Do Customer Emotions Affect Agent Speed? An Empirical Study of Emotional Load in Online Customer Contact Centers

Daniel Altman[a], Galit B. Yom-Tov[a], Marcelo Olivares[b], Shelly Ashtar[a], Anat Rafaeli[a]
[a]Technion—Israel Institute of Technology, [b]University of Chile

Problem Definition: Research in Operations Management has focused mainly on system-level load, ignoring the fact that service agents and customers express a variety of emotions that may impact service processes and outcomes. We introduce the concept of *emotional load*—the emotional demands that customer behaviors impose on service agents—to analyze how customer emotions affect service worker's behavior.

Academic / Practical Relevance: Most theories in Organizational Behavior literature predict that emotions expressed by customers reduce agent's cognitive abilities and therefore should reduce the agent's speed (e.g. by increasing the service time required to serve an angry customer). We aim to shed light on the magnitude of that phenomenon while addressing important econometric challenges. We also investigate an important mechanism that drives this relation, namely, agent effort. We discuss practical opportunities that arise from measuring emotional load, and how it can be used to enhance productivity.

Methodology: We measure the emotional load of agents using sentiment analysis tools that quantify positive/negative customer emotion expressions in an online chat-type contact center, and link it to agent behavior: response time, and the length and number of messages required to complete a service request. Identifying a causal effect of customer emotion on agent behavior using observational data is challenging because there are confounding factors associated to the complexity of service requests, which are related to both customer emotions and agent behavior. Our identification strategy uses panel data and exploits the variation across messages within a focal request, using fixed effects to control for unobserved factors associated to case complexity. Instrumental variables are also used to address issues of measurement error and other endogeneity problems; the instruments are based on exogenous shocks to agent performance indicators that have been studied in the service operations literature.

Results: Analyses show that emotional load created by negative customer emotions increases agent response time (RT), the length of the agent messages (a measure of effort) and the required number of messages needed to complete a service request. Emotional load and agent RT reciprocally effect each other, with long agent RTs and a high number of messages producing more negative customer emotion.

Managerial Implications: We suggest that the emotional content in customer communications should be an important factor to consider when assigning workload to agents in a service system. Our study provides a rigorous methodology to measure the emotional content from customer text messages and objectively evaluate its associated workload. We discuss how this can be used to improve staffing decisions and dynamic workload routing through real-time monitoring of emotional load.

---

1. Introduction

Toward understanding agent efficiency in service systems, research in Operations Management (OM) has investigated the impact of system load on agent efficiency (e.g., Kc and Terwiesch 2009, Song et al. 2015). Studies on the effects of system load are inconclusive, with some research indicating it increases efficiency and others showing it decreases efficiency (see Delasay et al. (2019) for a review of mechanisms that might explain this confusion). In the spirit of incorporating human behavioral aspects into OM research (Cho et al. 2019), we propose that the behavior of agents and emotions that customers express, generally ignored in operations research (Field et al. 2018), should be added to this discussion. We suggest that this salient aspect—of customer expressed emotion—can promote the understanding of agent performance-related behaviors (e.g., speed, effort) and help improve understanding and management of service delivery.

Research in Organizational Behavior (OB) describes the effects of emotions that people express toward other people, be it in negotiations (van Kleef et al. 2004), or in other forms of social interactions (Hareli and Rafaeli 2008). Lab experiments, for example, show that customer emotions affect the speed, accuracy and fatigue of service agents (Rafaeli et al. 2012). Studies also show that negative customer emotions lead to agent incivility (Walker et al. 2017), and that the amount and valence of emotions that customers express influence service agents (Grandey et al. 2004, 2010). Building on such research, we conceptualize Emotional Load as the ***amount of customer emotion that a service agent encounters and must handle***. Emotional load complements the construct of operational (or offered) load, recognizing and incorporating variability between people into analyses of service work. Emotional load adds an additional dimension to the load that service delivery agents experience.

This research offers five contributions to current research on service delivery. First, we propose that emotional load can be estimated by measuring emotions that customers express to agents. Second, we show the effects of emotional load on operational measures, notably agent response time to customers and number of turns a service interaction requires. We show these effects are above and beyond the effects of operational load. Third, we investigate one of the mechanisms that explains the effects of emotional load, agent effort. Fourth, we examine both the influence of customer emotions on service agents' behavior (i.e., response time) and the subsequent influence of that agent behavior (i.e., response time) on customer emotions, within the same data. Finally, we use automated sentiment analysis to analyze customer emotions in a large sample of authentic service conversations. Our analyses provide important foundations for evaluating efficiency and optimizing work allocation. To the best of our knowledge, this is the first study to recognize and analyze the dynamic nature of the emotionally-charged customer-agent conversations.

Our paradigm overcomes multiple biases and limitations of previous research (Donaldson and Grant-Vallone 2002), which was conducted primarily by OB scholars, and relied extensively on lab simulations (cf. Rafaeli et al. 2012, van Kleef et al. 2004), and self-report measures (cf. Wang et al. 2011). By using automated sentiment analysis (Thelwall 2013, Yom-Tov et al. 2018), we obtain unbiased measures of customer emotion from real-life data, and provide clear operational and managerial implications. We analyze *individual messages* within customer-agent conversations as instances of customer expression of emotion and agent work behavior. This focus offers high resolution into the dynamics *within* conversations. Also, our findings expand beyond the impact of negative customer emotions, which has been the focus of most research, to include also the effects of positive customer emotion (cf. Goes et al. 2018).

The context of our study is contact-center service, which is technology-mediated, and allows access to detailed data and measures of both agent and customer behavior (Rafaeli et al. 2017, 2019). Specifically, we analyze 141,654 customer-agent conversations from the archives of a large western transportation company. We empirically test the impact of emotional load created by customers on (a) agent response time to customers, (b) agent effort, and (c) number of turns/iterations required to complete the service.

Our main dependent variable is agent response time (RT) *to a specific message of a focal customer*. A key challenge we embrace is estimation of *causal* effects using the variation within service conversations. Our analyses show that higher emotional load, in the form of negative customer emotion, increases agent RT and the effect is 2.66 times larger in magnitude than the effect of agent multitasking, and stronger than system-level load (queue length). Negative customer emotion increases the length of text in agent replies by 4.3% and positive emotion increases the length of text in agent replies by 2%, compared to the text of neutral message. In addition, a one-point increase in negative customer emotion increases agent RT by 19.7%. Considering the reverse effect of agent RT on customer emotion, we show that if the agent doubles the RT, customer emotion decreases by about 0.1 standard deviations. This finding has implications for acceptable levels of in-service waits (i.e., waiting during customer length of stay) that result from concurrency decisions.

2. Context of the Study and Data Description

The current study is based on data provided by LivePerson Inc., a firm that offers a web-based service platform. The platform allows end customers to interact with agents of a service brand, through written "chat" messages. Customers who want to chat with a live agent enter a queue and wait for an available agent. Service chats comprise iterations of agent and customer written messages.

A feature unique to chat service platforms is that agents can simultaneously interact with multiple customers (maximum of 3 customers in our data). Agents waiting for a focal customer to respond can turn to interact with other customers. The implication is that if an agent is busy with one customer, his or her other concurrent customers must wait. Customers are not explicitly informed of this agent multitasking and do not know why an agent's response is delayed.

2.1. Data Description and Definitions

Our data includes 141,654 service conversations conducted from March 2016 to April 2017, by agents of a western transportation company. We use the terms "chat" and "conversation" interchangeably to refer to a full service interaction between an agent and a customer. Each conversation in the data includes agent and customer lines, as well as system lines, which are automatically generated, and not included in our analyses since they do not reflect any human input. The term "line" refers to a single parcel of text sent by a customer/agent (i.e., followed by pressing "enter") and "message" refers to one or more lines sent, uninterrupted, by a customer or agent. That is, a series of lines sent by an agent or customer are collapsed into one message. Figure 1 offers a schematic view of the simultaneous chats of one agent who is handling three customers, where each chat comprises multiple messages. Chats in our sample last on average 11.7 minutes ($SD = 9.46$), and include on average 5.40 customer messages ($SD = 3.54$) and 5.78 agent messages ($SD = 3.50$).

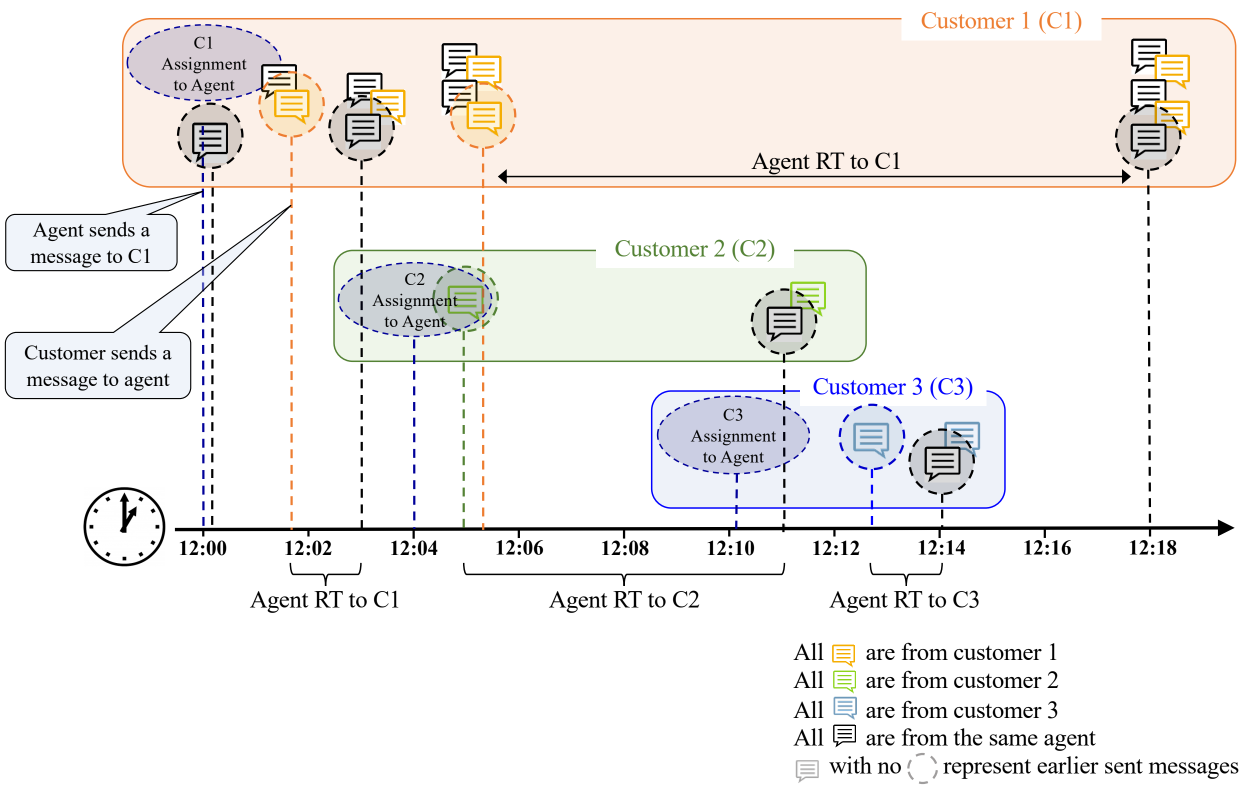
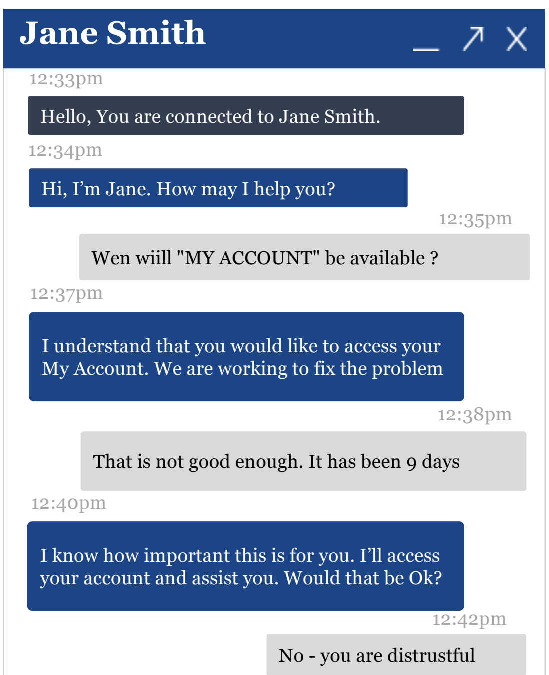


Figure 1 **Schematic View of Simultaneous Agent Chats with Three Customers.**

Each conversation is identified by a chat ID, agent ID, date, issue (sales or service), and time the customer waited in a queue before chatting. For each line within a chat our data contains the following: a time-stamp of when the line was sent, notation of who wrote the line (customer or agent), number of words, and an emotion score. To ensure privacy, our data do not include the text of the conversations or any demographic indicators of customers or agents.

2.2. Measuring Operational Features of Conversations

Figure 2 provides an example of a chat, its recorded data, customer emotion, and two computed variables: agent RT and number of turns. *Agent RT* is computed as the elapsed time between each customer message and the agent response. The number of turns is computed as the total number of customer-agent iterations.



Author	Agent Response Time (RT)	Number of Words	Emotion	Turn
System	-	-	-	-
Agent	-	8	-	1
Customer	-	6	0 (neutral)	2
Agent	75 sec	18	-	3
Customer	-	10	-1 (negative)	4
Agent	82 sec	19	-	5
Customer	-	4	-1 (negative)	6
Total Number of Turns	-	-	-	6

Figure 2 **An illustration of Agent-Customer Chat and Measures.**

We compute agent RT (rather than service time or customer length of stay (LOS)) because (1) agent RT translates directly into agent efficiency; (2) agent RT defines the customer wait time experience, which service delivery must minimize; (3) agent RT is free from endogenous time intervals, such as customer RT, so is preferable to LOS. We note that due to concurrency, agent

RT is a result of tasks being performed for a focal customer and for other customers. Customers are generally blind to agent's work processes. Customers can see a note indicating when the agent is typing to them; but we do not have the records of such notes, so could not include this in our analyses. Hence, our data does not allow a "clean" decomposition of agent RT into focal customer service time and service times to other customers. Therefore as a proxy for agent effort, we use the number of words in each agent message, similar to Goes et al. (2018). Using the meta-data described, we calculate the number of concurrent customers assigned to each agent, and control for agent multitasking by including this concurrency measure. We also control for the system load, using the number of customers waiting in the queue. This information is also available on the agent screen.

2.3. Measuring Customer Emotions in Conversations

We measure customer expressions of positive and negative emotions as two sides of a single scale (Fredrickson and Kahneman 1993, Gabriel and Diefendorff 2015), and use the terms "emotion" and "sentiment" interchangeably to refer to customer expressions of emotion. We reviewed multiple Sentiment Analysis tools (e.g., *LIWC*, Tausczik and Pennebaker (2010), *SentiStrength*, Thelwall (2013), *CustSent*, Yom-Tov et al. (2018), and *SentimentTreebank*, Socher et al. (2013)), and selected two tools —*SentiStrength* and *CustSent*—that offer the most accurate assessments of customer emotion in chat service. *SentiStrength* was developed to assess positive and negative emotion in short texts, and *CustSent* was designed to analyze sentiment in customer service conversations. Both tools utilize labeled dictionaries coupled with Natural Language Processing techniques, and have better accuracy than other tools in the customer service context: Yom-Tov et al. (2018) reports that *SentiStrength* has the highest recall, and *CustSent* has the highest precision values with customer service texts (see Appendix EC.1 for recall and precision data of the tools).[1]

These two tools assign a valence and intensity value for the emotion expressed in a message. Negative and positive signs represent negative and positive emotions, respectively. The score itself indicates the intensity of the emotion. *SentiStrength* sentiment scores range from -4 to +4. For example, the following text received a score of +1:

"That enabled me to access my account. Thanks, that's really helpful."

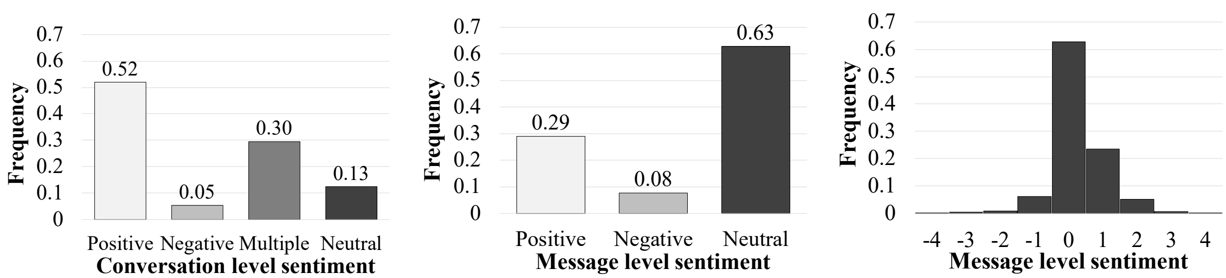
[1] Following reviewer queries, we considered also measuring agent emotions. We used the same tools to analyze agent sentiment, searching for instances of agent expression of negative emotion in a sample of about 200 agent messages. We found that all agent messages that express what the tools construed as negative emotions include some version of apology (e.g.,"I am so sorry you had to wait.") or reassurance (e.g., "Don't worry, we'll find the annoying mistake"). We found no agent messages with negative emotions such as anger or frustration. Thus, the sentiment score of agent messages is qualitatively different from the sentiment score of customer messages. We therefore do not develop hypotheses about agent emotion. We included agent sentiment in our analyses of the robustness tests, and the results did not change much (see §5.3).

In contrast, the following text received a score of -1:

"I don't know. I'm concerned about my credited miles."

CustSent has no hard limits on the sentiment scores, but in our data these scores range from -12 to +10. As reported below, to reduce measurement error we combine the scores of the two tools in our analyses (see Section 4).

Figure 3 describes the customer emotions evaluated by *SentiStrength* in our data. Figure 3(a) shows the proportion of chats having only positive emotion, only negative emotion, multiple emotion (both positive and negative) and neutral. More than 85% of chats include emotion, which positions emotion as a central feature of service. Figure 3(b) shows the proportion of customer messages that contain positive, negative or neutral expressions, and suggests that most messages within conversations are neutral. Both the chat and message analysis show that positive emotion is more commonly expressed than negative emotion. Figure 3(c), which shows the distribution of emotion intensity in messages, further confirms the higher prevalence of positive emotion.



(a) Full Service Conversations (b) Customer Messages (c) Sentiment Distribution

Figure 3 Frequency of Emotion in Customer Service (*SentiStregnth* Score)

Figure 4 graphically depicts the association between customer emotion and agent RT, showing a kernel smoothing of average agent RT (throughout the chat) as a function of average customer emotion. The apparent relationship between emotional load and agent RT is analyzed below by testing the causal effects in this relationship, while controlling multiple relevant factors. Next we formulate hypotheses that relate agent behavior and customer emotion and then test these hypotheses with an econometric framework in Section 4.

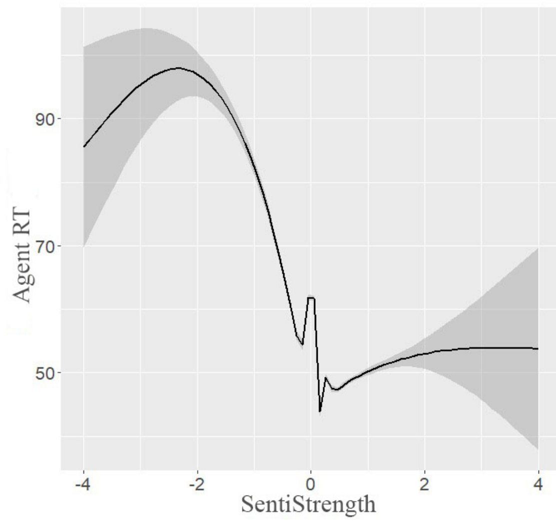


Figure 4 **Covariation of Agent RT and Customer Sentiment. Marked in gray is the 95% Confidence Interval.**

3. Theory Development

Operational and marketing perspectives typically consider customer emotions as responses to agent behavior and as indicators of customer satisfaction; mere outcomes of an interaction. In contrast, we view customer emotion as a unique source of load for service agents, and propose that such emotional load influences agent performance-related behavior, and specifically agent RT. Below, we first review literature that supports our emotional load theory in §3.1 and then, in §3.2, we discuss the opposite, and more prevalent, view that agent RT impacts customer emotion.

3.1. Effects of Customer Emotions on Agent Behavior

The Episodic Model of Affect and Performance (Weiss and Cropanzano 1996), positions work as a series of episodes in which emotional experiences vary, and influence work performance (Beal et al. 2005). The model suggests that emotional events at work (e.g., exposure to an angry customer), influence service agents, because they affect mental resources. This model is the foundation for our predictions. Experimental research has shown that customer rudeness and anger hamper service agents' performance of various tasks (Rafaeli et al. 2012), due to disruption of cognitive processes (Porath and Erez 2007). To illustrate, participants in a simulation of customer service work erred more when processing customer requests phrased in a hostile manner than when requests were phrased politely (Goldberg and Grandey 2007). Similarly, a series of lab studies showed that listening to verbally abusive customers hampered participants' ability to recall the content of the conversation (Rafaeli et al. 2012). Building on this, we expect that agents need extra time to resolve customer issues expressed with negative emotion.

Moreover, agents who encounter customer emotions must often suppress their own genuine emotions, and display organizationally appropriate responses (Geddes and Callister 2007), performing the demanding task known as "*Emotional Labor*" (Rafaeli and Sutton 1987, Grandey et al. 2010). The additional effort required to convey appropriate emotions likely requires extra time from the agent (Sutton and Rafaeli 1988). In this vein, when customer emotion is ***positive***, agent emotion corresponds to the appropriate response and so no extra effort is required for the agent to express their response. Additionally, positive customer emotion is replenishing, and improves agent motivation and available cognitive resources (Bakker and Demerouti 2007), both of which help agents solve customer issues more rapidly. Hence, our first hypothesis:

HYPOTHESIS 1. *The more negative emotion a focal customer expresses in a given message, the longer the agent RT in the subsequent agent message.*

Although multiple mechanisms could explain Hypothesis 1, we test what we see as a key mechanism—agent effort. Customer messages that include negative emotion, require additional communication effort (compared to positive/neutral messages), in addition to the effort required to generally resolve the customer issue (Geddes and Callister 2007). For example, agents must acknowledge customer frustration or dissatisfaction and may need to apologize to customers. These additional communication efforts will lengthen the agent text. Hence, our second hypothesis:

HYPOTHESIS 2. *The more negative emotion a focal customer expresses in a given message, the larger the agent effort in the subsequent agent message.*

We position agent effort as a mechanism through which customer emotion influences agent RT, suggesting that agent effort acts as a mediator. Hence our next hypothesis:

HYPOTHESIS 3. *Agent effort mediates the effect of customer emotion on agent RT.*

The hypotheses presented so far are relevant to the message-level (i.e., messages within conversations). A conversation-level analysis is relevant for considering the effects of customer emotion on the length of a conversation. Two effects confirmed by Rafaeli et al. (2012), Porath and Erez (2007) and others suggest that negative customer emotions will prolong a service conversation. Customer expressions of negative emotions hamper agents' cognitive processing and increase agent errors, which extend the length of a conversation. In addition, customers' negative emotions distract agents, leading to more agent inquiries as the agent seeks to understand the customer needs. Hence, our next hypothesis:

HYPOTHESIS 4. *The more negative emotion a focal customer expresses during a conversation, the greater the number of turns required to complete the conversation.*

We note that a competing hypothesis for Hypothesis 4, as suggested by Sutton and Rafaeli (1988), could be that customer expressions of positive emotion create more customer engagement, and extend the service conversation. Similarly, positive emotions create a more pleasant work environment, and may motivate agents to spend more time (and thus exchange more messages) in the conversation. But this competing theory does not have strong support in empirical research.

3.2. Effects of Agent Behavior on Customer Emotions

Another side of the customer-agent interaction is the influence of agent RT on customer emotions. Customers can construe an agent's RT as wait time, leading to expressions of negative customer emotion. Customers dislike waiting (Maister 1984, Larson 1987, Taylor 1994), so much so that people waiting often abandon a service (Mandelbaum and Zeltyn 2013, Allon et al. 2011). Importantly, agent RT in chat service can include delays and in-service waits due to concurrency of other customers (Goes et al. 2018). This can create unexplained waiting which may annoy and frustrate customers, evoking expressions of negative emotion. Hence, our next hypothesis:

Hypothesis 5. *Longer agent RT to a focal customer message creates more negative emotion in the subsequent customer message.*

Another way that agents might influence customer emotion is by investing effort. Customers are more satisfied when they feel that an agent works harder to resolve their issue (Groth et al. 2009). Customers seek specific cues to assess the agent effort, and cues such as the time and energy an agent spends on a customer can impact perceived effort above and beyond the outcome of the service (Mohr and Bitner 1995). In the context of chat-service, customers are detached from the service agent and cannot see when an agent is working toward solving their inquiry. However, customers can perceive invested agent time and effort through the length of an agent's message. If the number of words an agent writes is indeed a proxy for the customer's perception of agent effort, then we would expect that customers who encounter long messages will be more satisfied, and hence will express more positive emotions. In the same spirit, shorter messages would signal reduced agent effort and lead to customer dissatisfaction and the expression of more negative customer emotions. Hence, our next hypothesis:

Hypothesis 6. *An increase in agent effort (the number of words in an agent message) creates more positive customer emotion in the subsequent message.*

Finally, effects on customers may also accrue as a conversation unfolds. Customers participate in a service conversation to accomplish a goal or a set of goals. Long service conversations can make customers frustrated (Katz et al. 1991) and angry (Casado Diaz and Más Ruíz 2002). When a service conversation is very long, customers may strategically express "fake" anger, to signal dominance and toughness (Knutson 1996, Tiedens 2001). Customers can also perceive long service times as unprofessional (Anand et al. 2011, Casado Diaz and Más Ruíz 2002), since a longer conversation might signal that the agent is unable to solve the customer's problem. A sense of unprofessional service can translate into customer expressions of negative emotion. Hence, our final hypothesis:

HYPOTHESIS 7. *As the number of turns within a conversation increases, customers express more negative emotion.*

In the remainder of this paper, we empirically examine our hypotheses and decipher the complex relationship between customer expressed emotion and agent behavior.

4. Econometric Specification

This section develops an econometric framework to test the causal effects our hypotheses predict. An important challenge in the estimation arises from omitted factors related to the complexity of a focal case handled by an agent. Cases of higher complexity are likely to be associated with longer agent RTs, because they require more effort to handle. More complex cases are also likely to evoke negative customer emotion. We can include some observable proxies of case complexity in the model, but there are dimensions of complexity which cannot be measured and therefore become confounds that can bias the estimates. A second complication is reversed causality between agent behavior and customer emotion, whereby longer agent RTs may enhance customer frustration. This produces a simultaneity problem between customer emotion and agent behavior: cases that take longer to handle also tend to have negative customer emotions, and the causal relationships between the two is not clear (Manski 1993).

The empirical strategy we used to identify the causal effect of emotion on agent behavior is to exploit the panel structure of the data, using variation across the sequence of messages within a conversation as a source of identification. Let i index the customer-agent conversation associated to a case and let $NTurns_i$ denote the number of turns, with $t = 1 \dots NTurns_i$ representing each turn within that conversation. The variable EMO_{it} measures the emotion of a customer message in turn t, and RT_{it} the agent response time to a message t. RT is modeled as:

$$\log(RT_{it}) = \delta_i + \beta EMO_{it-1} + \gamma W_{it} + \tau ConvStage_{it} + u_{it}, \tag{1}$$

where δ_i is a fixed effect for the conversation, W_{it} are workload related factors that vary during the conversation and u_{it} is an error term. The coefficient of interest is β, which we predict to be negative according to Hypothesis 1. Other applications with similar data revealed that EMO has a positive trend during a conversation (Yom-Tov et al. 2018). To account for this trend, the covariate $ConvStage_{it} = t/NTurns_i$ is specified to capture the stage of conversation i where the focal turn t occurs. This control variable is included in the econometric models that are analyzed at the message (it) level.

The fixed effect δ_i controls for several unobserved factors that could lead to omitted variable bias. In particular, it captures the complexity of a case, which by definition does not vary during the case conversation. Because conversations last on the order of minutes (11.7 minutes on average), effects due to day of the week and hour of the day are also captured by δ_i. Because a conversation is handled by a single agent, all agent related factors are also absorbed by the fixed effect.

Previous work showed several mechanisms that relate workload to agent productivity (for a review see Delasay et al. (2019)). Workload can affect the speed of an agent's work by leading to fatigue, thereby reducing productivity and compliance with process standards (Dai et al. 2015). On the other hand, current and pending workload can put pressure on an agent to work harder and increase productivity (Kc and Terwiesch 2009, Tan and Netessine 2014). In settings with a shared queue among multiple agents, social loafing can lead agents to slow down when facing a long queue (Wang and Zhou 2016). To capture the effects of the customer queue, a covariate measuring the number of customers in the queue at the beginning of the RT interval, $NumInQueue_{it}$ is included as a control.

Agents in chat contact centers can simultaneously handle multiple conversations, a workload that can also create fatigue and pressure effects. Handling simultaneous conversations is a form of multitasking, known to also affect productivity (Kc 2013, Bray et al. 2016, Goes et al. 2018). The number of concurrent chats ($Concurrent_{it}$) is measured as an average during the RT interval. Given the dynamics of work assignment in contact centers, both $NumInQueue$ and $Concurrent$ can vary substantially during the course of a conversation, but are not controlled by the agent and are therefore considered exogenous; these two variables are the main covariates included in W_{it} (section 5.3 considers alternative measures of concurrency). Other workload related effects, such as the hours elapsed during the working shift, do not vary much during a conversation due to its relatively short duration, and are therefore absorbed in the fixed effect δ_i.

Identification in this model is driven by the variation in emotion across customer messages during the same conversation. Recall that one of the concerns regarding the identification of the causal effect of emotion on agent behavior was reverse causality: it is possible that EMO and RT affect each other. Our regression Model (1) exploits the sequencing of the messages to avoid this reverse causality. The variable RT_{it} is measured *after* the customer expresses emotion in his/her message in turn $t-1$, hence it could not have influenced EMO_{it-1}. Furthermore, the detailed model—including the conversation fixed-effect δ_i—controls for most of the omitted variables related to case and agent heterogeneity, providing a clean identification strategy.

A final concern for identification is measurement error in the EMO variable, which could lead to an attenuation bias in the associated coefficient. We started with one measure of EMO for the analysis *SentiStrength* (Thelwall 2013), and used a second measure, *CustSent* (Yom-Tov et al.

2018) to mitigate concerns about measurement error in the first measure. As noted in Section 2.3, there are differences between the two measures, both in the dictionaries they use and in the range of sentiment scores. Nonetheless, the two measures are highly positively correlated ($r = 0.63$, $p < 0.001$). Hence we use the second *CustSent* emotion measure as an Instrumental Variable (IV), which eliminates measurement error (Wansbeek and Meijer 2000), in all of the models with EMO as an independent variable.

4.1. Decomposing the Effect of Customer Emotions on Agent Behavior

The effect of customer emotion on agent RT can be direct or indirect (mediated) through agent effort (see Hypotheses 1 and 3). We use the number of words in agents' responses as a proxy for agent effort, similar to Goes et al. (2018). The number of words in an agent message ($NumWords_{it}$) is included in the specifications as follows:

$$\log(RT_{it}) = \delta_i + \beta_1 EMO_{it-1} + \beta_2 log(NumWords_{it}) + \gamma W_{it} + \tau ConvStage_{it} + u_{it}; \quad (2)$$

$$log(NumWords_{it}) = \delta_i + \beta_3 EMO_{it-1} + \gamma W_{it} + \tau ConvStage_{it} + v_{it}. \quad (3)$$

This specification, which includes the same control variables as Model (1), captures the direct and indirect effects of EMO on RT (Hypotheses 1 and 3). Coefficient β_3 captures the effect of customer emotion on agent effort in responding to the customer (Hypothesis 2), using $NumWords$ as a proxy for effort. This effect translates into an impact on RT because longer text requires more time to write ($\beta_2 > 0$). The coefficient β_1 captures other effects of emotion on RT, since $NumWords$ may not be a perfect proxy. Thus, β_1 may include other effort-related aspects not reflected in a longer agent message (e.g., scrutinized search in the CRM software). We do not have documentation of agents activity outside of the chat platform, and therefore cannot measure the full agent effort directly. Despite these limitations, Models (2) and (3) provide information about the alternative paths through which customer emotion affects agent behavior. As before, we correct for measurement errors and include the other control variables (with some abuse of notation, the same parameters are used for the controls' coefficients to facilitate reading).

Models (2) and (3) correspond to a mediation model where the effect of EMO on the agent RT can be decomposed into a direct effect (coefficient β_1) and an indirect effect through $NumWords$ (measured by $\beta_3 \times \beta_2$). A key assumption to identify the coefficients $\beta = (\beta_1, \beta_2, \beta_3)$ is that u_{it} and v_{it} are independent, that is, unobservable factors that affect $NumWords$ do not directly affect RT (conditional in all the controls of the model). Recall that the models include conversation fixed effects, which control for the case complexity and customer and agent characteristics; these controls are needed to justify this identification assumption. Under these conditions, Models (2) and (3) can be estimated as independent regression models (using IVs to mitigate the measurement error of EMO) to provide consistent estimates of the model parameters. However, calculating confidence intervals for the indirect effect $\beta_3 \times \beta_2$ is complicated because the two estimators are correlated due to sampling error. We use a bootstrapping approach to estimate the models and compute confidence intervals, using the methods developed by Hayes and Rockwood (2020) to conduct mediation analysis with panel data.

4.2. Effect of Customer Emotions on the Length of a Conversation

Section 4.1 discussed the effect of customer emotions on agent RT to each individual customer message. Hypothesis 4 extended our predictions to suggest that customer emotion affects the number of turns ($NTurns_i$) in a conversation.

Estimating the effect of emotions on the number of turns in a conversation requires a different approach. We propose here two identification strategies (Models (4) and (5)). First, we model the number of turns as a static variable that we measure at the conversation level. In this case, the basic unit of analysis is a conversation. The model includes the effect of emotion in the first customer message of the conversation, EMO_{i1}. According to Hypothesis 4, we expect that the coefficient of EMO, β_4, to be negative. One may be tempted to average the emotion across all messages in a conversation, but this is problematic due to the reverse causality problem discussed earlier: customer emotion affects agent behavior but agent behavior also affects customer emotion. Furthermore, measuring the impact of the customer emotion in the first message can be useful for balancing work allocation between agents (see Section 6). To account for agent workload we include average concurrency during a conversation ($Concurrent$), and the number of customers in queue when the conversation started ($NumInQueue$). Both indicators are exogenous, therefore, reversed causality is not a concern here. We use the following regression model to estimate the impact of customer emotion on the number of turns:

$$NTurns_i = \rho_{a(i)} + \beta_4 EMO_{i1} + \gamma W_i + \psi X_i + w_i. \tag{4}$$

The term $\rho_{a(i)}$ is a fixed effect of the agent serving chat i and w_i is an error term. The other covariates in Model (4) are discussed next. Since the model is estimated with a cross-section of conversations, it is important to control for case complexity. The number of words in the first customer message ($CustWords_1$) is an exogenous variable used to proxy the complexity of the case, included as a covariate with log transformation (to keep consistency with the previous models). To capture seasonal effects, a weekday-weekend dummy and hour of the day dummies are included ($IsWeekend$ and $HourOfDay$, respectively). The type of service case ($SrvType$) is controlled through a dummy variable. Finally, changes in agent behavior due to fatigue are controlled with dummy variables for each hour worked during the shift ($ShiftTime$). These covariates are included in the set of controls denoted by X_i.

As predicted by Hypothesis 4, customer emotion during the conversation can also affect the extension of the conversation, as measured by the number of turns. For this we use an alternative identification strategy. Consider a discrete-time duration model, where "periods" are represented by each turn in a conversation. Define y_{it} as a binary dependent variable which is equal to one if conversation i ends in turn t, and zero otherwise. The $\Pr(y_{it}=1)$ can be viewed as a hazard rate of the length of a conversation and can be modeled through a Probit model as:

$$\Phi^{-1}\left(\Pr(y_{it}=1)\right)=\rho_{a(i)}+\beta_5 EMO_{it-1}+\beta_6 EMO_{i1}+\gamma W_{it}+\phi X_i, \tag{5}$$

where W_{it} are the workload-related variables included in Models (1), (2) and (3), and $\Phi^{-1}(\cdot)$ is the inverse of the standard normal distribution. Since there is only one spell of messages for each conversation, this model cannot include conversation fixed effects (because $y_{it}=1$ only for the last turn of each conversation i). Therefore, the same control variables X_i from Model (4) are included in this model, to capture cross-sectional differences across conversations. Additional specifications were estimated including the emotion in the first customer message, EMO_{i1}, as a proxy for potential observable factors that could be correlated with the initial emotion of each conversation. The coefficient of interest in Model (5) is β_5, which measures the impact of customer emotion in the previous message on the hazard rate (likelihood of terminating the conversation); Hypothesis 4 predicts a positive effect, $\beta_5>0$ — the more positive the emotion, the shorter the conversation should be and the probability that the conversation will end in the next turn should increase. As before, EMO was instrumented in Models (4) and (5) to reduce measurement error.

4.3. Modeling the Effect of Agent Behavior on Customer Emotions

Our next hypothesis regards the influence of agent behavior on customer emotion (Hypothesis 5). The empirical strategy we used to test this hypothesis is as follows. First, we consider the following specification to estimate the effect of agent RT on customer emotion:

$$EMO_{it}=\delta_i+\alpha\log(RT_{it-1})+\tau ConvStage_{it}+e_{it}. \tag{6}$$

The unobservable e_{it} includes the quality of the agent response as perceived by the customer, which is difficult to control with the variables observed in the data. It is then plausible that agent RT is positively correlated with the quality of the agent response, since agents need to do time consuming work to properly address a customer issue. This positive correlation between RT and the error term induces a positive bias in the estimation of α. Our approach to correct for this bias is to use IVs that affect agent RT but do not *directly* affect customer emotion. Recall from Model (1) that RT is affected by the agent workload, W_{it}. In the context of this application, customers cannot directly observe the workload of the agent, thereby the effect of this workload can only affect emotion through the RT perceived by the customer. Measuring the effect of RT induced by variation in an agent's workload is also useful from a managerial perspective, as it provides insights on how workload management and staffing decisions can affect customer emotion. According to Hypothesis 5, we expect the coefficient α to be negative.

Model (6) can be further refined by including additional factors associated with agent effort, specifically, $NumWords$ and $Turn$:

$$EMO_{it} = \delta_i + \alpha_1 \log(RT_{it-1}) + \alpha_2 log(NumWords_{it-1}) + \alpha_3 Turn_{it} + \tau ConvStage_{it} + e_{it}. \tag{7}$$

The number of words ($NumWords$) in a message (our proxy for agent effort), is directly observable by the customer. Longer agent messages might be perceived by customers as increased agent effort, thereby generating positive emotion (see Hypothesis 6). We therefore expect the coefficient α_2 to be positive. As noted, customers cannot see all the activities performed by an agent during the RT, and may therefore interpret a long RT as lack of agent dedication, which would produce negative customer emotion. According to Hypothesis 7, customer emotion can also be affected by an extension of the conversation, which is captured through the variable $Turn_{it}$ (i.e., the ordinal count of turns in a conversation). Notice that $Turn_{it}$ and $ConvStage_{it}$ are correlated but not perfectly co-linear, hence their effects can be identified separately and with reasonable precision given the large sample size. A potential issue is that RT, $NumWords$ and $Turn$ can all be correlated with the complexity of the customer issue, since more complex issues require more effort from the agent and a longer conversation. But recall that the fixed-effect δ_i controls for case complexity, mitigating this omitted variable bias. As before, RT is instrumented with the workload-related exogenous variables W ($Concurrent$ and $NumInQueue$) in order to mitigate the endogeneity bias that can be generated by unobservable quality of the agent's response.

Table 1 summarizes the variables used in all the econometric models. The next section discusses further specification details, summary statistics and the estimation results.

5. Estimation Results

Table 2 reports summary statistics of the variables used in the estimation. The top panel shows the variables included in Models (1)–(3) and (5)–(7), with messages as the unit of analysis and the bottom panel shows variables of Model (4), with conversation as the unit of analysis. In both cases outliers were removed from the sample, in order to avoid influence of extreme cases on the estimation. The Max column indicates the cutoffs used for excluding outliers. In the message-level data, we defined outliers as observations with RT below the 5th percentile (below 8 seconds) and above the 95th percentile (above 1641 seconds). We removed observations where $NumWords$ was above the 95th percentile (387 words). We also removed conversations with data errors in the

Table 1 Labels and Coding of Study Variables

Variable	Description and coding
Dependent variables	
RT_{it}	Agent response time to a focal customer message in turn t of conversation i [seconds]
$NumWords_{it}$	Number of words agent wrote to a focal customer in turn t of conversation i (a proxy of agent effort)
$NTurns_i$	Number of iterations between customer and agent in conversation i (an iteration is counted when one party answers the second party)
EMO_{it}	Customer emotion in turn t of conversation i as measured by *SentiStrength*
W variables:	Agent workload
$NumInQueue_{it}$	Number of customers in queue at the beginning of turn t of conversation i
$Concurrent_{it}$	Weighted average of number of parallel chats handled by agent during turn t of conversation i
X variables:	Complexity of problem and time variables
$SrvType_i$	Type of service in conversation i: support (coded 0; 50.81%) or sales (coded 1)
$CustWords_{it}$	Number of words customer wrote in turn t of a conversation i
$ShiftTime_i$	Time that passed since an agent started the shift until the beginning of a conversation i [hours]
$HourOfDay_i$	Hour (8:00-23:00) of the conversation i
$IsWeekend_i$	Weekday: Mon-Fri (coded 0; 72.24%), Weekend: Sat-Sun (coded 1)
Other variables:	
$Turn_{it}$	Ordinal number of current turn t in a conversation i
$ConvStage_{it}$	Progress of conversation completed (Range 0-1)
$CustSent_{it}$	Second measure of customer emotion in turn t of conversation i

$ShiftTime$ and conversations that were conducted after the eighth hour of an agent's shift, to focus only on regular shifts (95% of conversations). The elimination of outliers and chats with missing data removed a total of 75,160 conversations from the analysis, leaving an effective sample size of 141,654 chats. Tables EC.2 and EC.3 in the Appendix show the inter-correlation among the variables. As a robustness check, all analyses were replicated with the outliers included in the sample (see §5.3).

5.1. Effect of Customer Emotions on Agent Behavior

Table 3 shows the estimation results of econometric Models (1), (2), and (3). Recall that Models (1)–(3) are at the message-level of analysis, and include fixed effects of the conversation, so the coefficients are estimated using variation across turns of each conversation. Models (2) and (3) are estimated using a mediation model based on Hayes and Rockwood (2020), using bootstrapping to compute the standard errors[2].

The results of Model (1) confirm a negative and statistically significant effect of EMO on RT, supporting Hypothesis 1. The key covariate for Model (1) is EMO, instrumented with $CustSent$. The effect is substantial, with a one point improvement in customer emotion (i.e., emotion becomes more positive) reducing RT by 20.6% (a 14 second reduction in average RT per message).

Other control variables also have significant effects on RT: $Concurrent$ has a positive effect, meaning that simultaneous conversations with multiple customers increase the RT to each focal

[2] This method is designed especially for panel data, drawing conversations with replacements from the data in the re-sampling procedure. For each re-sample, Equations (2) and (3) are estimated separately using 2SLS, which accounts for the nested nature of the data. Confidence intervals are calculated based on the empirical distribution of the estimates from each re-sample. See Hayes (2018) for a description of this bootstrap process.

Table 2 Descriptive Statistics of Study Variables

Variable	Mean	Median	SD	Min	Max
Message level (N=650,856)					
EMO [$SentiStrength$]	0.27	0.00	0.74	-4	4
$CustSent$	0.22	0.00	0.72	-12	10
RT [seconds]	65.25	47.00	66.10	8	1641
$log(RT)$	3.84	3.85	0.80	2.08	7.4
$NumWords$	34.58	27.00	26.16	1	387
$log(NumWords)$	3.30	3.30	0.72	0	5.96
$Concurrent$	2.33	2.47	0.72	1	3
$NumInQueue$	2.52	1.00	3.87	0	73
$ConvStage$ [%]	0.58	0.58	0.27	0.02	1
$ShiftTime$ [hours]	3.63	3.41	2.31	0	8.16
$Turn$	8.78	6.00	7.99	2	132
Conversation level (N=141,654)					
$NTurns$	10.18	8.00	7.01	2	114
$log(NTurns)$	2.14	2.08	0.6	0.69	4.74
EMO_1 [$SentiStrength$]	0.10	0.00	0.62	-4	4
$CustSent_1$	-0.04	0.00	0.51	-10	7.5
$Concurrent$	2.44	2.65	0.58	1	3
$NumInQueue$	3.13	2.00	4.00	0	72
$CustWords_1$	26.82	23.00	19.49	1	1131
$log(CustWords_1)$	3.00	3.14	0.88	0	7.03
$ShiftTime$ [hours]	3.47	3.26	2.27	0	7.76
$HourOfDay$	14.23	14.00	3.75	8	22

customer. $NumInQueue$ has a small positive effect, meaning that a longer queue of customers makes agents work slightly slower. The effect of $ConvStage$ is positive and suggests an increase in RT toward the end of conversations.

The second and third columns of Table 3 show the results for estimates of the mediation Model (2)–(3). Supporting Hypothesis 2, Model (3) shows that EMO increases log($NumWords$), though the magnitude of the effect is small. Model (3) also shows a negative effect of concurrent conversations: as agents increase multitasking, they write shorter messages to each customer.

Model (2), with log(RT) as a dependent variable, includes EMO and the logarithm of $NumWords$ as the main variables of interest. EMO has a negative and significant effect on RT, similar in magnitude to the estimates of Model (1): a one point increase in EMO reduces RT by 19.7%. The number of words in the message $NumWords$ has a large *positive* effect on RT, which is expected because a longer text takes more time to write. Doubling the length of an agent's message increases RT by 44.6%. The bottom panel shows the indirect effect of EMO on log(RT), with the significant mediation of log($NumWords$), supporting Hypothesis 3. Overall, a one point increase in EMO causes a 19.7% direct reduction in RT plus an indirect effect (through $NumWord$) that increases RT by 0.30%.

The effect of the other covariates in Model (2) are similar to those reported for Model (1), except for $ConvStage$ which now has a smaller magnitude: from 0.246 to 0.006. The longer RTs toward the end of the conversation appear to be partially explained by the length of the messages: Model (3) suggests that agent messages tend to be longer as the conversation progresses.

Table 3 Effect of Customer Emotion on Agent Behavior (Outliers Excluded, EMO_{t-1} is Instrumented using $CustSent_{t-1}$)

	Model(1) $log(RT)$	Model (3) $log(NumWords)$	Model (2) $log(RT)$	Model (3) $log(NumWords)$	Model (2) $log(RT)$
EMO_{t-1}	-0.206*** (0.0028)	0.007** (0.0024)	-0.197*** (0.0024)		
$EMO\_positive_{t-1}$				0.020*** (0.002)	-0.153*** (0.002)
$EMO\_negative_{t-1}$				0.043*** (0.004)	0.042*** (0.004)
$Concurrent_t$	0.057*** (0.0026)	-0.040*** (0.0024)	0.074*** (0.0024)	-0.040*** (0.002)	0.074*** (0.002)
$NumInQueue_t$	0.003*** (0.0007)	0.002* (0.0006)	0.003*** (0.0006)		
$ConvStage_t$	0.246*** (0.0043)	0.464*** (0.0037)	0.006† (0.0004)	0.458*** (0.004)	-0.033*** (0.004)
$log(NumWords_t)$			0.446*** (0.0014)		0.444*** (0.001)
$NumInQueue$ (chat level)				0.001*** (0.001)	0.001* (0.001)
Conversation Fixed Effect	Included	Included	Included	Included	Included
Constant	3.616*** (0.0070)	2.764*** (0.0064)	1.809*** (0.0099)	2.753*** (0.006)	1.831*** (0.01)
Indirect Effects					
EMO via $log(NumWords)$			0.003*** (0.0011)		
$EMO\_positive$					0.009*** (0.001)
$EMO\_negative$					0.019*** (0.002)
Observations	650,856	650,856	650,856	650,159	650,159

Standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ †$p < 0.1$

Note. The Macro we use in the mediation analysis (Hayes and Rockwood 2020) allows up to 3 message-level covariates and up to 3 chat-level covariates. This means that for the estimation with categorical emotion (last two columns) we had to aggregate $NumInQueue$ to the chat-level. This produced some additional missing values, and therefore reduced N size by 697 observations.

The last two columns of Table 3 replicate the estimation of Models (2) and (3) but include three categories of customer emotion: Negative ($EMO < 0$), Neutral (EMO equals to zero) and Positive ($EMO > 0$). The Neutral category is the excluded dummy. Model (3) suggests that positive customer emotions have a small positive effect on $NumWords$: messages with positive emotion (compared to neutral emotion) increase the number of words written by the agent by 2%. Similarly, negative customer emotion (compared to neutral emotion) increases $NumWords$ by 4.3%. Overall, the impact of customer emotion on the length of agents' messages is relatively small, but there is evidence that agents put in more effort when customers express emotions and the effort is greater when the emotion is negative.

The fifth column in Table 3 shows the estimation of Model (2) with categories of customer emotion, and confirms that the effect of customer emotion is negative, monotone, and economically significant. Messages with negative emotion receive RTs which are about 20% longer relative to messages with positive emotion. These results are consistent with the linear specifications. Interestingly, the largest effect is observed in $EMO\_positive$ when positive customer emotion reduces RT by 15.3% compared to neutral emotion.

We next discuss the estimation of Models (4) and (5), which assess the effect of EMO on the length of conversations as measured by the number of turns. The first column of Table 4 shows the estimation of Model (4), using a cross section of conversations and EMO_1—the emotion of the first customer message in the conversation—as the main covariate of interest (which is instrumented to reduce attenuation bias due to measurement error). The coefficient of EMO_1 is negative and statistically significant, where a one point reduction in customer emotion increases the number of turns in the conversation by 1.684, equivalent to 17% of the mean, which is economically significant. This result is aligned with those obtained in Models (1)–(3), providing further support that conversations with negative customer emotion tend to be longer and require more time from the agent. In terms of the other covariates, a higher number of concurrent customers handled by the agent during the conversation reduces the number of turns required to finish the case, suggesting that agents may be speeding-up to close cases faster when their workload is high. The number of words in the first customer message has a negative effect on the number of turns, and the effect of the number of customers in queue is small.

The second and third column in Table 4 report the estimates of two specifications of the hazard Model (5): they both include EMO_{t-1}—customer emotion in the previous message—but the third column also includes EMO_1 as an additional control variable (both specifications are estimated with a Probit model using IVs for EMO_1 and EMO_{t-1}). In both specifications, the lagged EMO has a positive effect on the probability of finishing the conversation. Hence, positive customer emotions are an indication that the conversation is closer to completion. One interpretation of this result is that customer emotion is a proxy for case complexity, where less complex cases—which take fewer turns to complete—are presented by customers expressing positive emotion. Including EMO_1 as a covariate rules out this explanation: the emotion in the first customer message controls for the initial emotion of the customer which can be related to the case complexity. We observe that the effect of lagged emotion EMO_{t-1} is very similar when including or excluding EMO_1 as a control variable, suggesting that the effect is not confounded by unobserved factors related to case

Table 4 Effect of Customer Emotion on the length of a conversation (Outliers Excluded, Both EMO_{t-1} and EMO_1 are Instrumented using $CustSent$)

	Model (4) $Nturns$	Model (5) $Pr(LastTurn)$	Model (5) $Pr(LastTurn)$
EMO_1	-1.684***		-0.033***
	(0.0689)		(0.0069)
EMO_{t-1}		0.571***	0.577***
		(0.0042)	(0.0044)
$Concurrent$ (chat level)	-1.237***		
	(0.0369)		
$Concurrent_{t-1}$		0.019***	0.019***
		(0.003)	(0.003)
$NumInQueue$ (chat level)	0.032***		
	(0.0047)		
$NumInQueue_{t-1}$		0.001	0.001
		(0.0005)	(0.0005)
$log(CustWords_1)$	-0.328***		
	(0.0211)		
$log(CustWords_{t-1})$		-0.016***	-0.015***
		(0.0018)	(0.0018)
$Turn_t$		-0.005***	-0.005***
		(0.0005)	(0.0005)
$IsWeekend$	-0.012	0.008	0.008
	(0.0415)	(0.0045)	(0.0045)
$SrvType$	6.192*	-0.277*	-0.278*
	(3.1458)	(0.1274)	(0.1275)
$ShiftTime$	Included	Included	Included
$HourOfDay$	Included	Included	Included
Agent Fixed Effect	Included	Included	Included
Constant	11.129***	-0.886***	-0.883***
	(1.6750)	(0.1454)	(0.1454)
Observations	141,654	518,437	518,437

Standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

complexity. The effect is also economically significant: changing EMO_{t-1} from -1 (negative) to 0 (neutral) increases the probability of ending the conversation from 0.08 to 0.2, on average.

Altogether, the results suggest direct and indirect paths through which customer emotion affects agent behavior. First, the agent spends more effort writing to customers with negative emotion (compared to neutral emotion), which increases RT. But this mechanism explains only a small fraction of the increase in RT. For agent messages of similar length, the results suggest that RT continues to be longer for customers with negative emotion relative to neutral and positive emotion. Moreover, conversations that start with more negative customer emotion tend to be longer. This effect persists through the conversation: in any turn during the conversation, the remaining extension of the case increases when the customer is expressing negative emotion.

5.2. Effect of Agent Behavior on Customer Emotions

Table 5 shows the results of the models with customer emotion EMO as the dependent variable (Models (6) and (7)). Recall this specification uses each message as a unit of analysis and includes fixed effects for the conversation, so the identification is based on variation across turns within a conversation. Two specifications were estimated, including different sets of covariates that measure distinct aspects of agent behavior. The specification reported in the first column corresponds to Model (6), which includes RT as the main covariate. Recall that this estimation is carried out using exogenous workload-related IVs ($Concurrent$ and $NumInQueue$) instrumenting RT, in order to remove the variation in RT that could be driven by the unobserved quality of the agent response (Section 5.3 discusses results of the estimation without IVs). The estimation suggests that doubling RT decreases customer emotion by 0.06, equivalent to less than 0.1 standard deviations, a relatively small effect.

Table 5 Effect of Agent Behavior on Customer Emotion (Outliers Excluded. $log(RT_{t-1})$ is Instrumented using $Concurrent_{t-1}$ and $NumInQueue_{t-1}$)

	Model (6) EMO	Model (7) EMO
$log(RT_{t-1})$	-0.062***	-0.427***
	(0.0145)	(0.0403)
$ConvStage_t$	0.896***	1.181***
	(0.0057)	(0.0079)
$log(NumWords_{t-1})$		0.200***
		(0.0185)
$Turn_t$		-0.016***
		(0.0003)
Conversation Fixed Effect	Included	Included
Constant	0.066	0.794***
	(0.0524)	(0.0943)
Observations	586,456	586,456

Standard errors in parentheses
$^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$

The second specification corresponds to Model (7), including RT, log($NumWords$) and the corresponding $Turn$ number as the main covariates and using the same IVs as in the previous specification to instrument RT. The results reveal that customer emotion becomes more positive for longer messages: doubling $NumWords$ increases EMO by 0.2. One interpretation of this result is that customers find longer agent messages to be more informative or a signal that the agent is paying attention to them, thereby improving their emotion. In addition, customer emotion tends to decrease for longer conversations: an increase by 10 turns (equal to the average number of turns in a conversation) reduces customer emotion by 0.16. Furthermore, controlling for these other measures of agent behavior reveals a larger effect of RT on customer emotion: doubling response time decreases EMO by 0.43, which is about half the standard deviation of the dependent variable.

Overall, the results suggest that customer emotion is affected by the different measures of agent performance-related behaviors, where the predominant effect is a negative effect of RT on customer emotion. The managerial implications of these results are discussed in Section 6.

5.3. Sensitivity Analysis and Alternative Specifications

We analyzed several alternative specifications of the models to verify the robustness of the empirical results, which are summarized in this section. All the result tables of these additional analyses are reported in the Online Appendix.

Models (1)–(4) in Tables 3 and 4 are estimated with 2SLS instrumenting EMO with an alternative sentiment measure, in order to mitigate attenuation bias due to measurement error. For robustness, the same specifications were estimated with Ordinary Least Squares (OLS), without instruments (Table 1, Online Appendix). The results reveal a negative effect of EMO on RT, which is statistically significant and smaller in magnitude compared to the estimations reported in Table 3 (coefficient is approx. -0.1 compared to -0.2). In Model (3) the effect of EMO on $NumWords$ is smaller in magnitude and not statistically significant. For Model (4), with $NTurns$ as dependent variable, the coefficient of EMO_1 changes from -1.684 to -0.250 (p-value < 0.001). Overall, these results are consistent with attenuation bias due to imprecise measurement of customer emotion, which can be corrected with the IV estimation proposed using an alternative sentiment measure as instrument.

Our analysis uses $SentiStrength$ as the main measure of customer emotion and uses $CustSent$ as an IV to correct for measurement error. An alternative approach is to combine both measures into one variable using factor analysis, where the first factor is used as a construct for customer emotion. Doing so yields similar results to those reported in Table 3 in terms of magnitude and statistical significance (Table 2, Online Appendix). One difference is that when factor analysis is used to measure the effect of emotion on number of words (Equation (3)) the effect is smaller in magnitude and significant at the 0.1 level. Consequently, the indirect effect of EMO on RT (via the number of words) is smaller in this specification, about half the magnitude compared to the original results. Models (1)–(4) were also estimated replacing EMO with the alternative $CustSent$ measure (see Table 13, Online Appendix). The results are similar to the main results of the paper.

In addition, we estimated Models (1)–(3) without log transformation to the dependent variable, and Model (4) with log transformation to the dependent variable. The results were similar in terms of the signs, magnitude and statistical significance (Table 3, Online Appendix).

Model (2) includes the number of words ($NumWords$), a proxy of agent effort, as a mediator. Another possible mediation is the emotion expressed in the agent's message, which we measured using the same sentiment analysis tools (see footnote 1). To check the robustness of our results, we included a second mediator—emotion in the agent's message. Including this variable in our analyses did not change our main results: the effect of EMO and $NumWords$ remained similar to those reported in Table 3 (see Table 4, Online Appendix), with a slightly smaller coefficient for EMO (drops from 0.2 to 0.16). The results suggest that the emotion in the agent's message is negatively related to RT. Emotions expressed by agents are highly influenced by organizational requirements regarding appropriate emotional displays. Additionally, agent expressed emotion is endogenous to agent RT, making it difficult to infer causality. Therefore, this analysis requires further investigation where agent expressed emotion is the main focus, and only presented here as a robustness test.

All the models related to agent behavior include $Concurrent$ as a control variable to account for the effects of multitasking. Our measure of concurrency is calculated based on number of simultaneous conversations assigned to an agent during the focal conversation. For robustness, we also estimated the models using two alternative definitions of concurrency; (i) the number of words and (ii) the number of messages written by the agent in parallel conversations. In all cases, concurrency has a positive effect on RT, corroborating that multitasking indeed increases the RT in a focal conversation (Table 5, Online Appendix). The effect of EMO on RT is similar to the main results (reported in Table 3) across all the specifications with alternative measures of concurrency. For our main analysis we preferred using the number of simultaneous conversations as a measure of concurrency because this is exogenous to the agent, whereas the number of words (or messages) written in parallel is endogenous.

The estimation of Model (1) is carried out using IVs and panel data, including fixed effects and assuming i.i.d. random errors. Examining the residuals of the model reveals serial correlation, and therefore the calculation of the standard errors may not be accurate. We estimated the same model clustering observations at the conversation level, which allows for arbitrary correlation within clusters. The standard errors were similar and the main conclusions do not change (Table 6, Online Appendix). In Table 3, the estimates of Models (2) and (3) use bootstrapped standard errors which account for correlation between the error terms within cluster.

Table 4 includes customer emotion as a linear predictor of the number of turns (Models (4) and (5)). For robustness, we also estimated the models including EMO in three levels, capturing positive, neutral and negative emotion (Table 7, Online Appendix). The results for Model (4) reveal a monotone non-linear effect of the emotion of the first message: taking neutral emotion as the base, positive emotion reduces $NTurns$ by 0.726 whereas negative emotion leads to an increase of 3.463 turns. In the survival Model (5), the coefficient associated to positive emotion in the previous message is positive and much larger in magnitude relative to negative emotion, consistent with a positive effect of emotion on the likelihood of ending the conversation (and thereby a shorter length of the conversation). The main conclusions remain when using a non-linear specification for the effect of customer emotion.

Recall that the results from Tables 3 and 4 are based on a sample without outliers. The same models were estimated with all the observations, including the outliers (Section 8, Online Appendix). Overall, the conclusions obtained from these results are similar. The coefficients associated to EMO in Models (1) and (2) continue to be negative and similar in magnitude. For Model (3), the effect of EMO on $NumWords$ is negative and significant, with a point estimate of -0.024. As discussed previously, the additional analysis reported in Table 3 showed that the effect of EMO on $NumWords$ appears to be non-linear, suggesting that the model with linear EMO is not well specified and less robust. This may explain why the EMO coefficient in the linear specification is sensitive to the definition of the sample (Table 8, Online Appendix). In Model (4), with $NTurns$ as dependent variable, the coefficient on EMO_1 , -1.677, is similar to the main analysis (Table 4).

Additional analysis was carried out to evaluate the robustness of Models (6) and (7) (main results reported in Table 5), with customer emotion used as the dependent variable. Recall that these models are estimated with IVs to address the endogeneity of RT, which is potentially correlated with unobservable factors associated to the quality of the response. The same models were estimated without IVs using OLS (Table 11, Online Appendix). The coefficient associated with $\log(RT)$ flips from negative to positive, with a point estimate close to 0.02 (with p-value < 0.001). This is consistent with the endogeneity bias that was conjectured: because RT is likely to be positively correlated with quality, which is unobservable, and is part of the error term. This generates a positive bias in the estimated coefficient of RT. Instrumenting RT with exogenous factors associated to agent workload helps to correct this bias. Models (6) and (7) were also estimated using OLS regression and replacing EMO with the alternative $CustSent$ measure (Table 12, Online Appendix). The results are similar to the main analysis.

Overall, the sensitivity analysis provides further support of the estimation results, showing that they are robust to alternative specifications.

6. Managerial Implications

The fact that customer emotion impacts agent behavior and that agent behavior impacts customer emotion suggests that emotional load should be monitored and taken into account in operational decisions. This section discusses some prescriptions for the design and control of service systems that are subject to the effects of emotional load. This is even more important considering findings that suggest that customer emotions reflect customer satisfaction (Yom-Tov et al. 2018, Ashtar et al. 2020) and the connection of the latter to organization profitability.

- **Performance goals, system design and staffing**

The results presented here should serve as a "call for awareness" that emotional load exists, varies within a service encounter, and impacts agent performance. A standard approach is to consider service time and case characteristics as the key dimensions of load, but our work suggests that customer emotion is another important factor. Dealing with more negative customers will require agents to spend more time to solve the customer issues and to cope with customers emotion. To evaluate the total effect of emotion, it is useful to compare the agent time required to handle angry (negative), neutral, and happy (positive) customers. To do that we define total throughput time by multiplying the average agent RT per turn by the average number of turns in a chat. The total throughput time required to handle a "negative" customer is 15.7 minutes (1.32 minutes times 11.9 turns); compared to 11.1 minutes for a neutral customer and 7.6 minutes to handle a "positive" one. This analysis suggests that the total amount of throughput time associated with a negative customer is 42% longer than the one associated with a neutral customer. Many contact centers measure agent performance by the number of calls an agent handles per hour (average concurrency divided by total throughput time). They should be aware that an agent can serve 12.6 neutral customers per hour but only 8.9 negative ones (assuming average concurrency = 2.33). Hence, the evaluation of service agents or teams who encounter a high proportion of negative customers (for example customer retention teams) should be based on adjusted targets of calls per hour. This is important if a contact center is considering a design change to incorporate skill-based routing (i.e. that each customer group is served by a separated agent-skill group).

Another way to think about the implication of emotional load is to consider how variations in customer mix impact the offered load, which is equivalent to the amount of staffing required to handle the arriving customer workload. We present in Table 6 a comparison between the offered load (calculated by the arrival rate, given in Appendix EC.3, multiplied by total throughput times divided by average concurrency) of the current mix of emotional messages in the contact center we analyzed vs. the offered load that the agents will need to handle if 10% of the messages transformed from being neutral to negative for some reason. Such a situation might arise after a company experiences failure in one of its products or services. This kind of change in the mix of customer emotion would increase the number of agents needed to handle customer issues by 4.4% and the amount of text written per day by 2.2%, assuming no change in arrival rate. This analysis suggest that customer emotion is an important factor that should be accounted for in staffing decisions.

- **Counterfactual analysis: the impact of emotional load on system-level performance**

Most organizations do not monitor customer emotions and do not adjust staffing to match variation in customer emotion. Here we would like to calculate the impact of an increase in emotional load on performance level, when staffing remains constant. We use the same scenario presented in

Table 6 Comparison of Working Hour Associated with Different Mix of Message Emotion

Emotion	Base Case			Alternative Case		
Type	Message percentage	Offered Load	Effort (NumWords)	Message percentage	Offered Load	Effort (NumWords)
Positive	29%	14.8	63,685	29%	14.8	63,685
Neutral	63%	47.1	162,652	53%	39.6	136,834
Negative	8%	8.5	25,103	18%	19.1	56,483
Total	100%	70.4	251,440	100%	73.7	257,002

Table 6. We use a simulation model that was calibrated to the operation of contact centers and developed by Castellanos et al. (2019). It is a version of an Erlang-A model that takes into account unique contact center features such as silent abandonment. For this counterfactual analysis, we estimated the simulation parameters using the data from February 2017. The simulation assumes that customers arrive to the contact center according to a non-homogeneous Poisson process with rate $\lambda_{d,t}$ ($\lambda_{d,t}$ is the arrival rate at day d and hour t). The customers are served by $n_{d,t}$ statistically identical agents, with average concurrency of c. Therefore, the number of service slots available at time (d,t) is $cn_{d,t}$. If there is no available slot, the customer waits in a First-Come-First-Serve queue. The customer has finite patience, assumed to be exponentially distributed with rate θ, which was estimated using the methodology developed in Yefenof et al. (2018) and that takes into account the fact that customer waiting time in chat systems is censored both from the right (by the offered wait) and from the left (by the fact that sometimes customers abandon the service without exiting the queue—they do not close the chat window but "disappear"). For more details about this silent abandonment phenomenon and its implication see Castellanos et al. (2019). In our data $\theta = 0.5$, and 70% of the customers indicate their abandonment in real time (30% abandon silently). Service times are assumed to be exponentially distributed with rate μ. Note that μ in this simulation is 1 divided by the throughput time of a conversation and equal to $\mu = 0.075$; in the counterfactual scenario with more negative customers, μ was adjusted so that the throughput time was 4.4% longer (as suggested by our empirical results).

The simulation predicts that the 10% change from neutral to negative messages increases the probability of abandonment by 2% and increases the expected waiting time by two minutes.

- **Routing policies designed to achieve load balancing or specialization**

Emotional load should also impact work allocation (routing) decisions. In our data, agents usually serve up to three (average 2.3) customers simultaneously. However, the load created by three negative customers differs dramatically from the load created by three positive customers. Specifically, dealing with three neutral customer messages is equivalent (in terms of workload) to dealing with only 2.5 negative messages or 3.7 positive messages. We suggest that like other measures of workload, emotional workload could be used in the design of dispatching rules commonly used in contact centers to *dynamically* adjust the workload of agents based on real-time assessments. The sentiment analysis tool used in this work allows for real-time monitoring of emotional load during service conversations. Previous research showed that there is a clear trend of sentiment within a conversation (Yom-Tov et al. 2018): customer sentiment usually starts negative, then becomes neutral and transforms to positive toward the end of the conversation. This positive trend is captured in our analysis by the variable *ConvStage* that monitors the conversation progress.

We suggest designing a routing policy that would *balance* both offered load and emotional load. The idea is that when a new conversation arrives it will be assigned to an agent that has the least overall load including offered, and emotional. Such a policy will dynamically allocate more capacity to agents that handle customers who consistently express negative emotion. This dynamic allocation can be based on Model (5), and would allow an agent to spend more time dealing with a negative customer which would be expected to improve customer emotion and overall customer satisfaction. This idea draws its intuition from Armony and Ward (2010) and Mandelbaum et al. (2012), who suggest adoption of allocating policy that is fair from the agent perspective (Carmeli et al. 2018).

In addition, in some contact centers customers write their inquiry before entering the queue (Castellanos et al. 2019). In such cases, we can assess the emotional load expected by that customer inquiry in real-time before assigning a new conversation to an agent and using the measure of emotional load we can also predict that customer needs. Model (4) supports the claim that this is indeed possible by showing that one can predict the number of turns that a chat will require using the customer sentiment of the first turn. This information can be used for designing a *"sentiment-based routing"* mechanism, analogous to skill-based routing. This routing mechanism could also assign an emotional call to the most appropriate agent group (e.g., customer retention team) trained to deal with particular customer emotion (e.g., anger).

- **Prioritization**

Our results show that longer agent RT hampers customer emotions. Therefore, operational policies that reduce agent RT will improve customer emotions. Such policies might be implemented in the following way: since agents handle multiple customers in parallel, they might miss expressions of negative customer emotion while they are interacting with other customers. Real-time monitoring can increase agents awareness by alerting them when an escalation in negative emotion occurs. For example sentiment engines can be designed to provide real-time monitoring of customer sentiment, and alert managers and agents of problematic situations (e.g., when the sentiment of a customer drops below a specific threshold). These alerts will enable agents to *prioritize* unsatisfied customers, reduce their RT, and improve customer satisfaction. Moreover, managers can use these alerts to identify extreme negative sentiment cases, and to provide agents with relevant assistance. This idea is now being implemented in some of the companies working with the LivePerson platform. In addition, given our finding that longer agent texts improve customer sentiment, agents should be made aware of the impact of message length and provided with indication of when customer sentiment is deteriorating, by alerts similar to those suggested above. In these cases agents should be trained to react by communicating their effort better in order to improve customer emotion.

7. Conclusions and Future Research

Our findings show that customer emotion impacts agent performance-related behaviors: agents respond more slowly and write more words to customers who express negative emotion, compared to positive or neutral emotion, supporting Hypotheses 1 and 2, respectively. Negative customer emotion increases agent RT directly and indirectly through agent effort (supporting Hypothesis 3). Most of the effect of customer emotion, however, is direct (see Table 3). This suggests that there may be additional mechanisms through which customer emotion impacts agent RT. One option is that our proxy of effort captures only a portion of agent effort. Future research should include data about other agent activities to fully understand the role of agent effort. Another option is that agents prioritize customers depending on their expressed emotions. For example, recent findings suggest that decision-makers' perceptions impact patient prioritization in Emergency Departments (Ding et al. 2019). We call for future research to continue this line of work to understand how service agents prioritize concurrent customers and whether emotional load impacts prioritization.

In addition, we showed that negative customer emotion prolongs the service interaction, supporting Hypothesis 4. This effect is large, and one possible mechanism may be agent errors (Rafaeli et al. 2012): when agents encounter expressions of negative emotion, they are more likely to make mistakes, extending the service encounter as a result. We cannot test this mechanism in the current dataset because we cannot automatically code agent errors in the data. We hope that future advancements in the field of Natural Language Processing will help researchers in pursuing this direction.

Overall, our findings suggest that negative customer emotions create a burden on agents, and that positive customer emotions may act as a source of motivation (Bakker and Demerouti 2007). The estimations suggest that the effects of emotional load are substantial, and comparable to workload factors. For example, a negative customer message increases agent RT in the subsequent message by 4.2% similar to adding another customer to an agent which increases RT by 7.4% (see Table 3). Our findings also show that the opposite direction of influence exists—an increase in agent RT or an increase in the number of turns, hampers customer emotion (supporting Hypotheses 5 and 7, respectively). An increase in agent effort, however, leads to greater positive customer emotion as predicted (Hypothesis 6).

Previous research relied almost solely on experimental manipulations with small samples and low-resolution self reported emotions, thus affording limited managerial insights. The current paper overcomes these limitations by using operational and objective measures of agent behavior and of customer emotion in real service conversations, measured at the resolution of a single message. We show that emotional load creates "micro-level influences," that occur at the level of a single message within the conversation between an agent and a customer. We theorize and show empirical effects of emotional load that goes beyond multitasking and queue length effects. Our analyses of a large data-set of conversations between agents and customers, empirically measure this type of load, and document its influence on critical Operations Management parameters including agent RT, agent effort, and the number of turns it takes to complete a conversation.

We introduce a new position for customer emotion in service—that of a potential source of load. This is in contrast to traditional Operations Management views, where customer emotion is treated as an outcome. The implicit assumption of past research was that customer happiness depends on their evaluation of the quality of service. We show, however, that customer emotions may also be a factor that determines the efficiency of the service. This suggests that the concept of load actually comprises multiple aspects, and that emotional load is one of them. This view of load accounts for pressures beyond the mere presence of a customer, and is based on factors inherent to the nature and content of individual service conversations.

Emotions provide information (data) about a social situation and the actors in it (van Kleef 2015). To date, these data have served only the service dyad: an agent and a customer. This dyad is engaged in co-production of value; both actors invest effort to resolve a specific issue. The ratio of the effort between the service interaction partners is dependent on context. For example, if a customer requests easy-to-get information, the ratio of effort will be close to 1. In contrast, if a customer has a complicated request, or if the customer creates high emotional load, the agent will likely need to invest more effort than the customer. As Roels (2014) showed, one can improve service system efficiency by considering the effort ratio and route customers to adequate service channels based on it. We therefore call for researchers and practitioners to view customer emotion as data that can aid them in designing service systems.

The type of data we use in the current paper is increasingly available in service organizations (i.e., full documentation of service). We highlight the opportunities that such data, coupled with automated sentiment analysis tools create for studying service delivery (Rafaeli et al. 2017, 2019) and improving the operations of contact centers. From a managerial perspective, our analyses suggest the importance of incorporating real time monitoring of the emotions of customers being handled by service agents. Beyond the technical count of the number of customers in the service system, service operations need to acknowledge the dynamics that customers bring to the system. This includes the types of problems that customers raise, the verbosity with which they communicate their problems, and the emotions that they attach to this communication. Failing to recognize such customer-induced states can lead to inaccurate planning models and sub-optimal service management.

References

Allon G, Federgruen A, Pierson M (2011) How much is a reduction of your customers' wait worth? An empirical study of the fast-food drive-thru industry based on structural estimation methods. *Manufacturing & Service Operations Management* 13(4):489–507.

Anand KS, Paç MF, Veeraraghavan S (2011) Quality—speed conundrum: Trade-offs in customer-intensive services. *Management Science* 57(1):40–56.

Armony M, Ward A (2010) Fair dynamic routing in large-scale heterogeneous-server systems. *Operations Research* 58(3):624–637.

Ashtar S, Rafaeli A, Yom-Tob GB (2020) High-peak, low-peak, end and trend: The influence of features of emotions expressed by customers and employees on customer evaluations of frontline service, working paper.

Bakker AB, Demerouti E (2007) The job demands-resources model: State of the art. *Journal of Managerial Psychology* 22(3):309–328.

Beal DJ, Weiss HM, Barros E, MacDermid SM (2005) An episodic process model of affective influences on performance. *The Journal of Applied Psychology* 90(6):1054–1068.

Bray RL, Coviello D, Ichino A, Persico N (2016) Multitasking, multiarmed bandits, and the Italian judiciary. *Manufacturing & Service Operations Management* 18(4):545–558.

Carmeli N, Yom-Tov GB, Mandelbaum A (2018) Multi-dimensional load: Balancing emotional and system load in service systems, working paper.

Casado Diaz AB, Más Ruíz FJ (2002) The consumer's reaction to delays in service. *International Journal of Service Industry Management* 13(2):118–140.

Castellanos A, Yom-Tov GB, Goldberg Y (2019) Silent abandonment in contact centers: Estimating customer patience with uncertain data, working paper.

Cho DD, Bretthauer KM, Cattani KD, Mills AF (2019) Behavior aware service staffing. *Production and Operations Management* 28(5):1285–1304.

Dai H, Milkman KL, Hofmann DA, Staats BR (2015) The impact of time at work and time off from work on rule compliance: The case of hand hygiene in healthcare. *Journal of Applied Psychology* 100(3):846–862.

Delasay M, Ingolfsson A, Kolfal B, Schultz K (2019) Load effect on service times. *European Journal of Operational Research* 279(3):673–686.

Ding Y, Park E, Nagarajan M, Grafstein E (2019) Patient prioritization in emergency department triage systems: An empirical study of the canadian triage and acuity scale (CTAS). *Manufacturing & Service Operations Management* 21(4):723–741.

Donaldson SI, Grant-Vallone EJ (2002) Understanding self-report bias in organizational behavior research. *Journal of Business and Psychology* 17(2):245–260.

Field JM, Victorino L, Buell RW, Dixon MJ, Meyer Goldstein S, Menor LJ, Pullman ME, Roth AV, Secchi E, Zhang JJ (2018) Service operations: What's next? *Journal of Service Management* 29(1):55–97.

Fredrickson BL, Kahneman D (1993) Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality and Social Psychology* 65(1):45.

Gabriel AS, Diefendorff JM (2015) Emotional labor dynamics: A momentary approach. *Academy of Management Journal* 58(6):1804–1825.

Geddes D, Callister RR (2007) Crossing the line(s): A dual threshold model of anger in organizations. *Academy of Management Review* 32(3):721–746.

Goes PB, Ilk N, Lin M, Zhao JL (2018) When more is less: Field evidence on unintended consequences of multitasking. *Management Science* 64(7):2973–3468.

Goldberg LS, Grandey AA (2007) Display rules versus display autonomy: Emotion regulation, emotional exhaustion, and task performance in a call center simulation. *Journal of Occupational Health Psychology* 12(3):301–318.

Grandey AA, Dickter DN, Sin HP (2004) The customer is not always right: Customer aggression and emotion regulation of service employees. *Journal of Organizational Behavior* 25(3):397–418.

Grandey AA, Rafaeli A, Ravid S, Wirtz J, Steiner DD (2010) Emotion display rules at work in the global service economy: The special case of the customer. *Journal of Service Management* 21(3):388–412.

Groth M, Hennig-Thurau T, Walsh G (2009) Customer reactions to emotional labor: The roles of employee acting strategies and customer detection accuracy. *Academy of Management Journal* 52(5):958–974.

Hareli S, Rafaeli A (2008) Emotion cycles: On the social influence of emotion in organizations. *Research in Organizational Behavior* 28:35–59.

Hayes AF (2018) *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (Guilford publications).

Hayes AF, Rockwood NJ (2020) Conditional process analysis: Concepts, computation, and advances in the modeling of the contingencies of mechanisms. *American Behavioral Scientist* 64(1):19–54.

Katz K, Larson B, Larson R (1991) Prescription for the waiting in line blues: Entertain, enlighten and engage. *Sloan Management Review* 32:44–53.

Kc DS (2013) Does multitasking improve performance? Evidence from the emergency department. *Manufacturing & Service Operations Management* 16(2):168–183.

Kc DS, Terwiesch C (2009) Impact of workload on service time and patient safety: An econometric analysis of hospital operations. *Management Science* 55(9):1486–1498.

Knutson B (1996) Facial expressions of emotion influence interpersonal trait inferences. *Journal of Nonverbal Behavior* 20(3):165–182.

Larson R (1987) Perspective on queues: Social justice and the psychology of queueing. *Operations Research* 35(6):895–905.

Maister D (1984) *The psychology of waiting lines* (Harvard Business School).

Mandelbaum A, Momcilovic P, Tseytlin Y (2012) On fair routing from emergency departments to hospital wards: QED queues with heterogeneous servers. *Management Science* 58(7):1273–1291.

Mandelbaum A, Zeltyn S (2013) Data stories about (im)patient customers in tele-queues. *Queueing Systems* 75(2–4):115–146.

Manski CF (1993) Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies* 60(3):531–542.

Mohr LA, Bitner MJ (1995) The role of employee effort in satisfaction with service transactions. *Journal of Business Research* 32(3):239–252.

Porath CL, Erez A (2007) Does rudeness really matter? The effects of rudeness on task performance and helpfulness. *Academy of Management Journal* 50(5):1181–1197.

Rafaeli A, Altman D, Gremler DD, Huang MH, Grewal D, Iyer B, Parsuraman A, Ruyter Kd (2017) The future of frontline research: Invited commentaries. *Journal of Service Research* 20(1):90–91.

Rafaeli A, Ashtar S, Altman D (2019) Digital traces: New data, resources, and tools for psychological-science research. *Current Directions in Psychological Science* 28(6):560–566.

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

Rafaeli A, Sutton RI (1987) Expression of emotion as part of the work role. *Academy of Management review* 12(1):23–37.

Roels G (2014) Optimal design of coproductive services: Interaction and work allocation. *Manufacturing & Service Operations Management* 16(4):578–594.

Socher R, Perelygin A, Wu J, Chuang J, Manning CD, Ng A, Potts C (2013) Recursive deep models for semantic compositionality over a sentiment treebank. *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 1631–1642 (Seattle, Washington, USA: Association for Computational Linguistics).

Song H, Tucker AL, Murrell KL (2015) The diseconomies of queue pooling: An empirical investigation of emergency department length of stay. *Management Science* 61(12):3032–3053.

Sutton RI, Rafaeli A (1988) Untangling the relationship between displayed emotions and organizational sales: The case of convenience stores. *Academy of Management Journal* 31(3):461–487.

Tan TF, Netessine S (2014) When does the devil make work? An empirical study of the impact of workload on worker productivity. *Management Science* 60(6):1574–1593.

Tausczik YR, Pennebaker JW (2010) The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology* 29(1):24–54.

Taylor S (1994) Waiting for service: The relationship between delays and evaluations of service. *The Journal of Marketing* 58(2):56–69.

Thelwall M (2013) Heart and soul: Sentiment strength detection in the social web with sentistrength. *Proceedings of the CyberEmotions* 5:1–14.

Tiedens LZ (2001) Anger and advancement versus sadness and subjugation: The effect of negative emotion expressions on social status conferral. *Journal of Personality and Social Psychology* 80(1):86.

van Kleef GA (2015) Social emotions life regulate the emotions as social information (EASI) model. *Current Directions in Psychological Science* 18(3):184–188.

van Kleef GA, De Dreu CKW, Manstead ASR (2004) The interpersonal effects of anger and happiness in negotiations. *Journal of Personality and Social Psychology* 86(1):57–76.

Walker DD, van Jaarsveld DD, Skarlicki DP (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.

Wang J, Zhou YP (2016) Impact of queue configuration on service time: Evidence from a supermarket. *Management Science* 64(7):3055–3075.

Wang M, Liao H, Zhan Y, Shi J (2011) Daily customer mistreatment and employee sabotage against customers: Examining emotion and resource perspectives. *Academy of Management Journal* 54(2):312–334.

Wansbeek TJ, Meijer E (2000) *Measurement error and latent variables in econometrics*, volume 37 (North-Holland).

Weiss HM, Cropanzano R (1996) Affective events theory: A theoretical discussion of the structure, causes and consequences of affective experiences at work. *Research in Organizational Behavior* 18(1):1–74.

Yefenof J, Goldberg Y, Wiler J, Mandelbaum A, Ritov Y (2018) Self-reporting and screening: Data with current-status and censored observations, working paper.

Yom-Tov GB, Ashtar S, Altman D, Natapov M, Barkay N, Westphal M, Rafaeli A (2018) Customer sentiment in web-based service interactions: Automated analyses and new insights. *In WWW '18 Companion: The 2018 Web Conference Companion, April 23–27*, 8 pages (New York, NY, USA: ACM).

This page is intentionally blank. Proper e-companion title page, with INFORMS branding and exact metadata of the main paper, will be produced by the INFORMS office when the issue is being assembled.

EC.1. Precision and Recall of Sentiment Analysis Tools (Yom-Tov et al. 2018)

Table EC.1 Comparing four models in detecting emotion in customer messages.

Emotion class	Model	Precision	Recall	F_1	$F_{0.5}$
Negative	SentiStrength	0.494	0.216	0.300	0.393
	CustSent	0.719	0.236	0.355	0.51
	Stanford	0.335	0.509	0.404	0.36
	LIWC	0.479	0.115	0.186	0.294
Positive	SentiStrength	0.813	0.677	0.739	0.781
	CustSent	0.866	0.569	0.687	0.784
	Stanford	0.546	0.339	0.418	0.486
	LIWC	0.491	0.717	0.583	0.524

EC.2. Correlation tables

Table EC.2 Pairwise Pearson Correlation: Message Level

		1	2	3	4	5	6	7	8	9	10	11
1	EMO ($SentiStrength$)	1.00										
2	$CustSent$	0.61‡	1.00									
3	RT	-0.04‡	-0.04‡	1.00								
4	$log(RT)$	-0.06‡	-0.07‡	0.84‡	1.00							
5	$NumWords$	0.07‡	0.09‡	0.46‡	0.40‡	1.00						
6	$log(NumWords)$	0.06‡	0.08‡	0.39‡	0.42‡	0.89‡	1.00					
7	$Concurrent$	0.00	0.01‡	-0.00†	0.03‡	-0.05‡	-0.03‡	1.00				
8	$NumInQueue$	0.01‡	0.01‡	0.01‡	0.01‡	0.01‡	0.01‡	0.14‡	1.00			
9	$ConvStage$	0.25‡	0.33‡	0.12‡	0.07‡	0.22‡	0.21‡	-0.05‡	0.02‡	1.00		
10	$ShiftTime$	0.00	0.00	-0.01‡	-0.01‡	-0.01‡	-0.01‡	-0.02‡	0.01‡	0.02‡	1.00	
11	$Turn$	0.09‡	0.12‡	-0.03‡	-0.05‡	0.01‡	0.01‡	-0.19‡	-0.02‡	0.42‡	0.05‡	1.00

Note: †$p < 0.05$, ‡$p < 0.01$.

Table EC.3 Pairwise Pearson Correlation: Conversation Level

		1	2	3	4	5	6	7	8	9	10	11	12	13
1	$NTurns$	1.00												
2	$log(NTurns)$	0.89‡	1.00											
3	EMO_1^a	-0.02‡	-0.01‡	1.00										
4	$CustSent_1$	-0.06‡	-0.06‡	0.43‡	1.00									
5	RT	-0.05‡	-0.06‡	-0.02‡	-0.04‡	1.00								
6	$log(RT)$	0.05†	0.08‡	-0.02‡	-0.05‡	0.94‡	1.00							
7	$NumWords$	-0.11‡	-0.10‡	0.01‡	0.01†	0.29‡	0.28‡	1.00						
8	$log(NumWords)$	-0.05‡	-0.01‡	0.01†	0.00	0.27‡	0.29‡	0.96‡	1.00					
9	$Concurrent$	-0.11‡	-0.08‡	0.01‡	0.01‡	0.04‡	0.05‡	-0.02‡	-0.02‡	1.00				
10	$NumInQueue$	-0.03‡	-0.03‡	0.01‡	0.00	0.01‡	0.01‡	0.02‡	0.02‡	0.09‡	1.00			
11	$CustWords_1$	-0.02‡	-0.01‡	0.00	-0.08‡	0.10‡	0.12‡	0.13‡	0.13‡	0.06‡	0.05‡	1.00		
12	$log(CustWords_1)$	-0.06‡	-0.05‡	0.00	-0.07‡	0.10‡	0.13‡	0.14‡	0.14‡	0.06‡	0.04‡	0.82‡	1.00	
13	$ShiftTime$	-0.00	-0.00	-0.01‡	-0.01‡	-0.02‡	-0.02‡	-0.02‡	-0.02‡	0.00	0.01‡	-0.01†	-0.01†	1.00

Note: †$p < 0.05$, ‡$p < 0.01$, $^a SentiStrength$.

EC.3. Arrival Rate

In this section we provide additional data needed for the calculations of the impact of our results on offered load. Table EC.4 provide a typical pattern of customer arrival rate per working hour in our data.

Table EC.4 Arrival Rate during a Working Day

Hour	λ
9	62.45
10	67.6
11	63.75
12	67.25
13	68.9
14	69.05
15	86.25
16	90.2
17	68.1
18	63.5
19	59.35
20	65.5
21	57.95
22	53.65