic example. On July 6, 2013, Mark Hamilton wrote an anti-foreclosure statement in front of a BoA branch and was told to leave by the police. He tweeted the following message from his Twitter handle @darthmarkh: *"Just got chased away by #NYPD 4 [for] 'obstructing sidewalk' while #chalkupy-ing with @CyMadD0x outside @bankofamerica HQ"*. Another user, @stevetimmis, responded, *"@darthmarkh @CyMadD0x @bankofamerica looks like you were really causing an obstruction."* These tweets, along with other users' follow-up tweets, triggered responses from BoA's customer service team on Twitter (i.e., @BofA_Help), three of which are shown in the top left panel of Figure 1. The top right panel of Figure 1 displays a tweet from @BofA_Help, which was posted at 4:09 PM on July 6, 2013, in response to a tweet posted by @OccupyLA at 3:59 PM. Whatever strategy or tool BoA was using in its Twitter service provision, the response's monotone nature clearly obfuscated customers about the agent's identity. Moreover, customers wondered whether *"the Bank of America Twitter account was run exclusively through autobots"*, as evidenced by many tweets throughout the flare-up (see the bottom panel of Figure 1; Coine & Babbitt, 2014, p.199).

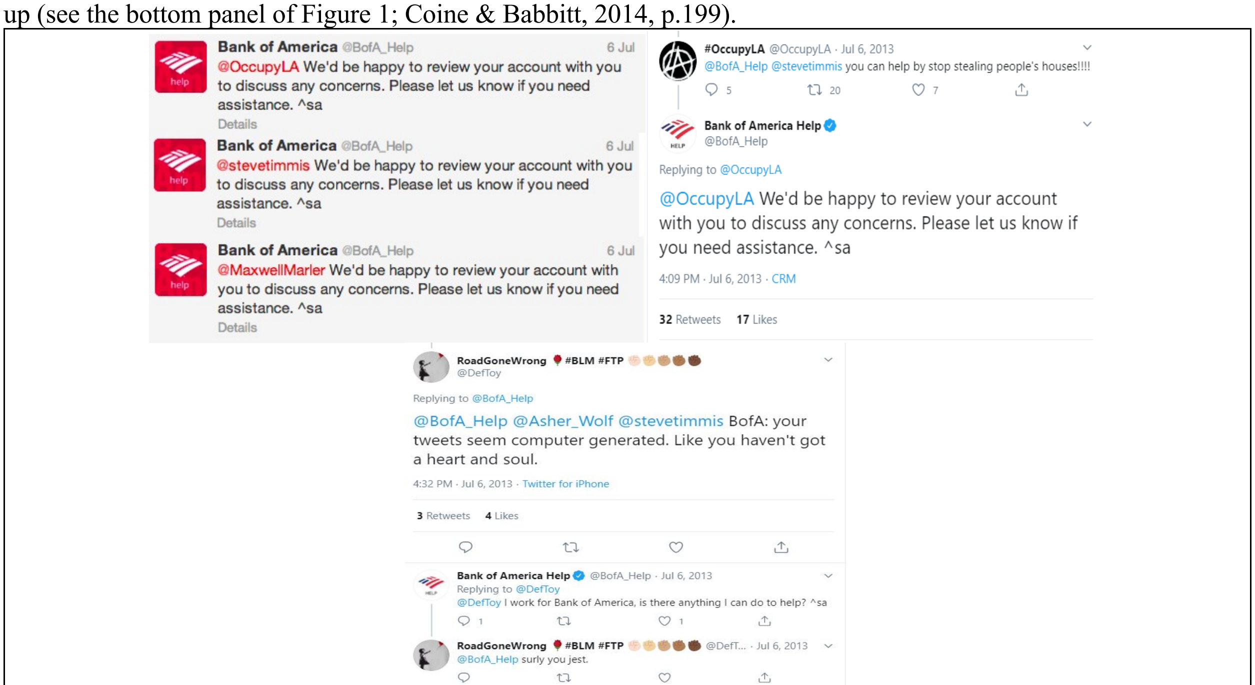


Figure 1. The Epic Failure of BoA's Twitter Customer Service in 2013

Given the increasing number of customers turning to text-based customer service for assistance and their confusion about agents' identities[1], it is important to understand *whether* and *how* customers' perceptions of social presence affect text-based service provision. Note that the conceptualization of social presence in the literature varies with the research context. One conceptualization that is particularly relevant to our research context is "*the subjective experience of being present with a 'real' person and having access to his or her thoughts and emotions*"

[1] To validate customers' confusion about agent identity, we surveyed 328 customers who had customer service experience on social media with the following question: "*Do you think social media customer service agents are human agents or algorithm-enabled agents?*" We found that 114 respondents (35%) were either uncertain about the human identity of agents or confident that agents were algorithm-enabled. This ratio is comparable with that of the randomized experiment described in Section Mechanism Tests (i.e., 38% of 400 respondents).

(Biocca, 1997; Kreijns et al., 2004; Oh et al., 2018). This definition emphasizes the psychological aspect of social presence, which aligns closely with customer perceptions in our context. To distinguish this from conceptualizations of social presence in different contexts (Cobb, 2009; Gunawardena, 1995; Kehrwald, 2008), we referred to this definition as ***humanization*** and interpreted it in our research context as the continuum of a customer's *perception* of being present with a "*real"* human agent during a customer service interaction.

To study how customers' perceived humanization levels affect their interactions with agents, we exploited a quasi-experiment that was induced by an identity cue policy from Southwest Airlines. On March 16, 2018, Southwest Airlines' customer service agents began including first names in service responses on Twitter. This sudden change, which we referred to as the ***signature experiment***, exemplified recent firms' efforts to "*let customers know they are talking to a person and not a bot*" (Perez, 2017). From the perspective of customers, Southwest Airlines' signature experiment was abrupt because there was no advanced notice or hint from the airline company; hence, customers cannot anticipate the change. Because an identity cue is a key contextual factor that influences social presence (Oh et al., 2018) and the inclusion of a first name (hereafter referred to as the ***signature***) provides an identity cue for the authorship of a response, we expected that customers' perceived humanization level would increase after the launch of the signature experiment.

Based on social presence theory, we proposed two sets of hypotheses to evaluate the impact of humanization on text-based customer service interactions. First, we considered customers' willingness to engage, which is a binary measure of whether customers would decide to continue a conversation upon receiving an agent's initial response to a service request. This engagement decision is crucial because a customer's follow-up is the prerequisite for an agent to continue the service. Conditional on a customer's further engagement, we examined whether the conversation led to a resolution. Second, because the inclusion of identity cues may cause unintended consequences, such as social media trolls due to the publicity and tractability of social media posts, we examined whether there was any change in customer verbal aggression, a well-recognized type of customer misbehavior that is detrimental to agents' cognitive and task performance (Grandey et al., 2004; Rafaeli et al., 2012).

We started the empirical analyses with a one-group before-and-after design (Claussen et al., 2013) to detect any changes in customer service interactions following the signature experiment. We found that an enhanced perceived humanization level increased customers' willingness to engage with service agents as well as the likelihood of reaching a resolution. Meanwhile, there was no evidence that customers behaved more or less aggressively after the signature policy. These findings remained consistent in various robustness checks and falsification tests. To minimize the chance that our findings were driven by unobserved time-varying factors, we then conducted difference-in-differences analyses by constructing a synthetic control group that was comparable to Southwest Airlines based on a donor pool consisting of four other major US airlines (Abadie et al., 2010; Abadie & Gardeazabal, 2003; Athey et al., 2021). To better control tweet-level confounders, we proposed a conversation-level two-way matching method, which was in the spirit of the synthetic control method but implemented at a more granular level. Both sets of results consistently supported our main findings.

To understand the underlying mechanism of the detected effects, we complemented the observational study with a randomized experiment on Amazon Mechanical Turk (AMT). By collecting data on and conditioning individuals' prior beliefs about agent identity, we discovered that the observed positive effect of identity cues on service outcomes came from customers who are uncertain or suspicious of an agent's human identity. Hence, adding identity cues updated customers' beliefs and lifted their perceived humanization level, which led to improved service outcomes. We also conducted a causal mediation analysis, which revealed that *trust* and *empathy* mediated the effect of identity cues on service outcomes. Specifically, identity cues enhanced customers' perceived levels of trust and empathy toward service agents, thereby increasing their willingness to engage and the likelihood of reaching a resolution.

To examine the mechanism behind the null effect of identity cues on customer verbal aggression, we conducted a mechanism test using observational data from Twitter and discovered the mediating role of customers' emotional states at the beginning of a conversation. Specifically, the inclusion of identity cues decreased the aggressiveness of customers who were initially emotion-focused but increased the aggressiveness of customers who were initially goal-oriented.

As the first study to connect social presence and service interactions in the context of social media customer service, this work contributes to the literature on online complaint management. By empirically verifying the critical role of individual heterogeneous prior beliefs about agents' identities and the theoretical mechanisms of trust and empathy, this study also contributes to social presence theory. For business practitioners, the positive impact of humanization on service interactions suggests that the human touch in customer service delivery is still much valued by customers. Our study thus sheds light on the looming question of how much automation firms should incorporate into customer service operations. Since call centers have long been perceived as cost centers, firms have been leveraging information technologies for decades to deliver customer service as cost-effectively as possible. The recent development of artificial intelligence (AI)-based chatbots presents the latest opportunity and probably the ultimate solution to such a quest for cost reduction.[2] Nonetheless, considering customers' inherent preference for engaging with human agents, a cost-oriented business strategy without a human touch in customer service provisions is not prudent.

The rest of the paper is organized as follows. We first review the relevant literature and then detail the development of hypotheses regarding the effect of humanization on customer service interactions. After describing the data and the key variables, we discuss identification strategies, report empirical analyses, and analyze the mechanisms. We conclude the paper by discussing its implications and limitations.

2. Literature Review

Our paper is related to the literature on online complaint management in terms of the research context and the literature on social presence in terms of the theoretical foundation. We review these two streams of literature separately.

2.1. Online Complaint Management

The literature on online complaint management can be categorized based on the nature of the underlying technology platform. The management response literature focuses on brand responses to consumer reviews posted on online review platforms, such as Yelp. The social media customer service literature focuses on the delivery of customer service through social media channels, such as Twitter. While users who post reviews are primarily motivated to share feedback with others and do not typically expect a response from the focal business, customers who complain to a brand via social media aim to address service concerns and often expect prompt responses from the brand, much like traditional customer service.

The management response literature mainly focuses on externality, such as the impact of management responses on the volume and valence of future reviews of a brand. While management responses have been shown to increase the volume of subsequent customer reviews (Chen et al., 2019; Proserpio & Zervas, 2017), their impact on review valence is inconclusive, as evidenced by positive and negative externalities on subsequent review sentiment (Chevalier et al., 2018; Wang & Chaudhry, 2018). Several studies have examined the impact of management responses on other dimensions, such as customer satisfaction (Gu & Ye, 2014; Zhao et al., 2020) and hotel performance (Kumar et al., 2018; Lee et al., 2016; Xie et al., 2014). Recent studies have investigated opinion management strategies for responding to online customer reviews (Proserpio et al., 2021; Yang et al., 2019). The current study offers new insights into this literature by introducing the element of social presence and examining how brand responses affect focal customers and service performance as opposed to externality.

The social media customer service literature has drawn increasing attention from information systems (IS) and marketing researchers in recent years and is closely related to the current study. One stream of this literature focuses on the customer side (Gans et al., 2021; Gunarathne et al., 2017; He et al., 2019; Ma et al., 2015). For example, Ma et al. (2015) found that redress seeking is a major driver of customer complaints. While service intervention can improve customers' relationships with a firm, it also induces more complaints. Another stream focuses on the firm side (Gunarathne et al., 2018, 2022; Hu et al., 2019; Mousavi et al., 2020; Sun et al., 2021). For example, Gunarathne et al. (2018, 2022) found that airlines selectively respond to customers based on their social media influence or racial identity. Our paper introduces an important and novel perspective to this literature by studying the implications of social presence in customer service delivery on social media.

2.2. Social Presence

Social presence was first introduced by Short et al. (1976) as a theoretical framework for understanding interactions that took place in different forms of media, such as face-to-face and computer-mediated interactions. Social presence was originally defined as "*the degree of salience of the other person in the interaction and the consequent salience of the interpersonal relationships*" (Short et al., 1976, p.65). Since its inception, different conceptualizations of social presence have been proposed for different research contexts. For example, studies of computer-mediated social networks have focused on the social aspect and referred to social presence as individuals' awareness of their social connections in a communication interaction (Cobb, 2009; Gunawardena, 1995; Kehrwald, 2008). Given the rapid advancements in automation technologies, several studies defined social presence as "*the subjective experience of being present with a 'real' person and having access to his or her thoughts and emotions*" (Biocca, 1997; Kreijns et al., 2004; Oh et al., 2018) to emphasize the psychological distance between individuals and their counterparts. We followed this relatively recent conceptualization of social presence and referred to it as

[2] AI is predicted to power 95% of all customer interactions by 2025, including live telephone and online conversations. For details, see https://www.financedigest.com/ai-will-power-95-of-customer-interactions-by-2025.html.

humanization to avoid confusion with other conceptualizations of social presence. More precisely, in our research context, we defined humanization as the continuum of a customer's perception of being present with a "*real"* human agent during a customer service interaction.

A major theme of the literature on social presence is identifying factors that can influence individuals' perceptions of social presence. We reviewed this literature based on whether the counterpart's identity (i.e., human or software) was known, a particularly important factor in the present study.

When the counterpart's identity is known to be software (e.g., a chatbot), the relevant literature is sometimes referred to as digital anthropomorphism, which aims to enhance individuals' perceived social presence. Anthropomorphism is defined as "*the tendency to attribute human characteristics to inanimate objects, animals, and others with a view to helping us rationalize their actions*" (Duffy, 2003). In the context of customer service, the literature has shown how specific features of service robots, such as physical features (e.g., facial configuration, voicing, and gestures) and skills, can shape customers' perceived anthropomorphism (Al-Natour et al., 2011; Salem et al., 2013; Yam et al., 2021; Zhang et al., 2010). Generally, customers tend to perceive more skilled conversational chatbots as more socially present and anthropomorphic than less skilled ones (Schuetzler et al., 2020). Digital anthropomorphism has been shown to affect customer behaviors, such as their intention to adopt service robots (Blut et al., 2021; Fan et al., 2016; Qiu & Benbasat, 2009; Sheehan et al., 2020). In the context of product recommendation, Verhagen et al. (2014) examined friendliness and expertise as determinants of social presence and personalization and studied their impacts on customers' interactions with virtual customer service agents. Several studies have suggested that anthropomorphism may lead to unintended consequences despite being generally beneficial (Duffy, 2003; Schanke et al., 2021; Schuetzler et al., 2018; Yam et al., 2021). Other studies have focused on building effective human-robot interactive systems, qualitatively discussing the potential value of social presence through case studies, and incorporating social presence as a factor in the design science framework (Gnewuch et al., 2017; Pereira et al., 2014; van Doorn et al., 2017; Yamaguchi et al., 2003).

When the real identity of a counterpart is not disclosed, identity cues or social context cues (e.g., demographic and personal traits) are key contextual factors that can influence individuals' perceptions of social presence (Oh et al., 2018). A few studies have concluded that one's perceived social presence crucially depends on identity cues in text-based computer-mediated communication. For instance, including support-seekers' portrait pictures and first-name IDs in online support forums can increase the perception of social presence and help them receive more person-centered and more polite support messages (Feng et al., 2016). Through an analysis of the 2006 Pew Internet and American Life Survey of parents and teens, Mesch and Beker (2010) found that a lack of social context cues in computer-mediated communication may lead to a dehumanization perception of unseen counterparts and cause high levels of online self-disclosure. Our study fits into this stream of literature because the real identity of customer service agents on Twitter is unknown. Within this literature, the closest work to ours is that of Cheng and Pan (2021), who investigated how the inclusion of agents' humanized profiles in agent responses affected the general sentiment of public tweets and the number of complaint tweets. Our research question differed from theirs because we focused on the outcomes of individual customer service encounters, as measured by customers' willingness to engage, service resolution rate, and verbal aggression. Because of this important difference in the research question, Cheng and Pan (2021) conducted their analyses at the tweet level using all types of public tweets, including those unrelated to customer service (e.g., compliments and marketing engagement). In contrast, the focus of our research question led us to analyze customer service-related encounters and distinguish customers' initial tweets from their follow-up tweets to agents' responses, thereby conducting the analyses at the conversation level.

3. Hypotheses Development

In this section, we hypothesize how identity cues affect customer-agent interactions in service encounters by enhancing the humanization level.

3.1. Humanization and Service Outcomes

According to Oh et al. (2018), the provision of identity cues (e.g., first names or portrait pictures) serves as a contextual factor that influences social presence. Providing identity cues offers clues to the "true" identity of a remote partner so that individuals feel the presence of a "real person". The impact of identity cues is even more salient in social media customer service because it is primarily text-based. Since a first name usually suggests human authorship, the inclusion of first names as identity cues should improve customers' perceptions of agents' humanization levels.

An enhanced humanization level may affect service outcomes through two channels: *trust* and *empathy*. The connection between humanization and trust is probably deeply rooted in our evolutionary past; humans needed to trust and rely on each other to compensate for physiological limitations and survive in the wild. Directly related to our research context, previous literature has demonstrated that social presence positively correlates with customers' trust in a service provider (Cyr et al., 2007; Gefen & Straub, 2003; Hassanein & Head, 2004; Lankton et al., 2015). For instance, in the context of business-to-consumer Internet-based services, Gefen and Straub (2003) showed that social presence significantly enhanced customers' trust.

Interpersonal trust is crucial for supportive communication because it helps establish supportive relationships (Mortenson, 2009). Naturally, customers are more willing to seek and receive help from trustworthy service providers, especially in an online environment in which they may feel insecure due to limited information about their counterparts. Therefore, a higher perceived level of humanization by customers should increase their willingness to engage due to their enhanced level of trust in the agents. Upon customers' follow-up engagement, enhanced humanization can further affect the resolution of a service encounter irrespective of the *actual* level of expertise because trust plays a critical role in persuasion (Hovland & Weiss, 1951; Sternthal et al., 1978), and a critical aspect of customer service is persuading customers to forgive a firm's service failure or defective product.

Enhanced humanization can also affect service outcomes through the channel of empathy—the unique human capacity to understand and feel what another person is experiencing within their frame of reference. According to social presence theory, enhanced humanization reduces the psychological distance between an individual and their counterpart (Biocca, 1997; Kreijns et al., 2004; Oh et al., 2018). With closer psychological distance, customers would feel that an agent is more likely to step into their shoes and resonate with their requests. Because a customer's perceived empathy of an agent positively contributes to the customer's satisfaction (Tax et al., 1998), the customer should be more willing to engage with an agent perceived as more humanized and, upon engagement, be more likely to accept the agent's reasoning or apology.

The aforementioned arguments led to the following hypothesis for empirical testing.

Hypothesis 1: *The presence of an identity cue improves customer service interactions by increasing: (a) customers' willingness to engage, and (b) the chance of reaching a resolution.*

3.2. Humanization and Customer Verbal Aggression

To understand how humanization might affect customer verbal aggression, it is important to first comprehend the underlying motives of customer complaining behavior, which can be categorized as *goal-oriented* or *emotion-focused* (Kowalski, 1996). Driven by different motivations, customers may react differently to enhanced perceptions of humanization.

Goal-oriented customers complain in order to seek redress or economic compensation rather than for venting. Their emotional dissatisfaction is not the primary driver of their service requests. Since customer aggressiveness is often driven by emotion, a goal-oriented customer's verbal aggression is not necessarily affected by an agent's level of humanization. However, goal-oriented customers may behave more aggressively toward more humanized agents as a strategic move to better achieve their goals, since agents are more susceptible to emotional pressure once their names have been publicized. Considering the vast publicity enabled by social media platforms, customers are fully aware of the benefits of publicly shaming a brand, as evidenced by their preference for public posts over private messages on social media (He et al., 2019).

Unlike a goal-oriented customer, the complaining behavior of an emotion-focused customer is evoked by frustration and the desire to express emotional dissatisfaction (Kowalski, 1996). With such motivation, customers complain because the act of venting *itself* makes them feel better. The aggressiveness of emotion-focused complaints can be alleviated by a higher degree of perceived humanization of customer service agents for two reasons. First, Bandura et al. (1975) suggested that the aggressiveness of complaints partly depends on how humanized the recipient of the actions is perceived to be and that people prefer not to behave cruelly toward those they perceive as more humanized because of human empathy. The presence of a customer service agent's identity cue tends to make a customer more conscious of the thoughts and emotions of the humanized agent, who is the recipient of the customer's action and, as a result, the customer behaves less aggressively.

Second, Derks et al. (2008) showed that the anonymous nature of the interaction is a key determinant of the relative ease and frequency with which emotions are expressed in computer-mediated communication, and individuals tend to be more overt and explicit in expressing negative emotions in the absence of visible others in more anonymous interactions. Because identity cues create a less anonymous setting, customers would increase negative appraisal (i.e., are more aware of and pay more attention to the potential negative consequences of their emotional reactions); as a result, they behave less aggressively toward agents. For these two reasons, we believe that the aggressiveness of emotion-focused complaints can be alleviated by a higher degree of perceived humanization of customer service agents.

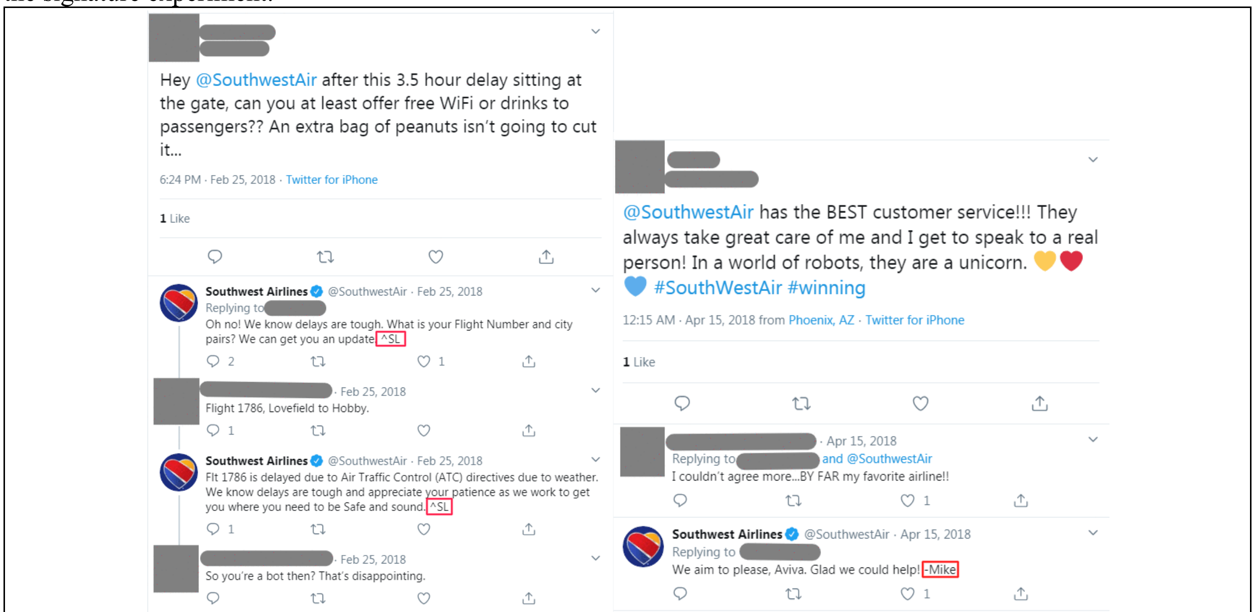
Since motivations can be mixed, a customer can be less aggressive towards more humanized agents—if an emotion-focused motive dominates a goal-oriented one—or more aggressive—if a goal-oriented motive dominates an emotion-focused one. In the latter case, humanization could have the side effect of hurting customer service employees' performance. As the literature suggests, customer service agents tend to respond to customers' aggressive behaviors by failing customer service provisions in return (Grandey et al., 2004; Rafaeli et al., 2012; Walker et al., 2017). For example, agents may respond to verbal aggression with aggressive language or simply ignore customers' requests. To examine potential unintended consequences of customers' enhanced humanization levels, we proposed the following competing hypotheses for empirical testing.

Hypothesis 2a: *The presence of an identity cue increases a customer's aggressiveness in a customer service interaction.*

Hypothesis 2b: *The presence of an identity cue decreases a customer's aggressiveness in a customer service interaction.*

4. A Quasi-Experiment

On March 16, 2018, the customer service agents of Southwest Airlines on Twitter began to include first names in each response to a customer tweet. Prior to this date, each agent's response was accompanied by a two-letter code following the carat/hyphen symbol. Figure 2 illustrates the difference in agent responses before and after the signature experiment.



Note: The left conversation is an example of customer service complaints ***before*** the signature experiment. The right conversation is an example ***after*** the signature experiment.

Figure 2. Sample Conversations Before and After the Signature Experiment

Figure 3 shows that the percentage of customer service agents using signatures switched from 0% to 100% on March 16, 2018; hence, the change was abrupt. More importantly, we did not find any advanced notice or discussion about the policy change on Southwest Airlines' official website; in its SEC filings, discussion forums, Twitter accounts; or in other news media. Therefore, the signature experiment was likely an exogenous shock, and customers did not anticipate the change. Furthermore, a search of various sources did not reveal any evidence of any events that co-occurred with the signature policy and could have affected customer service interactions. For instance, based on posts from the career discussion forum of Southwest Airlines (n.d.), no training sessions were provided for customer service representatives around the implementation of the signature experiment. In summary, the signature experiment offered us a quasi-experimental setting to investigate the effect of customers' perceived humanization level on customer service encounters.

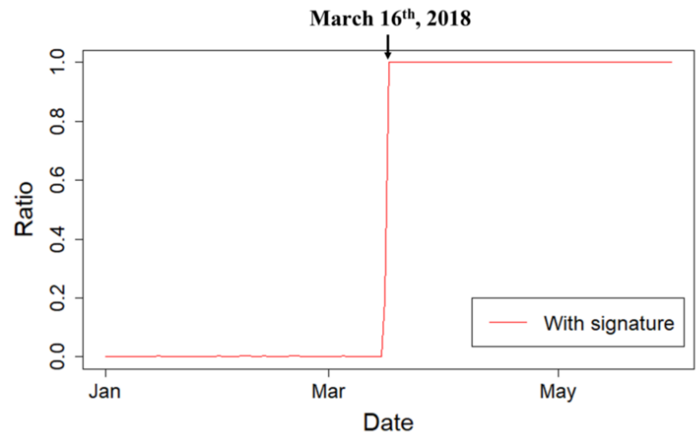


Figure 3. Daily Ratio of Agent Replies with A Signature

4.1. Data

The data consisted of all tweets received and posted by Southwest Airlines from February 16, 2018 to April 16, 2018. We used Twitter metadata to reconstruct entire conversations, starting with an initial tweet by a customer. To distinguish customer service-related conversations from other types of conversations, we defined customer service-related conversations as those that started with an inquiry or a complaint about an airline's service. Several types of conversations did not fall into the customer service category. For example, some customers have initiated a conversation with the airline to participate in a marketing event. Operationally, we hired an annotator to label 25,530 tweets and then trained a support vector machine (SVM) classifier using the labeled data. The performance of the SVM classifier is reported in Table A2 of Appendix B. Next, we used the classifier to identify customer service-related conversations in the data. We referred to these customer service-related conversations simply as *conversations*.

4.2. Variables

Table 1 provides a summary of the key variables and definitions. We considered three dependent variables in this study. To evaluate the hypothesis regarding the customer service outcomes (i.e., Hypothesis 1), we used *engagement* and *resolution* as outcome measures. To measure a customer's willingness to engage with an agent, we set *engagement* to one if a customer had follow-up interactions after an agent's first response to their initial tweet; otherwise, we set it to zero. As it is difficult to determine a resolution without customers' further engagement, we defined a binary variable, *resolution,* as whether a resolution was reached at the end of the conversation, conditional upon a customer's further engagement. We hired another annotator to read through all of the conversations and determine whether a resolution was reached. The labeling criteria are listed in Appendix A. To test the hypotheses regarding customer verbal aggression (i.e., Hypothesis 2a and 2b), we used the Python package *profanity-check* to construct *aggressiveness*, which captured customers' attitudes toward agents in customer service encounters. The higher the value, the more verbally aggressive the customer was.

The main source of endogeneity was a potential temporal shift in the quality of customer service provisions or the composition of customer service requests during the sample period. Our first line of defense against this threat to identification was to restrict the sample period to one month before and one month after the signature experiment. Doing so inevitably reduced statistical power, but also reduced potential bias because any temporal shift in service quality or service request composition should take some time.

Table 1. Conversation-level Summary Statistics

Variable	Obs.	Mean	S.D.	Definition
Outcome Measures				
engagement	8,214	0.40	0.49	Binary variable indicating whether a customer interacts with an agent after the agent's initial response
resolution	3,258	0.52	0.50	Binary variable indicating whether a resolution is reached at the end of a service encounter
aggressiveness	3,258	0.10	0.09	Continuous variable measuring how aggressive a customer is after an agent's first response
Circumstance				

initialAggressiveness	8,214	0.10	0.11	Continuous variable measuring a customer's aggressiveness in the initiated tweet
Customer Characteristics				
logFollowers	8,214	4.98	2.00	Log-transformed number of followers a customer has
logFollowings	8,214	5.49	1.46	Log-transformed number of followings a customer has
logUpdates	8,214	7.07	2.49	Log-transformed number of tweets a customer has posted
aggreeableness	8,214	7.37	1.08	Continuous variable measuring how agreeable a customer is
conscientiousness	8,214	-1.58	0.79	Continuous variable measuring how conscientious a customer is
extraversion	8,214	4.09	1.17	Continuous variable measuring how extrovert a customer is
neuroticism	8,214	1.59	0.91	Continuous variable measuring how neurotic a customer is
openness	8,214	-10.08	2.42	Continuous variable measuring how open a customer is
Agent Reply Quality				
responseTime	8,214	2.32	1.14	Log-transformed number of minutes it takes for an agent to respond to a customer
numReplies	8,214	1.34	0.69	Number of an agent's replies in a service encounter
avgWords	8,214	26.75	9.20	Average number of words per agent reply in a service encounter
DM	8,214	0.22	0.41	Binary variable indicating whether an agent mentions direct message in the responses
hello	8,214	0.15	0.36	Binary variable indicating whether an agent greets to a customer
gratitude	8,214	0.27	0.44	Binary variable indicating whether an agent appreciates a customer's tweets
apology	8,214	0.31	0.46	Binary variable indicating whether an agent apologizes to a customer
hedges	8,214	0.22	0.42	Binary variable indicating whether an agent shows uncertainty
please	8,214	0.21	0.41	Binary variable indicating whether an agent mentions "please" in the responses
request	8,214	0.01	0.09	Binary variable indicating whether the agent requests an action from the customer
Other Control				
googleTrend	8,214	87.73	6.72	The google trend index at time of Southwest Airlines
offlineIncident	8,214	0.07	0.26	Binary variable indicating whether there is any offline incident at time for Southwest Airlines
Note: *resolution* and *aggressiveness* are constructed only for conversations with customers' further engagement. Obs. Stands for the number of observations. S.D. stands for standard deviation.				

Our second line of defense was to control for a large number of conversation-level characteristics about the customer and the agent. On the customer side, we controlled for the characteristics of customers and their service inquiries. For every customer, we controlled for the number of followers, the number of followings, and the number of updates. Following the literature (Gunarathne et al., 2017; Yarkoni, 2010), we derived customers' Big Five personality traits (i.e., agreeableness, conscientiousness, extraversion, neuroticism, and openness) based on the Linguistic Inquiry and Word Count (LIWC) dictionary and customers' historical tweets on their public Twitter pages. We then extracted various features from customers' service inquiries. First, we used the latent semantic analysis (LSA) and the K-means clustering algorithm to group similar tweets into seven clusters, which was the optimal number of clusters suggested by the silhouette score (Rousseeuw, 1987). This allowed us to alleviate concerns regarding unobserved confounding factors associated with different types of customer service circumstances. Details about the clustering method are reported in Appendix C. We constructed *initialAggressiveness* as the proxy for customers' aggressiveness at the beginning of the conversation, which might affect the customer service interactions.

On the agent side, we controlled for *responseTime, numReplies,* and *avgWords* to capture the quality and efficiency of agents' interventions. We used a lexicon-based method to create *DM*, a binary variable that was equal to one if an agent mentioned keywords that are used to request a private conversation with a customer, such as "*direct message*". Following Yeomans et al. (2019), we created a list of dummy variables to quantify the politeness of agent replies, which could directly affect the effectiveness of a service interaction. For instance, *apology* indicated whether an agent apologized to a customer in a reply and *gratitude* indicated whether an agent expressed appreciation to a customer in a response.

Table 2 reports the balance checks of the conversation-level control variables before and after treatment. The absolute standardized differences were all below the threshold of 0.1, thus suggesting the covariate balance (Austin, 2009). Importantly, such a balance was naturally achieved without any matching, supporting the validity of the quasi-experiment setting and alleviating the concern of imbalance due to unobserved confounding factors.

Table 2. Balance Check

Variable	Pre-treatment		Post-treatment		Std. Diff.
	Mean	S.D.	Mean	S.D.	
initialAggressiveness	0.099	0.109	0.102	0.111	-0.027
logFollowers	4.998	1.979	4.967	2.016	0.016
logFollowings	5.517	1.446	5.468	1.468	0.034
logUpdates	7.120	2.466	7.028	2.511	0.037
agreeableness	7.397	1.077	7.345	1.089	0.048
conscientiousness	-1.574	0.768	-1.582	0.801	0.010
extraversion	4.114	1.162	4.071	1.185	0.037
neuroticism	1.585	0.908	1.596	0.920	-0.012
openness	-10.11	2.407	-10.06	2.436	-0.021
responseTime	2.309	1.146	2.327	1.142	-0.016
numReplies	1.316	0.628	1.367	0.735	-0.075
avgWords	26.65	9.360	26.83	9.074	-0.020
DM	0.217	0.412	0.220	0.414	-0.007
hello	0.272	0.445	0.271	0.445	0.002
gratitude	0.143	0.350	0.157	0.364	-0.039
apology	0.315	0.465	0.307	0.461	0.017
hedges	0.211	0.408	0.231	0.422	-0.048
please	0.201	0.401	0.211	0.408	-0.025
request	0.009	0.093	0.008	0.088	0.011

Note: This table reports the before-after difference in means of the key covariates in the analyses. S.D. stands for standard deviation. Std. Diff. stands for standardized difference. Austin (2009) suggested that an absolute standardized difference of 0.10 or more indicates that covariates are imbalanced between groups.

5. One-Group Before-And-After Analysis

To test the impact of the humanization level change, we started with the standard one-group before-and-after design following Claussen et al. (2013). We estimated the following linear probability model for dichotomous outcome measures (i.e., *engagement* and *resolution*) and ordinary least squares for the continuous outcome measure (i.e., *aggressiveness*) at the conversational level indexed by i:

$$Y_i = \beta_0 + \beta_1 signature_i + \beta_2 X_i + \beta_3 Z_t + \beta_4 HourofDay_t + \beta_5 DayofWeek_t + \beta_6 TimeTrend_t + \varepsilon_{i,t}$$

The main variable of interest is *signature*, whose coefficient β_1 captures the effect of enhanced humanization. We controlled for conversation-specific characteristics, X_i, which includes the circumstance, customers' characteristics, and agents' service quality. We also controlled for time-varying airline characteristics, Z_t, which includes the Google Trend index and the number of offline incidents at time t. We included the linear time trend, day-of-week fixed effects, and hour-of-day fixed effects to account for seasonality.

5.1. Baseline Results

Table 3 reports the regression results with four different estimation windows around the event date. The rationale for using different estimation windows is to alleviate the endogeneity concerns of unobserved confounding events or temporal shifts in unobserved confounding factors during the sample period. Shortening the estimation window alleviates such a concern at the expense of reduced statistical power.

Columns 1, 4, 7, and 10 show the estimation results when the dependent variable is a customer's willingness to engage. The coefficient of *signature* is positive and statistically significant, suggesting that including agents' signatures in the response increases customers' probability of engaging with agents. Columns 2, 5, 8, and 11 show the results when the dependent variable is *resolution*. Regardless of the difference in estimation windows, the coefficient of *signature* is positive and significant, indicating that enhanced humanization increases the likelihood of reaching a resolution. Therefore, **Hypothesis 1** is supported.

Columns 3, 6, 9, and 12 show the results regarding customers' attitudes change towards agents. Most coefficients of *signature* remain insignificant; hence, neither **Hypothesis 2a** nor **2b** is supported. The observed null effect is probably due to the mixing of customers' goal-oriented and emotion-focused motives, which affect customers' aggressiveness in opposite directions. Although we could not infer the exact composition of competing motives, the null effect at least suggests that the benefits of enhanced humanization do not come at the expense of an overall increased cost on customer service agents in the form of elevated customer verbal aggression.

Table 3. Baseline Results

	± 1 month			**± 3 weeks**		
	engagement	*resolution*	*aggressiveness*	*engagement*	*resolution*	*aggressiveness*
	(1)	(2)	(3)	(4)	(5)	(6)
signature	0.0842*** (0.0216)	0.0695** (0.0352)	-0.0093* (0.0050)	0.1177*** (0.0254)	0.0822* (0.0425)	-0.0083 (0.0058)
Controls	Y	Y	Y	Y	Y	Y
Time Trend	Y	Y	Y	Y	Y	Y
Seasonality FE	Y	Y	Y	Y	Y	Y
Observations	8214	3258	3258	5771	2249	2249
R^2	0.06	0.14	0.07	0.07	0.15	0.09
	± 2 weeks			**± 1 week**		
	engagement	*resolution*	*aggressiveness*	*engagement*	*resolution*	*aggressiveness*
	(7)	(8)	(9)	(10)	(11)	(12)
signature	0.2256*** (0.0378)	0.2082*** (0.0686)	0.0074 (0.0082)	0.2262*** (0.0632)	0.2810** (0.1071)	-0.0092 (0.0123)
Controls	Y	Y	Y	Y	Y	Y
Time Trend	Y	Y	Y	Y	Y	Y
Seasonality FE	Y	Y	Y	Y	Y	Y
Observations	3885	1518	1518	2010	744	744
R^2	0.08	0.16	0.10	0.10	0.19	0.12

Note: *** $p < 0.01$ ** $p < 0.5$ * $p < 0.1$. Robust standard errors in parentheses. This table reports one-group before-and-after regression results with different estimation windows using conversations from Southwest Airlines on Twitter. *Engagement* is a binary variable indicating whether a customer is willing to engage with an agent. *Resolution* is a binary variable indicating whether a resolution is reached at the end of the interaction. *Aggressiveness* is a continuous variable measuring a customer's attitude toward agents. For *resolution* and *aggressiveness*, we include only conversations with customers' further engagement in the sample.

6. Robustness Checks

Although intuitive, the one-group before-and-after research design may suffer from endogeneity issues due to time-varying confounding factors that are not systematically modeled. We alleviated such concerns with a series of falsification tests and robustness checks.

6.1. Falsification Tests with Pseudo Treatments

Besides varying the window size, another way to alleviate the concern of unobserved confounding events is to conduct falsification tests with pseudo treatments. First, we assumed two pseudo treatments: one was on March 1, 2018, two weeks before the event date, and the other was on March 8, 2018, one week before the event date (see Figure A2 of Appendix D for an illustration). If the identified effects were largely due to some unobserved performance-improvement initiatives before the event date, then, by assuming a pseudo treatment before the actual policy change, our econometric model would falsely detect similar effects as the baseline analysis. We estimated the regression model with those pseudo treatments and reported the results in Table A4 of Appendix D. From the insignificant coefficient estimates of *signature* in both tests, we concluded that there was no evidence that our main findings were driven by unobserved events before the signature experiment. Second, we considered a potential confounding factor that was right at or very close to the time of the signature experiment. If our findings were driven by unobserved seasonality specific to Southwest Airlines (e.g., an annual event by the airline around the time of the signature experiment), then we should falsely detect the humanization effect for Southwest Airlines on March 16, 2017 (i.e., exactly one year before the signature experiment). To implement this idea, we assumed a pseudo treatment on March 16, 2017 and conducted a falsification test using customer service-related conversations from Southwest Airlines in 2017. Table A5 of Appendix D shows the estimation results. Again, we did not find any significant effects on any of the outcome variables, which largely ruled out the possibility that our main results were due to unobserved seasonality at the time of the signature experiment.

6.2. Entropy Balancing and Coarsened Exact Matching

The one-group before-and-after design requires that the treated and control groups are comparable over time in the absence of treatment (Meyer, 1995). Although the balance check (see Table 2) indicated the comparability of the treated and control groups, we further balanced the sample using two matching methods and checked the robustness of our findings. One method was Entropy Balancing (EB), which relies on a maximum entropy reweighting scheme to produce a more balanced sample (Hainmueller, 2012). The other method was the coarsened exact matching (CEM), which coarsens the observed variables (i.e., the circumstance, customers' characteristics, and agent reply quality) before applying the exact matching on the coarsened data to determine the matches (Iacus et al., 2012).[3] The estimation results using both methods, which are consistent with the baseline results, are reported in Table A6 of Appendix D.

6.3. Synthetic Control

A standard approach in the literature for alleviating the concerns of unobserved time-varying confounding factors is to employ a difference-in-differences (DID) strategy. However, the critical parallel trend assumption for this approach was not supported for other major airlines (see Table A10 in Appendix E for a demonstration). An increasingly popular solution for overcoming a lack of natural control is to construct a synthetic control from the so-called donor pool of candidate controls (Abadie et al., 2010; Abadie & Gardeazabal, 2003; Athey et al., 2021; Xu, 2017). We adopted this synthetic control method to further test the hypotheses.

To this end, we considered customer service-related conversations from American Airlines, Delta Airlines, JetBlue Airlines, and United Airlines as potential control units. We included these airlines in the donor pool because each of them had either similar offline operations or a similar online presence to Southwest Airlines (see Table A9 of Appendix E), and none of these airlines initiated the signature policy during our sample period. We obtain all tweets received and posted by these airlines and determined the resolution of their service encounters using a supervised classifier trained with labeled data from Southwest Airlines. The performance of the classifier is reported in Table A3 of Appendix B. Since the synthetic control method required a panel setting, we constructed an airline-day panel by aggregating the conversation-level data at the daily level for each airline. Specifically, $engagement_{f,t}$ measured the ratio of conversations with customers' further engagements for airline f at day t, $resolution_{f,t}$ captured the ratio of conversations with a resolution for airline f at day t, and $aggressiveness_{f,t}$ indicated the average of customers' aggression toward airline f at day t. The control variables were similarly aggregated.

The criterion for evaluating the quality of the synthetic control was how well the treated unit matched the synthetic control in the pre-treatment periods (Abadie et al., 2010, 2015). As seen in Row 5 of Table A10, there were insignificant differences in the dependent variables between the treated and synthetic control before the treatment, suggesting a high quality of the synthetic control.

Table A7 of Appendix D reports the estimation results. As shown in Column 1, the positive and significant coefficient of $signature_{f,t}$ suggests that a higher proportion of customers are willing to interact with service agents after the launch of the signature experiment. Similarly, the significantly positive coefficient of $signature_{f,t}$ in Column 2 implies that a higher percentage of conversations reach a resolution after the launch of the signature experiment. Consistent with our baseline results, the insignificant coefficient of $signature_{f,t}$ in Column 3 shows that there is no evidence of increased (or decreased) verbal aggression by customers after the policy change. In summary, despite different units of analysis, the findings derived from the synthetic control method are consistent with the one-group before-and-after analysis.

6.4. Two-Way Matching

Since synthetic control was constructed at the aggregate level, conversation-level characteristics such as customer characteristics and agent response quality may not be fully accounted for. Therefore, we proposed a two-way matching method at the conversation level. We denoted the vector of the observed covariates and the outcome for a conversation relating to the treated airline (i.e., Southwest Airlines) before the treatment using X_{t0} and Y_{t0}, respectively. The observable covariates for X_{t0} included circumstance (*initialAggressiveness* and *content cluster*), customer characteristics (*logFollowers*, *logFollowings*, and *logUpdates*), and agent response quality (*responseTime*,

[3] To avoid the curse of dimensionality, we followed Broockman (2013) by selecting the following important covariates for the CEM procedure: *content cluster*, *logFollowers*, *DM, hello, gratitude, apology, hedges, please,* and *request.* We also conducted several robustness checks with different sets of covariates and the results remained qualitatively the same. The results are available upon request.

numReplies, *avgWords*, *DM*, *hello*, *gratitude*, *apology*, *hedges*, *please*, and *request*). For a conversation related to the treated airline after the treatment, we denoted the vector of observed covariates and the outcome using X_{t1} and Y_{t1}, respectively. In a similar fashion, we used the notations X_{c0}, Y_{c0}, X_{c1}, and Y_{c1} for conversations selected before and after the date of the signature policy from the donor pool.

$$
\begin{tikzcd}
& \textbf{Pre-treatment} & \textbf{Post-treatment} \\
\textbf{Treated} & (X_{t0}, Y_{t0}) \arrow[r, "2"] \arrow[d, "1"] & (X_{t1}, Y_{t1}) \arrow[d, "1"] \\
\textbf{Control} & (X_{c0}, Y_{c0}) & (X_{c1}, Y_{c1})
\end{tikzcd}
$$

Figure 4. Two-way Matching Procedure

The proposed two-way matching method is illustrated in Figure 4. First, for each (X_{t0}, Y_{t0}), we used the Mahalanobis distance and one-to-one matching to identify a matched conversation related to a control airline in the pre-treatment period, (X_{c0}, Y_{c0}). To ensure matching quality, we only kept the matched pairs whose distance between X_{t0} and X_{c0} was within a specified caliper. For instance, if the caliper was equal to three, the matched pairs whose distance is more than three times the standard deviation of all distances among the matched pairs will be dropped. The difference in outcomes for the treated group (i.e., Southwest Airlines) and the control group before the signature experiment, $\Delta Y_0 = Y_{t0} - Y_{c0}$, was then calculated. Similarly, by matching conversations in the post-treatment period, we obtained $\Delta Y_1 = Y_{t1} - Y_{c1}$. Next, we match each ($X_{t0}$, ΔY_0) with a (X_{t1}, ΔY_1) based on the Mahalanobis distance between X_{t0} and X_{t1}. For each matched tuple (X_{t0}, X_{t1}, ΔY_0, ΔY_1), we calculated the treatment effect as $\Delta Y_1 - \Delta Y_0$. A t-test was then conducted to check if the treatment effect was significantly different from zero.

Table A8 of Appendix D reports the results. Following previous studies (Aral et al., 2009; Goes et al., 2014), we employed two calipers in one-to-one matching. Take the first set of results (i.e., caliper = 3), for example. In Column 1, the significantly positive coefficient of *signature* suggests that, upon receiving a response, a customer is more willing to engage with a customer service agent perceived as more humanized. In Column 2, the significantly positive coefficient of *signature* suggests that a customer is more likely to reach a resolution with a customer service agent perceived as more humanized. Moreover, the insignificant coefficient of *signature* in Column 3, when the dependent variable is *aggressiveness*, again suggests the lack of evidence that enhanced humanization increases (or decreases) customer verbal aggression. In summary, the results from the proposed two-way matching analysis reinforce our results from the one-group before-and-after analysis and the synthetic control analysis.

7. Mechanism Tests

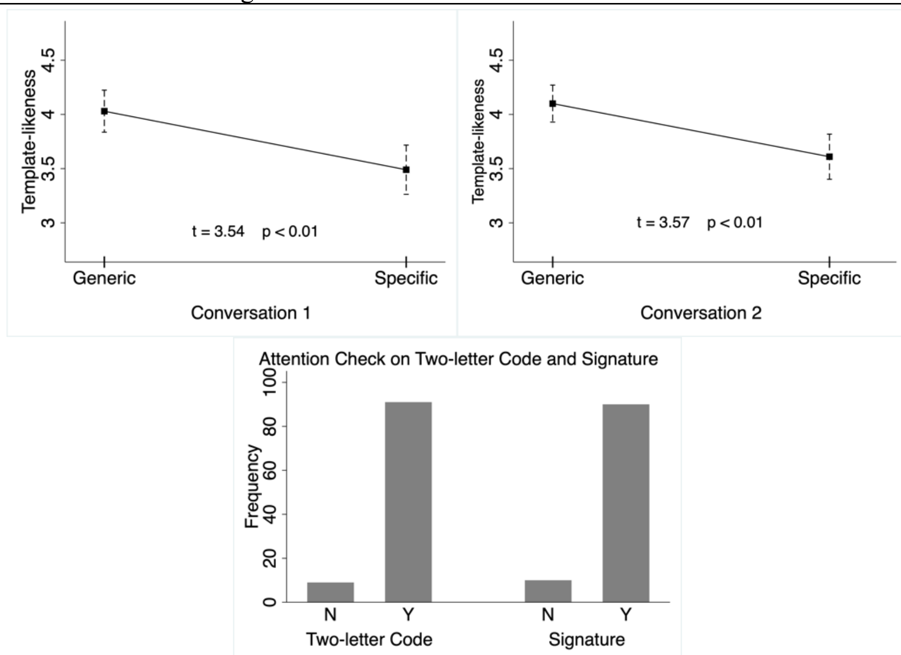
This section explores the mechanisms behind the effect of identity cues on service outcomes (i.e., Hypotheses 1 and 2).

7.1. Mechanism Test on Empathy and Trust

7.1.1. Experimental Design and Manipulation Check. To explore the mechanism underlying the effects of identity cues on customer engagement and satisfaction (i.e., Hypothesis 1), we conducted a between-subjects randomized experiment on AMT. We adopted a factorial design consisting of two factors: 1) whether an agent's response was specific or generic and 2) whether a signature was included at the end of a reply. For the first factor, we designed two levels of agent responses (i.e., specific reply and generic reply) based on real-world customer service conversations on Twitter. We showed each respondent two service encounters initiated by two types of common requests (i.e., flight delay and lost baggage). Table A12 in Appendix F shows four examples of conversations that differ from each other in terms of service requests and reply styles. The second factor consisted of two levels: a service response that was followed by a two-letter code (e.g., -MR and -RR) and one that was followed by a first name (e.g., -Michael and -Rachel).

To evaluate how participants perceived these two factors, we conducted a manipulation check via a pilot experiment on AMT. Specifically, we surveyed 200 respondents and randomly assigned them to one of four experimental conditions: (1) generic replies with two-letter codes, (2) generic replies with signatures, (3) specific replies with two-letter codes, and (4) specific replies with signatures. In each condition, respondents were presented with two customer service-related conversations, one about a flight delay and the other about a baggage delay. After presenting a conversation, we asked the respondents to answer the following question on a five-point Likert scale (1 = *strongly disagree* to 5 = *strongly agree*): "*Do you agree that the agent's responses are generic because they are template-like?*" After presenting both conversations, we asked respondents in the two-letter code condition the following question: "*Have you noticed the two-letter codes (e.g., -RR, -MR) at the end of agent responses?*" We asked respondents in the signature condition the following question: "*Have you noticed the signatures (e.g., -Rachel, -Michael) at the end of agent responses?*" These three questions were raised in separate pages of an online survey. Respondents were not allowed to return to previous pages.

The top two panels of Figure 5 display the results of a two-sample t-test of customers' perceptions of the agents' reply type (i.e., specific vs. generic). Regardless of the differences in the service encounter contexts, customers consistently perceived generic replies as more generic, suggesting the successful manipulation of the reply type. The bottom panel of Figure 5 shows the number of respondents who did and who did not notice the two-letter codes and the signatures. In total, 91 of the 100 respondents assigned to the two-letter code condition noticed the codes, and 90 of the 100 respondents assigned to the signature condition noticed the signature, validating their attention to the two-letter codes and signatures.



Note: The error bars of the estimate for treatment effects indicate the 95% confidence intervals.

Figure 5. Manipulation Check

7.1.2. Experimental Procedure. We recruited 600 AMT workers who owned Twitter accounts and randomly assigned them to one of four experimental conditions. At the beginning of the experiment, we randomly asked 400 respondents the following question, "*Do you think social media customer service agents are human agents or algorithm-enabled agents?*" Participants who viewed the question were asked to report their prior beliefs about agent identity on a five-point Likert scale (i.e., "*Extremely likely to be human agents,*" "*Likely to be human agents,*" "*Either human agents or algorithm-enabled agents,*" "*Likely to be algorithm-enabled agents,*" and "*Extremely likely to be algorithm-enabled agents*"). A customer's prior belief about an agent's human identity is a critical moderator in our study because identity cues work by enhancing the customer's perceived humanization level, and such an enhancement clearly depends on the customer's prior belief. For instance, compared to customers who firmly believe in the human identity of agents, suspicious customers could be more susceptible to the signature policy. We excluded this question from a randomly selected portion of respondents (i.e., 200 participants) so that we could investigate whether raising this identity question changed participants' awareness and diverted their responses. A statistically significant difference in the responses between customers *with* and *without* exposure to the identity question may cast doubt on the validity of further analyses.

The experiment consisted of several steps. First, all participants provided their demographic information. Then, two-thirds were randomly chosen to report their perceptions of service agents' identities. After that, we displayed the first part of the two service encounters (i.e., the initial customer request and the agent's first reply) and asked about the respondents' willingness to engage based on the agent's first response if they were the focal customer. Finally, we presented respondents with the complete conversation and asked them to evaluate the agent's humanization level, the perceived level of trust and empathy based on agent responses, and their overall satisfaction with the service encounter. We randomized the order of the service encounters to avoid any order effect and measured humanization, trust, empathy, and satisfaction on a five-point Likert scale using measurement items adapted from previous studies (Barger & Grandey, 2006; Coulter & Coulter, 2002; Einwiller, 2003; Fang et al., 2014; Garbarino & Lee, 2003; Jarvenpaa et al., 2000; Yoo & Alavi, 2001). Table A11 in Appendix F provides the full list of measurement items.

7.1.3. Results. After collecting the survey responses, we assessed the potential effect of the identity question using a two-sample t-test. The results failed to reject the null hypothesis that the exposure to the identity question affects respondents' evaluation of the humanization level (diff. = -0.07, SD = 0.07, $p > 0.10$), the engagement level (diff. = -0.01, SD = 0.08, $p > 0.10$), and satisfaction level (diff. = 0.02, SD = 0.10, $p > 0.10$). The similar responses between the two groups of participants corroborated the validity of the experimental design, implying that the collection of customers' prior beliefs about agents' identities does not alter participants' behaviors in survey responses. This laid the foundation for the follow-up test regarding customers' prior beliefs as a moderator of the effect of the signature policy.

For the rest of the analyses, we focused on the 400 respondents who were exposed to the identity question. The top left panel of Figure 6 displays the distribution of respondents' prior beliefs about agents' identities. As shown, 38% of the 400 respondents were either uncertain about the agent's identity or confident that the agent was algorithm-enabled. The balance check results reported in Table 4 suggest that there was no significant difference between treatment conditions in terms of demographic attributes, such as age, gender, and education.

Table 4. Balance Check on Demographics

	Specific vs. Generic			Signature vs. Two-letter Code		
	Specific	Generic	Diff.	Signature	Two-letter Code	Diff.
Age	2.95 (0.06)	2.93 (0.05)	0.02 (0.09)	2.92 (0.06)	2.96 (0.07)	-0.04 (0.09)
Gender	1.46 (0.04)	1.49 (0.04)	-0.03 (0.05)	1.51 (0.04)	1.44 (0.04)	0.07 (0.05)
Education	3.15 (0.06)	3.11 (0.06)	0.04 (0.09)	3.12 (0.06)	3.13 (0.06)	-0.01 (0.09)

Note: Standard errors are reported in the parentheses. The number of observations for each treated and control group is 200.

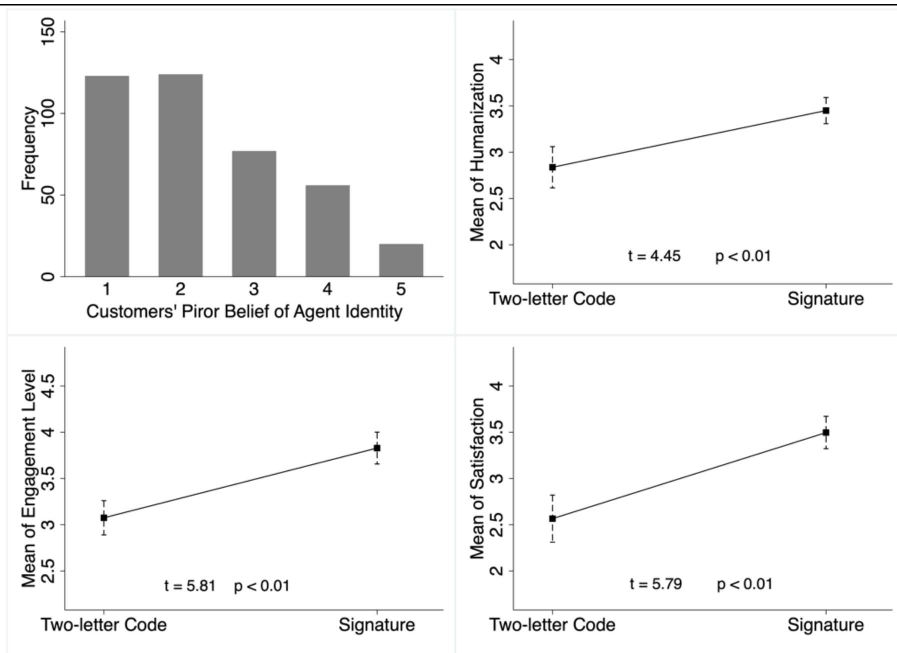
To explore whether the inclusion of signatures could update customers' perceptions, we examined the effect of signatures conditional on customers' prior beliefs about service agents' identities. We then conducted two-sample t-tests to examine the effect of agent signature conditional on individuals' prior beliefs. The estimation results reported in Row 1 of Table 5 suggest that there was barely any effect on customers who firmly believed in the agent's human identity. In contrast, for customers who were uncertain or suspicious about the agent's human identity (Row 4 of Table 5), we observed significant treatment effects. Specifically, customers perceived agents who included signatures in their replies as more humanized, and they became more willing to engage with agents and reached higher levels of satisfaction. Figure 6 visually demonstrates the positive treatment effects of signatures on customers who were suspicious about agents' human identity. In other words, the observed effect of agent signatures worked by updating the perceptions of customers who were either uncertain or suspicious about the human identity of agents. Since a generic response is more likely to be perceived as coming from non-human agents than a specific response, we further decomposed the effect of agent signatures by reply styles. As shown in Rows 5 and 6 of Table 5, despite having the same direction for all outcome measures, the treatment effects were significantly stronger when the agent's reply style was generic.

Table 5. Effects of Agent Signature Conditional on Customers' Prior Beliefs of Agents' Identity

	Prior Belief	Reply style	Obs.	Humanization			Engagement			Satisfaction		
				Treated	Control	Diff.	Treated	Control	Diff.	Treated	Control	Diff.
(1)	Human	Both	247	3.35 (0.07)	3.45 (0.07)	-0.10 (0.10)	3.96 (0.08)	3.93 (0.08)	0.03 (0.12)	3.41 (0.11)	3.67 (0.09)	**0.26* (0.14)**
(2)	Human	Generic	116	3.13 (0.12)	3.29 (0.12)	-0.16 (0.17)	3.75 (0.12)	3.76 (0.14)	-0.01 (0.19)	3.14 (0.16)	3.49 (0.16)	-0.35 (0.23)
(3)	Human	Specific	131	3.57 (0.08)	3.57 (0.08)	0.004 (0.11)	4.18 (0.10)	4.07 (0.10)	0.12 (0.14)	3.68 (0.13)	3.80 (0.11)	-0.12 (0.16)
(4)	Uncertain or algorithm-enabled	Both	153	3.45 (0.07)	2.84 (0.11)	**0.61*** (0.14)**	3.83 (0.09)	3.08 (0.09)	**0.75*** (0.13)**	3.50 (0.09)	2.57 (0.13)	**0.93*** (0.16)**

(5)	Uncertain or algorithm-enabled	Generic	84	3.42 (0.10)	2.48 (0.15)	**0.94*** (0.19)**	3.68 (0.14)	2.81 (0.13)	**0.87*** (0.19)**	3.51 (0.13)	2.09 (0.15)	**1.43*** (0.21)**
(6)	Uncertain or algorithm-enabled	Specific	69	3.48 (0.10)	3.38 (0.13)	0.10 (0.16)	3.97 (0.10)	3.47 (0.10)	**0.50*** (0.14)**	3.48 (0.12)	3.28 (0.17)	0.20 (0.20)

Note: * *p<0.10* ** *p<0.05*, *** *p<0.01*. Standard errors are reported in the parentheses. Obs. Stands for the number of observations. The treated group corresponds to respondents who receive agent replies with a signature and the control group corresponds to two-letter codes.



Note: The x-axis of the top left panel shows the distribution of customers' prior beliefs of agent identity before the experiment: 1 corresponds to "*Extremely likely to be human agents,*" 2 corresponds to "*Likely to be human agents,*" 3 corresponds to "*Either human agents or algorithm-enabled agents,*" 4 corresponds to "*Likely to be algorithm-enabled agents,*" and 5 corresponds to "*Extremely likely to be algorithm-enabled agents*". The rest plots show the treatment effects of the signature experiment on participants who were either uncertain or suspicious about the human identity of agents. The error bars of the estimates for treatment effects indicate the 95% confidence intervals.

Figure 6. Effects of Agent Signature When Customers are Uncertain or Suspicious about Agents' Human Identity

7.1.4. Causal Mediation Analysis. To empirically test the mechanisms (i.e., trust and empathy) underlying the effect of agent signature on engagement and satisfaction, we followed Li et al. (2022) by conducting a causal mediation analysis. We estimated the following equations and decomposed the effect of agent signatures into two effects—the natural direct effect (NDE) and the natural indirect effect (NIE):

$$M = \alpha_0 + \alpha_1 * T + \varepsilon_1$$

$$Y = \beta_0 + \beta_1 * T + \beta_2 * M + \beta_3 * T * M + \varepsilon_2$$

where M is the mediator, T is the treatment variable, and Y is the outcome variable. NDE measured how much the outcome would change on average if the treatment condition would change from two-letter codes to agent signatures, but the values for the mediator remain fixed at the level when the treatment condition is a two-letter code (i.e., $E[M \mid T = 0]$). NIE measured how much the outcome would change on average if the treatment condition were agent signatures, but the mediator is changed from the level it would have with two-letter codes (i.e., $E[M \mid T = 0]$) to the level it would have with agent signatures (i.e., $E[M \mid T = 1]$).

We used the *paramed* package in STATA to compute NDE, NIE, and the corresponding confidence intervals. Table 6 reports the estimation results, and Tables A13 and A14 of Appendix G report the corresponding regression results. As the first row of Table 6 shows, the estimated NDE is 0.30 and statistically significant at the 95% level (the bootstrapped 95% confidence interval lies above 0), and the estimated NIE is 0.45 and statistically significant. The evidence suggests that the inclusion of agent signatures increased customers' willingness to engage by enhancing their trust in the agent. In other words, trust mediated the effect of agent signatures on customer engagement. Moreover, the mediation effect of trust is a complimentary mediation because direct and indirect effects are both significantly positive (Zhao et al., 2010).

Similarly, the results in the second row of Table 6 suggest that empathy mediated the effect of signatures on engagement in a complementary way. Besides these single-mediator analyses, we conducted a multiple mediation analysis by jointly considering trust and empathy as mediators. As shown in Row 3 of Table 6, direct and indirect effects both remain positive and significant, suggesting that trust and empathy together mediated the effect of agent signatures on customer engagement. In other words, the inclusion of agent signatures increased customer engagement by improving customers' perceived levels of trust and empathy from the agent. In a similar vein, the last three rows of Table 6 report the results of causal mediation analyses when the outcome is customer satisfaction and indicate that trust and empathy together mediated the effect of agent signatures on customer satisfaction.

Table 6. Causal Mediation Analysis

Dependent Variable	Mediator	Direct Effect	Indirect Effect
Engagement	Trust	0.30 [0.01, 0.59]	0.45 [0.26, 0.65]
Engagement	Empathy	0.44 [0.15, 0.72]	0.32 [0.15, 0.49]
Engagement	Trust + Empathy	0.44 [0.11, 0.77]	0.29 [0.09, 0.50]
Satisfaction	Trust	0.40 [0.13, 0.67]	0.53 [0.35, 0.70]
Satisfaction	Empathy	0.53 [0.28, 0.79]	0.40 [0.24, 0.56]
Satisfaction	Trust + Empathy	0.63 [0.36, 0.90]	0.30 [0.10, 0.49]

Note: The 95% confidence intervals are reported in the brackets. All confidence intervals are bootstrapped with bias correction.

Additionally, we conducted a sensitivity analysis using the *mediation* package in STATA to examine how robust the detected mediation effect was to unobserved confounding factors that could affect the outcome and the mediator (Hicks & Tingley, 2011). Figure A3 of Appendix G reports the results of the sensitivity analysis. The sensitivity parameter is the correlation between the residual terms of the mediator and the outcome model (i.e., ε_1 and ε_2), which would not be zero if there are unobserved confounders (Imai et al., 2010). We can see that the sensitivity parameter needs to be large enough for the detected mediation effects to be zero. Take the top-left panel of Figure A3, for example. For the average mediation effect to disappear or change its sign, the sensitivity parameter must be greater than 0.40.

7.2. Mechanism Test on Expressiveness

Since neither Hypothesis 2a nor Hypothesis 2b was supported, we explored potential explanations for the null effect of humanization on customers' aggressiveness. It is challenging, if possible at all, to test the underlying mechanism through a controlled experiment because verbal aggression is a much more subjective and personal experience than the other outcome variables (i.e., engagement and resolution). Hence, we again resorted to the observational data from Southwest's signature experiment.

We implemented the mechanism test by distinguishing expressive from non-expressive customers, corresponding to the emotional-focused and goal-oriented customers in Hypothesis 2a and 2b (Kowalski, 1996). To this end, we used a customer's initial tweet to construct a binary variable *expressive* which was set to one if the customer was purely venting without mentioning any remedy.[4] As it is unlikely to fully infer customers' underlying motivations based on textual information, tweets that were not identified as expressive may be either primarily goal-oriented or had mixed motivations. Table 7 reports the results with different estimation windows.

Table 7. Heterogeneous Effect of Humanization on Aggressiveness

	aggressiveness

[4] We used the term "expressive" to be consistent with the literature convention, which refers to emotion-focused complaints as expressive and goal-oriented complaints as instrumental (Kowalski, 1996). Regarding implementation, we built an SVM classifier based on a labeled data set (i.e., 1700 tweets with labels indicating whether they were expressive). The precision, recall, and F1 score of this classifier on the test set are 0.76, 0.85, and 0.80, respectively. We conducted two sets of analyses to alleviate concern about measurement errors. First, we generated the expressive measure using another classification model: the multinomial logit model (MLC). The estimation results remained consistent. Second, we considered the expressive measures constructed by SVM and MLC as replicate measures of consumer expressiveness and then applied SIMEX to estimate the regression model. The results remained qualitatively the same (Carroll & Stefanski, 1990; Cook & Stefanski, 1994).

	± 1 month	± 3 weeks	± 2 weeks	± 1 week
	(1)	(2)	(3)	(4)
signature	-0.0042 (0.0055)	-0.0010 (0.0065)	0.0198** (0.0096)	-0.0020 (0.0124)
expressive	0.0195*** (0.0046)	0.0221*** (0.0054)	0.0269*** (0.0069)	0.0248*** (0.0080)
signature × *expressive*	-0.0109* (0.0060)	-0.0157** (0.0071)	-0.0244*** (0.0090)	-0.0163* (0.0092)
Controls	Y	Y	Y	Y
Time Trend	Y	Y	Y	Y
Seasonality FE	Y	Y	Y	Y
Treatment Effect for the Expressive Type	-0.0151** (0.0062)	-0.0166** (0.0071)	-0.0046 (0.0090)	-0.0183 (0.0134)
Observations	3258	2249	1518	744
R^2	0.08	0.10	0.11	0.13

Note: *** *p < 0.01* ** *p < 0.5* * *p < 0.1*. Robust standard errors in parentheses. *Aggressiveness* is a continuous variable measuring a customer's attitude toward agents. For *aggressiveness*, we include only conversations with customers' further engagement in the sample.

We are interested in the coefficient of the interaction term between *signature* and *expressive.* The coefficient estimates are negative and significant in all four estimations, implying that the treatment effect of humanization on *aggressiveness* indeed depends on the complaint type. Compared with customers who are primarily goal-oriented or have mixed motivations, those who are purely emotion-focused are more likely to act less aggressively toward more humanized agents. The estimated treatment effect for the expressive type supports our theoretical arguments that verbal aggression of emotionally charged customers can be alleviated by an enhanced level of humanization. Take Column 1, for example. The aggressiveness for the expressive type reduces by 0.0109 compared to the non-expressive type and the average treatment effect for the expressive type is -0.0151 and statistically significant ($p < 0.05$).

8. Conclusion

Despite the growing popularity and many advantages of text-based customer service, customers may develop the perception that they are communicating with a dehumanized subject (Mesch & Beker, 2010; Oh et al., 2018) which may affect the effectiveness of service provisions. We hypothesized and validated through a quasi-experiment on Twitter, that with the inclusion of identity cues, customers are more willing to engage, and upon engagement, more likely to reach a resolution. Furthermore, we found no evidence of an overall increase in customer verbal aggression towards more humanized agents, despite the theoretical prediction that this could happen. These findings remained consistent in a series of robustness checks and falsification tests. We also conducted a randomized experiment, the findings of which suggested that the observed positive effect of identity cues was driven by customers who were ex ante uncertain or suspicious about the human identity of agents. Moreover, we found supporting evidence for the mechanisms underlying Hypothesis 1 via the randomized experiment and the mechanism underlying Hypothesis 2a and 2b via the quasi-experiment on Twitter.

8.1. Contribution to the Literature

Our study contributes to the IS and marketing literature by investigating whether a customer service agent's identity cue affects customer service interaction; this is an important and timely question given the rise of social media customer service and the current age of advanced algorithms. Unlike traditional customer service centers, a lack of social presence while delivering customer service via text may confuse customers about the human identity of agents, as evidenced by Twitter's efforts to help brands convince customers that they are talking to human agents rather than bots (Perez, 2017), and by the fact that around 38% of the participants in our randomized experiment questioned the human identity of social media customer service agents. Therefore, it is both natural and urgent to understand whether the absence or presence of identity cues has implications for the effectiveness of customer service delivery. To the best of our knowledge, the literature concerning social presence and customer service has only focused on virtual customer service agents whose robotic identity is known to customers. Therefore, our study significantly extends the current literature and connects social presence theory with the burgeoning practice of delivering customer service on social media.

In addition, our study contributes theoretically to the literature on social presence from three perspectives. First, through a randomized experiment, we empirically validated the moderating effect of customers' prior beliefs by demonstrating that customers who are ex ante doubtful about agents' human identities are more susceptible to the effect of identity cues, suggesting that the power of identity cues works by updating customers' prior beliefs about the human identities of agents. This insight highlights the importance of accounting for individual belief heterogeneity in social presence theory. Second, the current study deepened our theoretical understanding of social presence by revealing two fundamental mechanisms driving its effect, at least in the context of customer service. Specifically, we demonstrated that introducing identity cues increased customers' willingness to engage and overall satisfaction by enhancing the level of trust and empathy they perceived in customer service agents. Third, our theorization of the effect of identity cues on customer attitude (i.e., aggressiveness) is new to the literature. The empirical evidence on the moderating effect of whether a complaining customer is goal-oriented or emotion-focused supported the validity of our theorization.

8.2. Contribution to Practice

Our paper also provides several important insights for practitioners. First, the specific empirical setting of the present study suggests that a simple policy change of including agent signatures in service responses can go a long way toward creating more engaging and satisfying interactions. Meanwhile, such a policy does not appear to have the unintended consequence of increasing customer verbal aggression. Given the ease of implementing such a policy, we encourage all firms to consider adopting this strategy when they deliver customer service through text. It should be noted that such a cost-effective strategy is not obvious without a rigorous empirical study. Indeed, few firms have implemented such a strategy on Twitter to date. As of May 2022, four years after Southwest Airlines' signature policy, only 12 of the 35 airlines[5] we checked had adopted the signature policy.

Second, our study highlights the importance of customers' perceptions and their inherent preferences for human agents in text-based customer service. The randomized experiment suggests that a significant portion of customers nowadays suspect agents' algorithm-enabled identities, which might have resulted from the recent trend of automating customer service (e.g., chatbots). The observed effect of identity cues works by updating customers' perceived humanization levels, which demonstrates the value of humanization or, equivalently, bias against machines. Since customers tend to have an inherent need to engage with a human agent when reaching out for help, our findings offer a cautionary tale that firms should resist the temptation to remove humans from the provision of customer service, especially in the current age of AI. Instead of replacing humans with algorithms for the delivery of customer service, we recommend a human-AI collaboration strategy where AI completes peripheral tasks and supports human experts in their handling of the main tasks (Berente et al., 2021; Jain et al., 2021; Rai et al., 2019). For example, human customer service agents can focus on incorporating their human touch and algorithm-enabled analytics into algorithm-drafted responses to complaining customers, thereby striking the right balance between efficiency and effectiveness.

Finally, the results from our mechanism tests suggest that as long as an identity cue can enhance the perceived level of trust and empathy, especially for customers who are uncertain about the human identity of customer service agents, such a cue can lead to more effective customer service. Therefore, firms can go beyond the inclusion of a first name by, for example, including a portrait of an agent in their response. Such a visual identity cue could provide an even stronger signal of humanization and potentially drive even better customer service interactions.

8.3. Limitations and Future Research

Our work has several limitations that are worth noting and offer opportunities for future research. First, the observational study was based on data from the U.S. airline industry. Future works could examine the external validity of our findings using data from other industries or other countries. Second, one of the outcome measures, *resolution*, was based on a third-party evaluation of customer service interactions. Such a measure has value in capturing a bystander's perception of service provisions; however, a more direct measure would be customer-based satisfaction scores. Although the randomized experiment alleviates this concern in our study, future studies may consider extending our research by obtaining customer satisfaction scores in a field study. Third, data limitations and identification challenges preclude us from investigating the long-term effects of identity cues on the effectiveness of customer service delivery on social media. It would be interesting to explore such a question when data is available. Finally, future works could consider alternative formats of identity cues and their heterogeneous impacts.

---

[5] Within the following list of major international airlines, airlines with the signature policy are highlighted in bold: AirAsia, airBaltic, Air France, **Alaska Airlines**, Alitalia, American Air, **British Airways**, Cathay Pacific, Delta, **Deutsche Lufthansa**, **easyJet**, Emirates, **Frontier Airlines**, **Garuda Indonesia**, Gulf Air, Iceland Air, JetBlue, **Jetstar Airways**, Kenya Airways, KLM, Korean Air, **Lufthansa USA**, Malaysia Airlines, Philippine Airlines, Qatar Airways, **SAS - Scandinavian Airlines**, South African Airways, **Southwest Airlines**, Spirit Airlines, **Swiss Intl Air Lines**, United Airlines, **Virgin Atlantic**, Virgin Australia, Vueling Airlines, and WestJet.

9. Acknowledgements

The authors sincerely thank the senior editor, the associate editor, and three anonymous reviewers for their many insightful and constructive suggestions, which significantly contributed to this research. The authors also appreciate the feedback from attendees of 2020 INFORMS Annual Meeting, the 2020 Conference on Information Systems and Technology, and the 2021 Hawaii International Conference on Systems Sciences.

Funding. This research was supported by the Singapore Ministry of Education (MOE) Academic Research Fund (AcRF) Tier 1 grant.

10. References

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, *105*(490), 493–505. http://doi.org/10.1198/jasa.2009.ap08746

Abadie, A., Diamond, A., & Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, *59*(2), 495–510. http://doi.org/10.1111/ajps.12116

Abadie, A., & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque Country. *American Economic Review*, *93*(1), 113–132. http://doi.org/10.1257/000282803321455188

Al-Natour, S., Benbasat, I., & Cenfetelli, R. (2011). The adoption of online shopping assistants: Perceived similarity as an antecedent to evaluative beliefs. *Journal of the Association for Information Systems*, *12*(5), 347–374. http://doi.org/10.17705/1jais.00267

Aral, S., Muchnik, L., & Sundararajan, A. (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences*, *106*(51), 21544–21549. http://doi.org/10.1073/pnas.0908800106

Athey, S., Bayati, M., Doudchenko, N., Imbens, G., & Khosravi, K. (2021). Matrix completion methods for causal panel data models. *Journal of the American Statistical Association*, *116*(536), 1716–1730. http://doi.org/10.1080/01621459.2021.1891924

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), 1228–1234. http://doi.org/10.1080/03610910902859574

Barger, P. B., & Grandey, A. A. (2006). Service with a smile and encounter satisfaction: Emotional contagion and appraisal mechanisms. *Academy of Management Journal*, *49*(6), 1229–1238. http://doi.org/10.5465/amj.2006.23478695

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly*, *45*(3), 1433–1450. https://doi.org/10.25300/MISQ/2021/16274

Biocca, F. (1997). The cyborg's dilemma: Progressive embodiment in virtual environments. *Journal of Computer-Mediated Communication*, *3*(2). http://doi.org/10.1111/j.1083-6101.1997.tb00070.x

Blut, M., Wang, C., Wünderlich, N. V., & Brock, C. (2021). Understanding anthropomorphism in service provision: A meta-analysis of physical robots, chatbots, and other AI. *Journal of the Academy of Marketing Science*, *49*(4), 632–658. http://doi.org/10.1007/s11747-020-00762-y

Broockman, D. E. (2013). Black politicians are more intrinsically motivated to advance blacks' interests: A field experiment manipulating political incentives. *American Journal of Political Science*, *57*(3), 521–536. http://doi.org/10.1111/ajps.12018

Carroll, R. J., & Stefanski, L. A. (1990). Approximate quasi-likelihood estimation in models with surrogate predictors. *Journal of the American Statistical Association*, *85*(411), 652–663. http://doi.org/10.1080/01621459.1990.10474925

Chen, W., Gu, B., Ye, Q., & Zhu, K. X. (2019). Measuring and managing the externality of managerial responses to online customer reviews. *Information Systems Research*, *30*(1), 81–96. http://doi.org/10.1287/isre.2018.0781

Cheng, H.-T., & Pan, Y. (2021). I'm not a chatbot: An empirical investigation of humanized profiles of social media customer service representatives. *Proceedings of the 54th Hawaii International Conference on System Sciences*, pp. 4167–4176. http://doi.org/10.24251/HICSS.2021.506

Chevalier, J. A., Dover, Y., & Mayzlin, D. (2018). Channels of impact: User reviews when quality is dynamic and managers respond. *Marketing Science*, *37*(5), 688–709. http://doi.org/10.1287/mksc.2018.1090

Claussen, J., Kretschmer, T., & Mayrhofer, P. (2013). The effects of rewarding user engagement: The case of Facebook apps. *Information Systems Research*, *24*(1), 186–200. http://doi.org/10.1287/isre.1120.0467

Cobb, S. C. (2009). Social presence and online learning: A current view from a research perspective. *Journal of Interactive Online Learning*, *8*(3), 241–254.

Coine, T., & Babbitt, M. (2014). *A world gone social: How companies must adapt to survive*. Amacom.

Cook, J. R., & Stefanski, L. A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association*, *89*(428), 1314–1328. http://doi.org/10.1080/01621459.1994.10476871

Coulter, K. S., & Coulter, R. A. (2002). Determinants of trust in a service provider: The moderating role of length of relationship. *Journal of Services Marketing*, *16*(1), 35–50. http://doi.org/10.1108/08876040210419406

Cyr, D., Hassanein, K., Head, M., & Ivanov, A. (2007). The role of social presence in establishing loyalty in e-service environments. *Interacting with Computers*, *19*(1), 43–56. http://doi.org/10.1016/j.intcom.2006.07.010

Derks, D., Fischer, A. H., & Bos, A. E. (2008). The role of emotion in computer-mediated communication: A review. *Computers in Human Behavior*, *24*(3), 766–785. http://doi.org/10.1016/j.chb.2007.04.004

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, *42*(3), 177–190. http://doi.org/10.1016/S0921-8890(02)00374-3

Einwiller, S. (2003). When reputation engenders trust: An empirical investigation in business-to-consumer electronic commerce. *Electronic Markets*, *13*(3), 196–209.

Fan, A., Wu, L. (Laurie), & Mattila, A. S. (2016). Does anthropomorphism influence customers' switching intentions in the self-service technology failure context? *Journal of Services Marketing*, *30*(7), 713–723. http://doi.org/10.1108/JSM-07-2015-0225

Fang, Y., Qureshi, I., Sun, H., McCole, P., Ramsey, E., & Lim, K. H. (2014). Trust, satisfaction, and online repurchase intention. *MIS Quarterly*, *38*(2), 407–428. https://doi.org/10.25300/MISQ/2014/38.2.04

Feng, B., Li, S., & Li, N. (2016). Is a profile worth a thousand words? How online support-seeker's profile features may influence the quality of received support messages. *Communication Research*, *43*(2), 253–276. http://doi.org/10.1177/0093650213510942

Gans, J. S., Goldfarb, A., & Lederman, M. (2021). Exit, tweets, and loyalty. *American Economic Journal: Microeconomics*, *13*(2), 68–112. http://doi.org/10.1257/mic.20180085

Garbarino, E., & Lee, O. F. (2003). Dynamic pricing in internet retail: Effects on consumer trust. *Psychology & Marketing*, *20*(6), 495–513. http://doi.org/10.1002/mar.10084

Gefen, D., & Straub, D. (2003). Managing user trust in B2C e-services. *E-Service*, *2*(2), 7–24. http://doi.org/10.2979/esj.2003.2.2.7

Gnewuch, U., Morana, S., & Maedche, A. (2017). Towards designing cooperative and social conversational agents for customer service. *Proceedings of the 38th International Conference on Information Systems (ICIS)*, Seoul, South Korea.

Goes, P. B., Lin, M., & Au Yeung, C. (2014). "Popularity effect" in user-generated content: Evidence from online product reviews. *Information Systems Research*, *25*(2), 222–238. http://doi.org/10.1287/isre.2013.0512

Grandey, A. A., Dickter, D. N., & Sin, H.-P. (2004). The customer is not always right: Customer aggression and emotion regulation of service employees. *Journal of Organizational Behavior*, *25*(3), 397–418. http://doi.org/10.1002/job.252

Gu, B., & Ye, Q. (2014). First step in social media: Measuring the influence of online management responses on customer satisfaction. *Production and Operations Management*, *23*(4), 570–582. http://doi.org/10.1111/poms.12043

Gunarathne, P., Rui, H., & Seidmann, A. (2017). Whose and what social media complaints have happier resolutions? Evidence from Twitter. *Journal of Management Information Systems*, *34*(2), 314–340. http://doi.org/10.1080/07421222.2017.1334465

Gunarathne, P., Rui, H., & Seidmann, A. (2018). When social media delivers customer service: Differential customer treatment in the airline industry. *MIS Quarterly*, *42*(2), 489–520. http://doi.org/10.25300/MISQ/2018/14290

Gunarathne, P., Rui, H., & Seidmann, A. (2022). Racial bias in customer service: Evidence from Twitter. *Information Systems Research*, *33*(1), 43–54. http://doi.org/10.1287/isre.2021.1058

Gunawardena, C. N. (1995). Social presence theory and implications for interaction and collaborative learning in computer conferences. *International Journal of Educational Telecommunications*, *1*(2), 147–166.

Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis*, *20*(1), 25–46. http://doi.org/10.1093/pan/mpr025

Hassanein, K. S., & Head, M. M. (2004). Building online trust through socially rich web interfaces. *Proceedings of the 2nd Annual Conference on Privacy, Security and Trust, Fredericton, New Brunswick, Canada*, pp. 15–22.

He, S., Lee, S.-Y., & Rui, H. (2019). Open voice or private message? The hidden tug-of-war on social media customer service. *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 6638–6647. http://doi.org/10.24251/HICSS.2019.795

Hicks, R., & Tingley, D. (2011). Causal mediation analysis. *The Stata Journal*, *11*(4), 605–619. http://doi.org/10.1177/1536867X1201100407

Hovland, C. I., & Weiss, W. (1951). The influence of source credibility on communication effectiveness. *Public Opinion Quarterly*, *15*(4), 635–650. http://doi.org/10.1086/266350

Hu, Y., Tafti, A., & Gal, D. (2019). Read this, please? The role of politeness in customer service engagement on social media. *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 4466–4475. http://doi.org/10.24251/HICSS.2019.541

Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, *20*(1), 1–24. http://doi.org/10.1093/pan/mpr013

Imai, K., Keele, L., & Yamamoto, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science*, *25*(1), 51–71. http://doi.org/10.1214/10-STS321

Jain, H., Padmanabhan, B., Pavlou, P. A., & Raghu, T. S. (2021). Editorial for the special section on humans, algorithms, and augmented intelligence: The future of work, organizations, and society. *Information Systems Research*, *32*(3), 675–687. http://doi.org/10.1287/isre.2021.1046

Jarvenpaa, S. L., Tractinsky, N., & Vitale, M. (2000). Consumer trust in an Internet store. *Information Technology and Management*, *1*(1), 45–71.

Kehrwald, B. (2008). Understanding social presence in text-based online learning environments. *Distance Education*, *29*(1), 89–106. http://doi.org/10.1080/01587910802004860

Kowalski, R. M. (1996). Complaints and complaining: Functions, antecedents, and consequences. *Psychological Bulletin*, *119*(2), 179. http://doi.org/10.1037/0033-2909.119.2.179

Kreijns, K., Kirschner, P. A., Jochems, W., & Van Buuren, H. (2004). Determining sociability, social space, and social presence in (a) synchronous collaborative groups. *CyberPsychology & Behavior*, *7*(2), 155–172. http://doi.org/10.1089/109493104323024429

Kumar, N., Qiu, L., & Kumar, S. (2018). Exit, voice, and response on digital platforms: An empirical investigation of online management response strategies. *Information Systems Research*, *29*(4), 849–870. http://doi.org/10.1287/isre.2017.0749

Lankton, N. K., McKnight, D. H., & Tripp, J. (2015). Technology, humanness, and trust: Rethinking trust in technology. *Journal of the Association for Information Systems*, *16*(10), 880–918. http://doi.org/10.17705/1jais.00411

Lee, Y. J., Xie, K., & Besharat, A. (2016). Management response to online WOM: Helpful or detrimental? *Proceedings of 22nd Americas Conference on Information Systems*, San Diego, U.S.A.. http://doi.org/10.2139/ssrn.2919553

Li, X., Grahl, J., & Hinz, O. (2022). How do recommender systems lead to consumer purchases? A causal mediation analysis of a field experiment. *Information Systems Research*, *33*(2), 620–637. http://doi.org/10.1287/isre.2021.1074

Ma, L., Sun, B., & Kekre, S. (2015). The squeaky wheel gets the grease—An empirical analysis of customer voice and firm intervention on Twitter. *Marketing Science*, *34*(5), 627–645. http://doi.org/10.1287/mksc.2015.0912

Mesch, G. S., & Beker, G. (2010). Are norms of disclosure of online and offline personal information associated with the disclosure of personal information online? *Human Communication Research*, *36*(4), 570–592. http://doi.org/10.1111/j.1468-2958.2010.01389.x

Meyer, B. D. (1995). Natural and quasi-experiments in economics. *Journal of Business & Economic Statistics*, *13*(2), 151–161. http://doi.org/10.1080/07350015.1995.10524589

Mortenson, S. T. (2009). Interpersonal trust and social skill in seeking social support among Chinese and Americans. *Communication Research*, *36*(1), 32–53. http://doi.org/10.1177/0093650208326460

Mousavi, R., Johar, M., & Mookerjee, V. S. (2020). The voice of the customer: Managing customer care in Twitter. *Information Systems Research*, *31*(2), 340–360. http://doi.org/10.1287/isre.2019.0889

Oh, C. S., Bailenson, J. N., & Welch, G. F. (2018). A systematic review of social presence: Definition, antecedents, and implications. *Frontiers in Robotics and AI*, *5*(114), 1–35. http://doi.org/10.3389/frobt.2018.00114

Pereira, A., Prada, R., & Paiva, A. (2014). Improving social presence in human-agent interaction. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1449–1458. http://doi.org/10.1145/2556288.2557180

Perez, S. (2017, February 23). Twitter debuts custom profiles for businesses' agents, so people don't think they're talking to bots. *TechCrunch*. https://social.techcrunch.com/2017/02/22/twitter-debuts-custom-profiles-for-businesses-so-people-dont-think-theyre-talking-to-bots/

Proserpio, D., Troncoso, I., & Valsesia, F. (2021). Does gender matter? The effect of management responses on reviewing behavior. *Marketing Science*, *40*(6), 1199–1213. http://doi.org/10.1287/mksc.2021.1313

Proserpio, D., & Zervas, G. (2017). Online reputation management: Estimating the impact of management responses on consumer reviews. *Marketing Science*, *36*(5), 645–665. http://doi.org/10.1287/mksc.2017.1043

Qiu, L., & Benbasat, I. (2009). Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems. *Journal of Management Information Systems*, *25*(4), 145–182. http://doi.org/10.2753/MIS0742-1222250405

Rafaeli, A., Erez, A., Ravid, S., Derfler-Rozin, R., Treister, D. E., & Scheyer, R. (2012). When customers exhibit verbal aggression, employees pay cognitive costs. *Journal of Applied Psychology*, *97*(5), 931–950. http://doi.org/10.1037/a0028559

Rai, A., Constantinides, P., & Sarker, S. (2019). Next generation digital platforms: Toward human-AI hybrids. *MIS Quarterly*, *43*(1), iii–ix. https ://misq.org/misq/downloads/

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65. http://doi.org/10.1016/0377-0427(87)90125-7

Salem, M., Eyssel, F., Rohlfing, K., Kopp, S., & Joublin, F. (2013). To err is human (-like): Effects of robot gesture on perceived anthropomorphism and likability. *International Journal of Social Robotics*, *5*(3), 313–323. http://doi.org/10.1007/s12369-013-0196-9

Schanke, S., Burtch, G., & Ray, G. (2021). Estimating the impact of "humanizing" customer service chatbots. *Information Systems Research*, *32*(3), 736–751. http://doi.org/10.1287/isre.2021.1015

Schuetzler, R. M., Giboney, J. S., Grimes, G. M., & Nunamaker Jr, J. F. (2018). The influence of conversational agent embodiment and conversational relevance on socially desirable responding. *Decision Support Systems*, *114*, 94–102. http://doi.org/10.1016/j.dss.2018.08.011

Schuetzler, R. M., Grimes, G. M., & Scott Giboney, J. (2020). The impact of chatbot conversational skill on engagement and perceived humanness. *Journal of Management Information Systems*, *37*(3), 875–900. http://doi.org/10.1080/07421222.2020.1790204

Sheehan, B., Jin, H. S., & Gottlieb, U. (2020). Customer service chatbots: Anthropomorphism and adoption. *Journal of Business Research*, *115*, 14–24. http://doi.org/10.1016/j.jbusres.2020.04.030

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. Toronto; London; New York: Wiley.
Southwest Airlines. (n.d.). *Southwest careers discussion forum*. Retrieved August 1, 2022, from https://community.southwest.com/t5/Southwest-Careers/bd-p/Careers
Sternthal, B., Dholakia, R., & Leavitt, C. (1978). The persuasive effect of source credibility: Tests of cognitive response. *Journal of Consumer Research*, *4*(4), 252–260. http://doi.org/10.1086/208704
Sun, S., Gao, Y., & Rui, H. (2021). Does active service intervention drive more complaints on social media? The roles of service quality and awareness. *Journal of Management Information Systems*, *38*(3), 579–611. http://doi.org/10.1080/07421222.2021.1958548
Tax, S. S., Brown, S. W., & Chandrashekaran, M. (1998). Customer evaluations of service complaint experiences: Implications for relationship marketing. *Journal of Marketing*, *62*(2), 60–76. http://doi.org/10.1177/002224299806200205
UJET. (n.d.). *Optimizing channels for customer support whitepaper*. UJET. Retrieved August 1, 2022, from https://ujet.cx/resources/white-papers/optimizing-channels-for-customer-support/
van Doorn, J., Mende, M., Noble, S. M., Hulland, J., Ostrom, A. L., Grewal, D., & Petersen, J. A. (2017). Domo arigato mr. Roboto: Emergence of automated social presence in organizational frontlines and customers' service experiences. *Journal of Service Research*, *20*(1), 43–58. http://doi.org/10.1177/1094670516679272
Verhagen, T., Van Nes, J., Feldberg, F., & Van Dolen, W. (2014). Virtual customer service agents: Using social presence and personalization to shape online service encounters. *Journal of Computer-Mediated Communication*, *19*(3), 529–545. http://doi.org/10.1111/jcc4.12066
Walker, D. D., van Jaarsveld, D. D., & Skarlicki, D. P. (2017). Sticks and stones can break my bones but words can also hurt me: The relationship between customer verbal aggression and employee incivility. *Journal of Applied Psychology*, *102*(2), 163–179. http://doi.org/10.1037/apl0000170
Wang, Y., & Chaudhry, A. (2018). When and how managers' responses to online reviews affect subsequent reviews. *Journal of Marketing Research*, *55*(2), 163–177. http://doi.org/10.1509/jmr.15.0511
Xie, K. L., Zhang, Z., & Zhang, Z. (2014). The business value of online consumer reviews and management response to hotel performance. *International Journal of Hospitality Management*, *43*, 1–12. http://doi.org/10.1016/j.ijhm.2014.07.007
Xu, Y. (2017). Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis*, *25*(1), 57–76. http://doi.org/10.1017/pan.2016.2
Yam, K. C., Bigman, Y. E., Tang, P. M., Ilies, R., De Cremer, D., Soh, H., & Gray, K. (2021). Robots at work: People prefer-and forgive-service robots with perceived feelings. *Journal of Applied Psychology*, *106*(10), 1557–1572. http://doi.org/10.1037/apl0000834
Yamaguchi, H., Suzuki, K., & Ramamoorthy, C. V. (2003). The humanization, personalization and authentication issues in the design of interactive service system. *Journal of Integrated Design and Process Science*, *7*(4), 1–11.
Yang, M., Zheng, Z., & Mookerjee, V. (2019). Prescribing response strategies to manage customer opinions: A stochastic differential equation approach. *Information Systems Research*, *30*(2), 351–374. http://doi.org/10.1287/isre.2018.0805

Yarkoni, T. (2010). Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality*, *44*(3), 363–373. https://doi.org/10.1016/j.jrp.2010.04.001

Yeomans, M., Kantor, A., & Tingley, D. (2018). The politeness package: Detecting politeness in natural language. *The R Journal*, *10*(2), 489–502. http://doi.org/10.32614/RJ-2018-079

Yoo, Y., & Alavi, M. (2001). Media and group cohesion: Relative influences on social presence, task participation, and group consensus. *MIS Quarterly*, *25*(3), 371–390. http://doi.org/10.2307/3250922

Zhang, T., Kaber, D. B., Zhu, B., Swangnetr, M., Mosaly, P., & Hodge, L. (2010). Service robot feature design effects on user perceptions and emotional responses. *Intelligent Service Robotics*, *3*(2), 73–88. http://doi.org/10.1007/s11370-010-0060-9

Zhao, X., Lynch Jr, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research*, *37*(2), 197–206. http://doi.org/10.1086/651257

Zhao, Y., Wen, L., Feng, X., Li, R., & Lin, X. (2020). How managerial responses to online reviews affect customer satisfaction: An empirical study based on additional reviews. *Journal of Retailing and Consumer Services*, *57*, Article 102205. http://doi.org/10.1016/j.jretconser.2020.102205

11. Author Biographies

Yang Gao is an Assistant Professor of Information Systems at the School of Computing and Information Systems, Singapore Management University. He received his Ph.D. from the University of Rochester in 2021. His current research focuses on the management of customer voices on social media. His works have been accepted for publications in journals, including *Information Systems Research* and *Journal of Management Information Systems*.

Huaxia Rui is the Xerox Chair Professor at Simon Business School and is an inaugural Distinguished Researcher and Scholar of the Center of Excellence in Data Science, at the University of Rochester. He is interested in data science and causal inference. His works have been published in leading academic journals such as *Information Systems Research*, *MIS Quarterly*, *Journal of Management Information Systems*, *Management Science*, *Production and Operation Management*, and *Journal of Financial Economics*.

Shujing Sun is an Assistant Professor at the Naveen Jindal School of Management, the University of Texas at Dallas. She received her Ph.D. from the University of Rochester in 2020. Her research interests center around the economics of technological innovations, with a current focus on healthcare information technology and social media analytics. She has published in *Information Systems Research* and *Journal of Management Information Systems*.

12. Appendices

12.1. Appendix A: Criteria for Labeling the Resolution of Conversations

The annotator is asked to read through a conversation carefully and then determine whether there is a resolution of the service interaction based on the following criteria.

- If a customer expressed satisfaction or gratitude in the end, the outcome of the conversation should be labeled as a resolution. Take the first conversation (ID 1-1 ~ 1-5) listed in Table A1, as an example. Since the customer explicitly appreciated the agent for updating the flight information, the outcome of the conversation is a resolution. On the contrary, the customer in the second example conversation (ID 2-1 ~ 2-6) still thought the company was incompetent and unaccountable after multiple interactions with the agent, which indicates a service failure.
- If a customer sent private messages to the airline, the outcome of the conversation would be labeled as a resolution. The rationale is that private communications can reduce the risk of negative externality when firms fail to address complaints publicly (He et al., 2019). Therefore, a focal customer's agreement to communicate via private messages at least signals a "public" resolution to bystanders. The third sample conversation (ID 3-1 ~ 3-3) listed in Table A1 is an example.

- If a customer didn't explicitly express gratitude but the agent successfully addressed the question or complaint, then the outcome of the conversation would be labeled as a resolution. Take the fourth conversation (ID 4-1 ~ 4-6) listed in Table A1, for instance. The agent provided detailed information to answer the customer's question of why the flight was canceled.

For a small set of conversations that do not fall into any of the above circumstances, the annotator should follow two steps: 1) make his/her judgment based on factors, such as customers' sentiment changes and agents' efforts and reasoning; 2) discuss with the authors and determine the label according to all readers' votes.

Table A1. Sample Conversations for Labeling Criteria

ID	Role	Content	Resolution
1-1	Customer	@SouthwestAir what's going on with Flight 40 out of Dallas to Chicago?	✓
1-2	Agent	@customer Hey there, Courtney. It looks like inclement weather and associated Air Traffic Control delays have held your aircraft up, but we are working to get your flight to Chicago another aircraft ASAP. We hope to have you on your way soon. -Adrienne	
1-3	Customer	@SouthwestAir On another aircraft? So there's something wrong with the plane as well? I'm seeing 7:20 departure time	
1-4	Agent	@customer Nope! Nothing wrong, just that it has been held up with delays, and we are hoping to get you an alternative plane to get you out sooner. We are currently anticipating an ETD of about 7:20pm. -Adrienne	
1-5	Customer	@SouthwestAir Thanks for the update	
2-1	Customer	@SouthwestAir you leave me stranded on St. Patricks day at an airport with nothing open. No offer of anything except the payment of my time. Sweet. Fun. Awesome	×
2-2	Agent	@customer We regret the disappointment today, Brent. We know that your time is precious, and we appreciate your patience today! -Nicole	
2-3	Customer	@SouthwestAir Yeah that's not it. Ill use a different airline for now on. Thanks for your patience	
2-4	Customer	@SouthwestAir Are you going to cover my costs for the extra $ it will cost me for this delay?	
2-5	Agent	@customer While we don't cover interim expenses, you're welcome to reach out to us once your flight is completed so we can assist at that time. -Nicole	
2-6	Customer	@SouthwestAir Airlines. The only industry that can waste your time, cost you $ and there is no accountability	
3-1	Customer	@SouthwestAir I was charged for wifi on my phone, but I still can't get a connection for wifi on my phone. What should I do?	✓
3-2	Agent	@customer Hey, Mister. Our apologies for the inconvenience. Please DM the email address you used when you purchased the WiFi, and we'll follow-up. ^KD	
3-3	Customer	@SouthwestAir I've sent the message	
4-1	Customer	@SouthwestAir what gives? Cancelled flight in line at the airport and on infinite hold with customer service. #nothappy	✓
4-2	Agent	@customer It's not our intention to disappoint you, Kim. Please speak to an Agent to get rebooked. We can refund the unused portion of your flight if you are able to find alternate transportation. Please DM if you need further assistance. ^SJ https://t.co/mQmfkXW4oV	
4-3	Customer	@SouthwestAir Wondering why the delta flight at the same time is on time?	
4-4	Customer	@customer Mind sending your flight details so we can look into this? ^SJ	
4-5	Agent	@SouthwestAir 224 from ATL to MCI	
4-6	Customer	@customer Thank for the additional information. Our records indicate Flight #224 is canceled due to weather. We cannot speak to other carriers. Our number one concern is operating a safe flight for our Customers and Employees. I am sorry for frustration this has caused. ^SJ	

12.2. Appendix B: Text Classifiers

Table A2. Performance of the SVM Classifier on Customer Service-Related Conversations Using 10-Fold Cross Validation

	Precision (0)	Recall (0)	F1(0)	Precision (1)	Recall (1)	F1 (1)
Fold 1	0.90	0.85	0.87	0.91	0.94	0.92
Fold 2	0.89	0.82	0.85	0.89	0.93	0.91
Fold 3	0.91	0.83	0.89	0.89	0.95	0.92
Fold 4	0.88	0.84	0.86	0.90	0.93	0.91

Fold 5	0.88	0.84	0.86	0.90	0.93	0.92
Fold 6	0.89	0.83	0.86	0.90	0.93	0.92
Fold 7	0.89	0.84	0.86	0.90	0.94	0.92
Fold 8	0.90	0.82	0.86	0.90	0.94	0.92
Fold 9	0.90	0.85	0.87	0.91	0.94	0.92
Fold 10	0.90	0.84	0.87	0.91	0.95	0.93
Average	**0.89**	**0.84**	**0.87**	**0.90**	**0.94**	**0.92**
Note: The (0) symbol represents conversations irrelevant to customer service, and the (1) symbol represents conversations related to customer service.						

Table A3. Performance of the SVM Classifier on Conversation Resolution Using 10-Fold Cross Validation

	Precision (0)	Recall (0)	F1(0)	Precision (1)	Recall (1)	F1 (1)
Fold 1	0.74	0.73	0.73	0.82	0.82	0.82
Fold 2	0.74	0.72	0.73	0.80	0.81	0.80
Fold 3	0.73	0.70	0.72	0.80	0.82	0.81
Fold 4	0.77	0.72	0.74	0.81	0.85	0.83
Fold 5	0.76	0.73	0.75	0.81	0.84	0.82
Fold 6	0.77	0.73	0.75	0.80	0.83	0.81
Fold 7	0.73	0.70	0.72	0.78	0.81	0.79
Fold 8	0.73	0.76	0.74	0.82	0.81	0.81
Fold 9	0.74	0.72	0.73	0.80	0.82	0.81
Fold 10	0.74	0.69	0.72	0.78	0.82	0.80
Average	**0.75**	**0.72**	**0.73**	**0.80**	**0.82**	**0.81**
Note: The (0) symbol represents conversations ended without a resolution, and the (1) symbol represents conversations ended with a resolution.						

12.3. Appendix C: Text Clustering

We first preprocess the tweets by removing numbers, punctuation, and stop words. Then we create a term-document matrix with the term frequency-inverse document frequency (TF-IDF) as the weighting scheme. Considering the high dimensionality of text data, we use the singular-value decomposition (SVD) technique to reduce the dimensionality of the term-document matrix. Next, we implement the K-means algorithm on the term-document matrix to group similar tweets into clusters. To determine the appropriate number of clusters, we run the algorithm with different numbers of clusters specified and then calculate the average silhouette scores (Rousseeuw, 1987). As shown in Figure A1, the best silhouette score is obtained for seven clusters. To further evaluate the performance of the clustering algorithm, we create a word cloud for each cluster and manually examine if a topic can be inferred from the cloud. We can see a separation in the types of service requests from the world clouds: flight booking, delay, or cancellation; baggage; inflight amenities; and other specific requests. The detailed word clouds are available upon request. The separation is generally aligned with the top customer service requests according to Southwest (see https://www.southwest.com/html/customer-service/index.html?clk=GFOOTER-SERVICE). Therefore, the identified clusters are informative, and we introduce them into our regression model as valid controls.

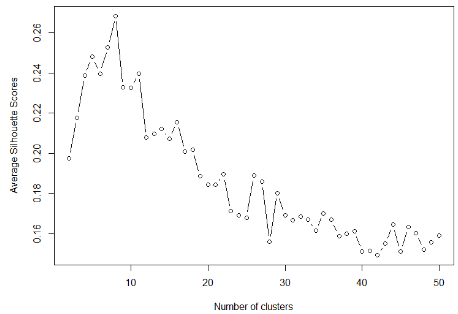
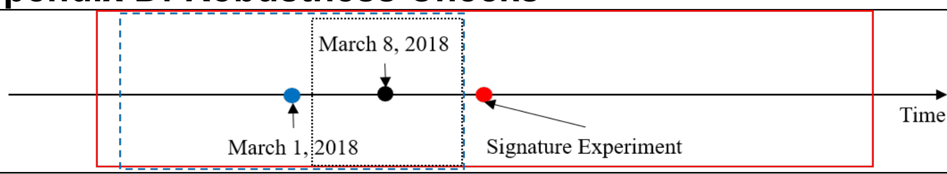


Figure A1. Average Silhouette Scores

12.4. Appendix D: Robustness Checks



Note. The solid box represents the time window used for the baseline analysis (i.e., one month before and after the signature experiment), the two dashed and dotted boxes represent the time windows used for the falsification tests in the pre-treatment period.

Figure A2. Falsification Test with Pseudo-treatments

Table A4. Falsification Test with Pseudo Treatment at Different Times Before the Signature Experiment

	March 1, 2018			March 8, 2018		
	engagement	*resolution*	*aggressiveness*	*engagement*	*resolution*	*aggressiveness*
	(1)	(2)	(3)	(4)	(5)	(6)
signature	0.0501 (0.0407)	0.0358 (0.0661)	-0.0026 (0.0126)	0.2056 (0.1975)	-0.1895 (0.3636)	0.0972 (0.0645)
Controls	Y	Y	Y	Y	Y	Y
Time Trend	Y	Y	Y	Y	Y	Y
Seasonality FE	Y	Y	Y	Y	Y	Y
Observations	3513	1315	1315	1649	574	574
R^2	0.09	0.14	0.10	0.12	0.21	0.20

Note: *** $p < 0.01$ ** $p < 0.5$ * $p < 0.1$. Robust standard errors in parentheses. Columns 1 – 3 report results with a pseudo treatment on March 1, 2018, and the estimation window as two weeks before and after the pseudo treatment. Columns 4 – 6 report results with a pseudo treatment on March 8, 2018, and the estimation window as one week before and after the pseudo treatment.

Table A5. Falsification Test with Pseudo Treatment at Southwest Airlines (in 2017)

	Southwest Airlines – March 16, 2017		
	engagement	*resolution*	*aggressiveness*
	(1)	(2)	(3)
signature	-0.0085 (0.0213)	-0.0076 (0.0352)	0.0014 (0.0059)
Controls	Y	Y	Y
Time Trend	Y	Y	Y
Seasonality FE	Y	Y	Y
Observations	8013	2936	2936
R^2	0.08	0.16	0.07

Note: *** $p < 0.01$ ** $p < 0.5$ * $p < 0.1$. Robust standard errors in parentheses. This table reports the results with a pseudo treatment on March 16, 2017 (i.e., one year before the implementation of the signature policy), for Southwest Airlines.

Table A6. Entropy Balancing and Coarsened Exact Matching						
	EB			CEM		
	engagement	*resolution*	*aggressiveness*	*engagement*	*resolution*	*aggressiveness*
	(1)	(2)	(3)	(4)	(5)	(6)
signature	0.0620*** (0.0229)	0.0670* (0.0380)	-0.0086 (0.0057)	0.0791*** (0.0249)	0.0853** (0.0420)	-0.0077 (0.0051)
Controls	Y	Y	Y	Y	Y	Y
Time Trend	Y	Y	Y	Y	Y	Y
Seasonality FE	Y	Y	Y	Y	Y	Y
Observations	8214	3258	3258	7733	3110	3110
R^2	0.01	0.02	0.01	0.02	0.07	0.06

Note: *** $p < 0.01$ ** $p < 0.5$ * $p < 0.1$. Robust standard errors in parentheses. Columns 1 – 3 report regression results with the sample reweighted by Entropy Balancing (EB). No observation is dropped because the sample weights are all above zero. Columns 4 – 6 report regression results after the application of Coarsened Exact Matching (CEM).

Table A7. Synthetic Control Analysis			
	$engagement_{f,t}$	$resolution_{f,t}$	$aggressiveness_{f,t}$
	(1)	(2)	(3)
$signature_{f,t}$	0.018** (0.008)	0.016** (0.007)	-0.0012 (0.0029)
Controls	Y	Y	Y
Firm FE	Y	Y	Y
Time FE	Y	Y	Y
Observations	300	300	300

Note: *** $p < 0.01$ ** $p < 0.5$ * $p < 0.1$. Bootstrapped standard errors in parentheses.

Table A8. Two-way Matching Analysis						
	caliper = 3			caliper = 4		
	engagement	*resolution*	*aggressiveness*	*engagement*	*resolution*	*aggressiveness*
	(1)	(2)	(3)	(4)	(5)	(6)
signature	0.0314** (0.0157)	0.0878*** (0.0331)	-0.0023 (0.0057)	0.0290** (0.0148)	0.0823*** (0.0277)	-0.0019 (0.0049)
Observations	3629	797	797	4138	1130	1130

Note: *** $p < 0.01$ ** $p < 0.5$ * $p < 0.1$. We use calipers to exclude the matched pairs whose distance is larger than the threshold. For instance, if the caliper is specified as three, the matched pairs whose distance is more than three times the standard deviation of all distances among matched pairs will be dropped.

12.5. Appendix E: Evaluation of the Synthetic Control

Table A9. Comparison of Offline Operation and Online Presence Among Airlines in the US

Airline	Passengers	Fleet Size	Destinations	Daily Departures	Followers
Southwest Airlines	162,681,000	736	111	4,000	2.1M
American Airlines	215,182,000	1,440	352	5,400	1.5M
Delta Airlines	204,000,000	1,170	290	6,700	1.5M
United Airlines	162,443,000	1,391	371	5,000	1M
JetBlue Airlines	42,727,694	270	102	925	1.9M
Spirit Airlines	34,537,000	157	78	400	112.7k

Frontier Airlines	21,689,000	104	110	270	123.1k
Hawaiian Airlines	11,751,003	61	32	212	188.7k

Note: The 2nd through the 5th columns report the statistics of airlines' offline operations, which are from Wikipedia (see https://en.wikipedia.org/wiki/List_of_largest_airlines_in_North_America). Given the enormous impact of COVID-19 on the airline industry, we report the passenger volumes as of 2019. The other three statistics as of March 9th, 2021. The numbers of followers on Twitter are as of May 23rd, 2021.

Table A10. Paired t-tests of Differences in the Pre-treatment Period (in %)

	$engagement_{f,t}$	$resolution_{f,t}$	$aggressiveness_{f,t}$
Treated – American Airlines	-0.545	1.278	-1.373***
Treated – Delta Airlines	11.064***	4.615***	0.083
Treated – United Airlines	4.230***	5.516***	-0.090
Treated – JetBlue Airlines	-7.847***	-5.621**	0.356
Treated – Synthetic Control	3.846e-8	6.562e-8	1.682e-8

Note: *** $p < 0.01$ ** $p < 0.05$ * $p < 0.1$.

12.6. Appendix F

Table A11. List of Measurement Items

Theoretical Constructs	Items	Source
Humanization (Social Presence)	• There was a human contact with the agent. • There was a sense of personalness with the agent. • There was a sense of sociability with the agent. • There was a sense of human warmth with the agent. • There was a sense of human sensitivity with the agent.	(Yoo & Alavi, 2001)
Trust	• The agent is consistent in quality and service. • The agent is keen on fulfilling customer needs. • The agent is honest. • The agent wants to be known as one that keeps promises and commitments. • The agent has customers' best interests in mind. • The agent is trustworthy. • The agent has high integrity. • The agent is dependable.	(Einwiller, 2003; Fang et al., 2014; Garbarino & Lee, 2003; Jarvenpaa et al., 2000)
Empathy	• The agent is caring. • The agent is warm. • The agent is friendly.	(Coulter & Coulter, 2002)
Satisfaction	• How satisfied do you think the customer feel with the agent's response? • How satisfied do you think the customer feel with the overall interaction?	(Barger & Grandey, 2006)

Table A12. Experiment Service Encounters by Service Request and Reply Style

	Specific reply	Generic reply
Flight delay	**- Customer:** Flight to Phoenix delayed 3 hours. Should've just drove there. Thanks a lot @Airlines! **- Airline:** We're sorry for the trouble. What is your flight number? We would be happy to get an update for you. -MR **- Customer:** 2033 to PHX and 1568 from PHX to LAS **- Airline:** Our apologies for the delay. It looks like Flight #2033 is delayed due to Air Traffic Control directives, and our Team will do their best to make up time where they can. We appreciate you hanging in there with us today, and we hope to have you on your way soon. -MR	**- Customer:** Flight to Phoenix delayed 3 hours. Should've just drove there. Thanks a lot @Airlines! **- Airline:** We never like to keep you waiting. We appreciate your patience while we work to get you going ASAP. -MR **- Customer:** I need a free flight. Or a discount at least! **- Airline:** While we're unable to offer you a free flight, you're more than welcome to

		follow up with us post travel so we can look further into your flight details. -MR
Lost baggage	**- Customer:** I don't understand how @Airlines lost/delayed my bag. I really wish another airline had a direct flight to my city. So annoyed. **- Airline:** We don't like hearing this. Have you filed a baggage report? We'll do our best to have you reunited with it soon. Can we be of any further assistance in the meantime? -RR **- Customer:** How about being available to talk to customers you've wronged? 45 minute wait to talk to corporate customer service? Again, Alaska airlines is looking better and better. **- Airline:** Did you speak to Customer Relations? We'll be glad to address your concerns here and see what else can be done to try and make it right. You're welcome to DM us your confirmation number. -RR	**- Customer:** I don't understand how @Airlines lost/delayed my bag. I really wish another airline had a direct flight to my city. So annoyed. **- Airline:** We're sorry for the inconvenience. I assure you our Baggage Team will do all they can to locate your bag as soon as possible. -RR **- Customer:** How about being available to talk to customers you've wronged? 45 minute wait to talk to corporate customer service? Again, Alaska airlines is looking better and better. **- Airline:** We hope to reunite you with your baggage soon. Thank you for your patience. -RR

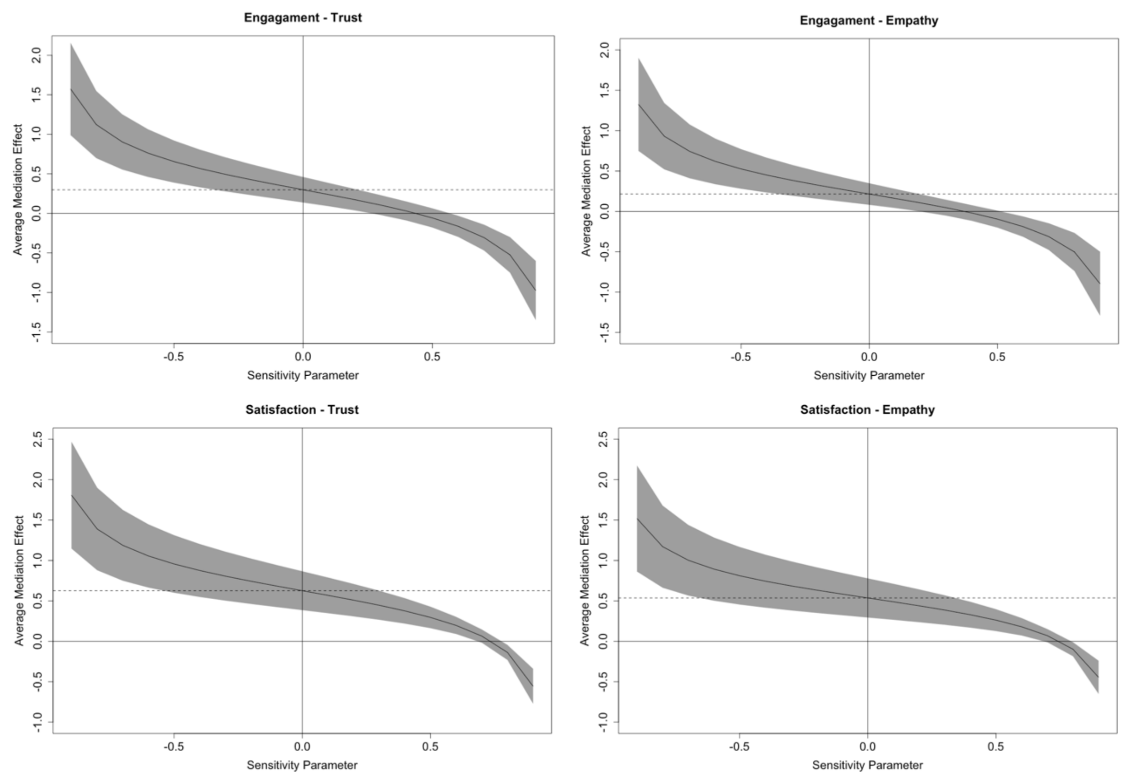
12.7. Appendix G: Causal Mediation Analysis

Table A13. Regression Results of Causal Mediation Analysis (Mediator: *Trust*)

	Trust	***Engagement***		***Satisfaction***	
	(1)	(2)	(3)	(4)	(5)
signature	0.64*** (0.12)	0.45*** (0.12)	-0.79 (0.61)	0.30*** (0.12)	1.11* (0.57)
Trust		0.47*** (0.10)	0.36*** (0.09)	0.98*** (0.06)	1.05*** (0.09)
signature × *Trust*			0.35** (0.17)		-0.22 (0.16)
R^2	0.16	0.34	0.36	0.63	0.63
Note: * *p<0.1* ** *p<0.05*, *** *p<0.01*. Robust standard errors in parentheses.					

Table A14. Regression Results of Causal Mediation Analysis (Mediator: *Empathy*)

	Empathy	***Engagement***		***Satisfaction***	
	(1)	(2)	(3)	(4)	(5)
signature	0.59*** (0.13)	0.54*** (0.12)	-0.35 (0.59)	0.39*** (0.12)	1.59*** (0.52)
Empathy		0.37*** (0.08)	0.29*** (0.13)	0.91*** (0.07)	1.02*** (0.08)
signature × *Empathy*			0.25 (0.16)		-0.34** (0.14)
R^2	0.12	0.29	0.31	0.64	0.65
Note: * *p<0.1* ** *p<0.05*, *** *p<0.01*. Robust standard errors in parentheses.					



Note: This figure reports the sensitivity analysis of the detected mediation effects to the violation of the sequential ignorability assumption following Imai et al. (2010). Take the top-left panel as an example, the average mediation effect would disappear or change its sign if the sensitivity parameter is greater than 0.40.

Figure A3. Sensitivity Analysis