Opportunities, Tools and New Insights:
Evidence on Emotions in Service from Analyses of Digital Traces Data

Anat Rafaeli
Galit Yom Tov
Shelly Ashtar
Daniel Altman
Technion – Israel Institute of Technology

Chapter forthcoming in
Research on Emotions in Organizations. Volume 16
Theme of volume "Emotions and Service in the Digital Age".

Editor: Charmine E. J. Härtel

January 2020

Table of Contents

Digital Age Opportunities, Tools and New Insights: Evidence of Emotions in Service from Analyses of Digital Traces Data ................................................................. Error! Bookmark not defined.

Opening: Digital traces organic data: New and unique research opportunities ... Error! Bookmark not defined.

Terms, Resources and Tools for Digital Service Research about Emotion Error! Bookmark not defined.


1. The magnitude of the data Error! Bookmark not defined.
2. The nature of variables available in digital traces data. Error! Bookmark not defined.


Available Data Error! Bookmark not defined.

Figure 1. Number of words in customer messages and number of customer messages in service interactions, in a sample of customer interactions [Airline, n= 25,714] Error! Bookmark not defined.

New insights from analyses of digital traces of emotions in customer service conversations Error! Bookmark not defined.
Figure 1. Frequencies of expressed positive and negative emotion in customer messages [Retail data, n=3,659,053] ................................................................. Error! Bookmark not defined.

Figure 2. Frequencies of emotion in service conversations [Retail, n =439,585] ... Error! Bookmark not defined.

Figure 3. Distribution of sum of positive and negative customer emotions in full conversations [Airline, n= 21,122] ........................................................................................................ Error! Bookmark not defined.

b. What customer emotions do service agents really encounter in their service delivery work?

Figure 4. Customer emotion in first line and last line of a conversation with a service agent [Airline, n= 25,668] ................................................................. Error! Bookmark not defined.

Figure 5. Customer emotion expressed in different sections of service conversations, at different hours of the day and different days of the week [Telecommunication, n= 390,438] .......... Error! Bookmark not defined.

c. What customer emotions do service agents encounter over the course of a work shift? .. Error! Bookmark not defined.

Figure 6. Customer emotion encountered by one employee during a workday [within agent analysis, n= 124 conversations] ................................................................. Error! Bookmark not defined.

Figure 7. Total (sum of all) positive emotions and negative emotion that all customers express over the course of a workday [Airline, n= 11,591] ....................... Error! Bookmark not defined.

d. Relating customer expression of emotions to customer evaluations of the service conversation.......................................................................................... Error! Bookmark not defined.

Figure 8. Customers’ expressed emotions during service conversation by level of customer satisfaction with the performance of the service agent after the service conversation had ended [Airline, n= 6,973]..................................................................................... Error! Bookmark not defined.

Figure 9. Customers expressed emotions during service conversation by level of customer overall satisfaction with the service agent as measured by the Net Promoter Score (NPS) [retail, n= 123,554] ........................................................................................................ Error! Bookmark not defined.

Figure 10. Customer emotion in sections of interaction by customer evaluations of whether their service issue had been resolved [Telecommunication, n=390,438]. .. Error! Bookmark not defined.

e. Customer emotions in different industries?................................. Error! Bookmark not defined.

Figure 11. Customer emotion in service conversations in firms from three industries ........ Error! Bookmark not defined.

f. Relating customer expression of emotion to service agents’ behaviors...... Error! Bookmark not defined.

Figure 12. Relationship between emotion in customer messages and employee response time (RT) in subsequent message. (Airline, n= 1,447,070 customer messages in 208,210 service conversations) ..................................................................................... Error! Bookmark not defined.

Discussion................................................................................................................. Error! Bookmark not defined.

Toward Future Research ..................................................................................... Error! Bookmark not defined.

References ............................................................................................................. Error! Bookmark not defined.
Opportunities, Tools and New Insights:
Evidence on Emotions in Service from Analyses of Digital Traces Data

Abstract

No one can live and work in the 21st century without digital service interactions. We use archival traces of digital service interactions to examine and document the experience of emotion in service conversations between customer and service agents. We use automated sentiment analyses to decipher objectively what happens in terms of customer and employee expressions of emotions during service interaction. Our data also show the research benefits of digital age technologies, which provide a gold mine of archives of service conversations and unique opportunities to promote our understanding of emotions in service delivery. Our analyses show what emotions customers or service agents feel or express, finding expression of much more positive emotion than currently recognized. We also show how emotions evolve through service conversations, showing that this evolution can predict customer satisfaction. We further show the emotions that service agents encounter during their work shifts, which helps explain the high level of employee burnout. And we test the reciprocal emotional influence between customers and service employees. Our analyses give new insights into emotions in service delivery and into new opportunities afforded by digital service platforms.

Digital traces organic data: New and unique research opportunities

Can anyone live and work in the 21st century without digital service interactions? We buy through Amazon or Ali Express, book flights and hotels through Expedia or Booking, and communicate with service agents through chats, texts, Facebook, Twitter and email. Emotion in service is equally ubiquitous – it is rare to talk about service without having someone intervene to recount a frustrating or annoying service situation.

At the same time, the experience of emotion in service is still very abstract. What exactly happens in a service interaction? What emotions do customers or service agents feel or express? How do emotions evolve through service conversations? What customer emotions do service agents encounter during their work shifts? And how do these emotions affect customers and agents?

The prevalence of service in modern life has been accompanied by increasing research attention to emotion in service. Available research of emotions in service has relied primarily on self-report data (e.g., Grandey, Dickter, & Sin, 2004; Groth & Grandey, 2012), qualitative explorations or field work based on observations (e.g., Pugh, 2001), and experimental manipulations (Cheshin, Amit, & Van Kleef, 2018; Goldberg & Grandey, 2007; Rafaeli et al., 2012). Now digital age 21st century technologies afford new sources of data, and new approaches to data collection and analyses, and provide fascinating opportunities for new insights.

For a long time, service has been conducted through telephone call centers, where “calls are recorded for quality assurance” (and perhaps for legal reasons). Such recordings offered invaluable access to actual communication between agents and customers. However, utilizing this resource traditionally relied on the labor-intensive process of transcribing the conversations and manually coding
key themes (e.g., Gabriel & Diefendorff, 2015; Rafaeli, Ziklik, & Doucet, 2008). But increasingly, modern-day technologies afford tools for automatic recording and retrieval of the full data comprising service interactions. Additionally, traditional service media (face-to-face, telephone) are increasingly being replaced with sophisticated technology-mediated encounters. One such development is services delivered through written messages (chats, texting, twitter) Communication can be through corporate websites, Twitter or Facebook, or through mediators, such as http://LivePerson.com, a company that sells other firms tools for text-based service communication between customers and service agents.

From a research perspective, these digital age technologies provide a gold mine of archives of service conversations and unique opportunities to promote our understanding of emotions in service delivery.

Digital age service delivery also provides the opportunity to access direct and accurate measures of meta-data about service conversations. Not only are the full content of the service conversations accessible, they can also be matched with when the conversation occurs, how long it lasted, what else happened before or after the conversation, and more. Exploiting the full potential of this data requires tools and methodologies mostly foreign to OB researchers. Making the most of the data can be achieved through collaborations with Computer Science, Operations Research, and Data Science colleagues. Such interdisciplinary work is still rare in the Organizational Behavior community, and the goal of this chapter is to encourage researchers to pursue such collaborations, and to show how this can expand the scope of research on emotion in service delivery. In this spirit, we next briefly review the foundations of the new form of research that we advance. Then we describe the context and tool of research that we have conducted, followed by new insights about emotion in customer service that our analyses unveiled. Our chapter only begins to show some possibilities afforded by analyses of archives of digital service conversations.

Terms, Resources and Tools for Digital Service Research about Emotion

A key feature of digital service research is the use of raw data generated in genuine service conversations between agents and customers. This means that the research does not begin with a collection of data based on a pre-defined research design but rather the relevant data must be extracted with special and appropriate tools. Service conversations are archived by the digital platforms through which they occur and can be retrieved and analyzed using relevant automated tools. We have referred to these data as Digital Traces (Rafaeli, Ashtar, & Altman, 2019), because the data trace actual, spontaneous and genuine behaviors of service agents and customers. The data are continuously documented by digital devices (e.g., corporate servers, social media platforms (e.g., Twitter)). Digital traces of the expressions and behaviors of customers and service agents are naturally occurring data points, that can be used to study various research questions and it is the researchers challenge (and privilege) to identify and construct research variables from available data. Others have labeled such data as Organic Data (Xu, Zhang, & Zhou, 2019), which we find a very useful term, because it emphasizes that the data are generated naturally as behaviors occur. In the study of service, organic data depict verbal and non-verbal expressions as conversations unfold. Organic data document people’s spontaneous behavior, with no intervention and potential bias due to researchers’ predictions or planned research design. Importantly, organic data can be extracted without too much technical training in Computer Science (Harlow & Oswald, 2016).

Digital traces of various elements of human behavior are already analyzed in the emerging trend called “Computational Social Science” (Alvarez, 2016). Mainstream social scientists are gradually joining
this trend, by doing and publishing social science research using data extracted from social media, mobile phones and other digital marvels (e.g., Dai, Milkman, Hofmann, & Staats, 2015; Jones & Silver, 2019; Salganik, 2017; Staats, Dai, Hofmann, & Milkman, 2016). There is also some penetration of somewhat similar sources of data referred to as Online Panel Data into management research (Porter, Outlaw, Gale, & Cho, 2019; Walter, Seibert, Goering, & O’Boyle, 2019). Our goal of this chapter is to attract researchers of emotion in service to join this trend by informing and exciting them about the potential horizons it opens.

Computational Social Science research has touched on the study of emotion. Groups of Computer science and information systems researchers have been—for over ten years now—attempting to develop tools for automatic detection and identification of the presence of emotion, and various features of emotion in human expression and behavior (e.g., Pang & Lee, 2008; Wilson, Wiebe, & Hoffmann, 2005). For whatever reason, a lot of this work speaks about “sentiment” rather than “emotion,” but the growth in the number of published papers in Computer Science, Information Systems and Robotics that study “sentiment” or “emotion” is exponential (see Rafaeli et al., 2019). Importantly, this growth is not boosted by research done by psychologists or other social scientists, but rather by computer scientists who attempt to study psychology or social science questions. For example, Paltoglou et al. (2010) describe an analysis of emotions in informal textual communication in cyberspace. Thelwall (2013) describes detection of sentiment strength in the social web. Settanni and Marengo (2015) describe emotion expressed in Facebook posts. Lucas et al. (2017) go as far as claiming to study emotional dynamics, and Waddell (2016) describes algorithms that can tell managers how their employees are feeling. The question that we as psychologists and organizational behavior researchers must ask ourselves is whether such research makes us irrelevant. We reject such a claim, because we have a rich understanding of the theory of emotion, and especially the nuanced complexities of emotion in service. Thus, on the contrary, we call here for researchers of emotion in service to join the digital age by integrating digital age data and tools into our research.

Some Computational Social Scientists study customer service. Misopoulos et al. (2014) describe using Twitter archives to identify customer service experiences with the airline industry. Herzig et al. (2016) analyzed Twitter interactions to identify specific emotions (e.g., anger, frustration) in service conversations of employees and customers. In this vein, Hu et al. (2018) describe creation of a tone-aware chat-bot for interactions with customer requests on social media. Balducci and Marinova (2018) refer to analyses of “Unstructured Data (UD)” using novel technologies for promoting theoretical developments in customer service (see also Abney, Pelletier, Ford, & Horky, 2017; Misopoulos et al., 2014). Psychological literature is beginning to report research using such data (e.g., Jones & Silver, 2019; Jones, Wojcik, Sweeting, & Silver, 2016). We have previously encouraged (Rafaeli et al., 2017, 2019), and call here again for more consideration and integration of these new data sources and tools into research on emotion in customer service.

A pivotal development for computational research on emotion, is the emergence of tools enabling automated analyses of emotion. The collective term Sentiment Analysis refers to an amalgam of methods which allow automated detection of emotion in text and speech (e.g., Cambria, Das, Bandyopadhyay, & Feraco, 2017). One tool recognized by researchers in psychology and in organizational behavior is the Linguistic Inquiry and Word Count tool, known as LIWC (Tausczik & Pennebaker, 2010). Less known to social science researchers is SentiStrength (Thelwall, 2013), which—like the LIWC—is relatively inexpensive and easily implemented by any determined social scientist or aspiring graduate student. A more sophisticated tool – which is continuously updated to include state of
the art information systems developments is technically named the “Recursive Neural Tensor Network” (RNTN) model, and more popularly recognized as “the Stanford Tool” (Socher et al., 2013).

In the remainder of this chapter we report on our own research of emotion in service delivery, research that was enabled by automated sentiment analysis of archives of service conversations. We briefly describe the context and tool of our research, as a prelude to sharing new insights about emotion in service delivery that these analyses avail.

### Digital Analyses of Emotion in Customer Service

Our empirical work analyzed organic digital traces data of service conversations conducted through written chats. We obtained the data from a firm that maintains a platform for text-based interactions between customers and brands. LivePerson (LivePerson (http://LivePerson.com), serves 18,000 business customers, who communicate with their customers through chat. The LivePerson platform facilitates 25 million service conversations a month, accumulating archives of organic data. Our research started with development of a tool for analyzing emotions in these data, and then used this tool to obtain unobtrusive insights about emotions in service conversations.

Available sentiment analysis tools have been developed and validated with specific types of verbal data, mostly various types of reviews, such as movie, restaurant or hotel reviews. The tools have not been tested or validated for the study of emotion in customer service. This limitation is important because the unique features of service conversations require adapting available tools to the study of emotion in customer service.

A first challenge our research therefore embraced was testing and adapting previously defined rules and guidelines for conducting sentiment analysis of service conversations. Our collaboration began with us as researchers guiding the LivePerson Data Science team in how to create a protocol for detection and quantification of emotion in service text conversations, and how to assess the validity of the emotion detection protocol. This validation process, as described in greater detail in Yom-Tov et al. (2018), began by testing the accuracy of state-of-the-art sentiment analysis tools – The Stanford tool (Socher et al., 2013), SentiStrength (Thelwall, 2013), and LIWC (Tausczik & Pennebaker, 2010) for identification of emotions in customer chats. These tests found an unacceptably low level of precision.

This low level of validity led us to examine the unique features of text-based conversations. We found that data used for developing and validating sentiment analysis tools --movie and other reviews-- typically comprise unambiguous, straightforward opinions, and logical and well thought out text. In contrast, customer service conversations are mostly made up of short sentences, often not adhering to proper spelling or grammar, and containing ample slang, typos, and spelling mistakes. Service conversations also contain obscenities and extensive use of punctuation, symbols, emoticons and capitalization, features that express various aspects of emotion. Some work on identification of emotion in Twitter texts attempts to consider such features (e.g., Lucas et al., 2017), but little work has examined emotion expressed by customers or agents in a service setting.

We therefore proceeded to adapt the available tools to the unique nature of customer service texts (see Yom-Tov et al., 2018). In brief, the adaptation included revisions of the lexicon of the emotional value of words and terms that are used (Taboada, Brooke, Tofiloski, Voll, & Stede, 2011), addition of non-verbal icons and images such as 😊😊, and some added features taken from Natural Language Processing (NLP), like those used in other models of emotion analysis (e.g., Buechel & Hahn,
These efforts were fruitful and we validated performance of the tool across multiple types of customers and industries.

Once developed, the tool could be used to automatically analyze large samples of customer service conversations in different types of brands, and with different types of research foci. As elaborated next these analyses are unique in four ways: (1) they rely on analyses of large samples of actual expressions of customers, suggesting high external validity; (2) they are done automatically, with no human intervention, so offer high reliability, and minimal biases due to human error; (3) they provide access to new variables and analyses that previous research on emotion in customer service could not access without a major investment of time and effort; (4) they provide data on different firms, time periods, service agents and customers, allowing for comparisons and insights at a level of granularity that exceeds anything done in previous research.

Studying Emotion in Customer Service by Analyzing Chat Based Service Conversations

1. The magnitude of the data

A large magnitude of data is the first benefit of the new data and analyses that we promote. For example, the analyses described below are based on data retrieved from an archive comprising 1.14 million service conversations (or some 14 million text messages). Moreover, the data is obtained from multiple firms, and represents multiple service agents and customers who conversed at different times. Specific studies necessarily use samples from this enormous data set. These samples completely overshadow typical data sets used in prevailing research of emotion in customer service, where samples are often a few hundred at best. Importantly additional samples of data can be retrieved easily for comparison or to replicate as a test of the robustness of any given finding.

The organic nature of the data frees researchers from the reliance on service agent or customer self-reports. Data also span wide ranges of time and resolutions from minutes and days to months. Thus, the data allows exploration of the evolution of emotion and its effects over time. To illustrate, we report below on patterns of customer expression of emotion over the course of conversations, during the shift of a specific service agent. The breadth of the data allows us to unravel some fundamental issues regarding emotions in customer service, including issues that previous research constraints prohibited. For example, we report data that offer insights into what emotions customers actually display to service agents, as opposed to the customer emotions that agents remember or recall, which is what self-report data represent. We also report data on emotions displayed at different times of day, or different days of the week, and data on emotions expressed as service conversations unfold, from the perspective of the customer and the perspective of the service agent. In addition, and in contrast to most published research which reports on one firm or industry, the analyses we report refer to different firms and industries and include some comparisons between firms and industries. These comparisons have not been a major research goal for us but could become a future focus.

Of course, our data show dynamics of emotions in chat service conversations which is not exactly the same as the direct, absolute real time nature of phone or face to face conversations. But these data nonetheless provide informative insights into the genuine dynamics of service conversations. Moreover, new tools are being developed as we write this chapter that can avail similar analyses of other forms of service delivery.

2. The nature of variables available in digital traces data.
The organic nature of digital traces means that they come as they are, based on decisions of the designers or the entities that maintain the digital platform from which they are extracted. A major disadvantage is that researchers have limited control over the full set of variables that they can assess. For example, demographic variables—often expected and presumed essential in reports on behavioral science research—are not available to us in the analyses reported below. A key reason this information is not available is privacy. Barbaro and Zeller (2006) aptly illustrated the major effort that must be invested into disguising private information, to ensure that data of a specific person cannot be identified.

On the other hand, digital traces data has the advantage of being highly diversity in the information that they do include. Some information is automatically available, notably information unavailable to researchers until now without major and costly coding efforts. For example, data includes measures for the length of service conversations, the number of words exchanged by the parties and the response speed of participants (customers and service agent). All of this information provides fine-tuned access to the nature of the service situation. In our projects, for example, we use the number of words customers express as indices of the amount of cognitive load or complexity of the customer request that an agent must handle. Similarly, we use the number of words that agents post as an index of the effort a customer situation demanded (Altman, Ashtar, Olivares, & Yom-Tov, 2019). We also use the number of messages written in the course of a conversation as a measure of how long a conversation took, and the number of messages written by each participant as the relative contribution of each participant.

Empirical Insights: Analyses of Digital Traces of over a million service chat conversations

Available Data

From the archives of service conversations maintained by LivePerson we can trace the employee and the customer in each conversation, the number of customer words, the number of customer lines in the interaction, and an emotion score of the customer text. The emotion score for each sentence in the conversation was generated by automatic analyses using the tool we developed together with the firm as described above (see also Yom-Tov et al., 2018). The emotion score varies from -7 to +7. A score of zero (0) indicates No Emotion, presence and intensity of positive and negative emotions in each of the customer messages is coded on a scale of 1 to 7 for Positive Emotions, and -7 to -1 for Negative Emotions.

As reported below, for some of our analyses we aggregated these customer message-level scores to analyze emotions at the conversation-level of analysis. We define Positive conversations as including at least one positive message and no negative messages, Negative conversations as including at least one negative message and no positive messages and Multiple Emotion conversations as including at least one positive and one negative message. No emotion conversations are those where all the messages are assigned a score of 0 (zero). As reported below, these emotion scores can then be aggregated to represent the emotion in a full conversation, or to study changes in emotion over time within a conversation, or over time within an agent’s work shift. Importantly, the emotion scores represent the emotions that agents and customers express. We do not have access to the emotions people feel. From a theoretical perspective, however, we can view this as direct access to the emotion that a partner to the conversation experienced; partners have access only to what the other person expressed, not to what he or she felt (Hareli & Rafaeli, 2008). Expressed emotions are therefore most likely what influences the way a conversation unfolds.
Some of the data included three customer satisfaction variables, which were solicited after the conversations had ended: (1) **Customer Assessment of Employee Performance**, namely customer responses to the question “Thinking about the employee that you just chatted with, how would you rate him/her?” (1-5 scale, 1=poor to 5=excellent). (2) **Net Promoter Score** (NPS; Reichheld, 2003), an index of customer satisfaction based on customer responses to the question: “How likely is it that you would recommend our company to a friend or a colleague?” (0-10 scale, 0=Not likely to recommend, 10 = highly likely to recommend). (3) Customer assessment of **First Contact Resolution** (FCR; Hart, Fichtner, Fjalestad, & Langley, 2006), assessed by means of customer responses to the question “Was your query resolved in this interaction?” (Closed response scale: Yes/No).

Figure 1 describes the scope of the data in terms of the number of words in customer messages, and the number of customer messages in service conversations.

![Figure 1. Number of words in customer messages and number of customer messages in service interactions, in a sample of customer interactions [Airline, n= 25,714]](image)

The companies in the sample were selected to represent three different industries and include a telecommunication company (n=677,936), a retail company (n=439,585), and an airline (n=25,714). These samples are not huge, but they are substantially larger than typical samples in research on emotion in customer service. The data are large enough to provide insights that are highly likely to be representative of the population, and any findings with such data likely have external validity. Moreover, the data include a lot more granularity than most other research on emotion in customer service.

To retain the privacy and anonymity, we do not know anything about the agents or the customers, and cannot report demographic information. The data does include, however, an ID code identifying (and keeping anonymous) the agent and customer in each conversation. This means we can

---

1 These data are available only for part of the data because not all customers respond to the post-service surveys.
2 As noted, Figure 1 describes one company—an airline—and presented here as an overarching picture of the diversity in length of customer chat service conversations. This diversity is rarely mentioned or recognized in research on emotion in service, yet likely relevant to the full understanding of the complexity of service work.
3 For brevity we do not repeat any of our reported results here for the different firms in our sample. Thus Figure 1 reports only about the data of airline customers. Other figures report randomly about different firms that we analyzed.
track multiple conversations of the same customer or agent. As described below, we can therefore trace
the pattern of emotion customers express over the course of conversations, or the full load of customer
emotions an agent experiences over the course of a shift.

These data afford additional statistics, beyond number of words. Like number of words, the
statistics may or may not directly regard emotion, but we believe they are relevant and should be
considered in the context of research on emotion in service delivery. First, agents in the sample can
converse with more than one customer at a time, up to a maximum of 3 customers\(^4\). During a work-shift
an agent can converse with between 22 and 124 customers (Mean=64, SD=27). Second, the duration of
conversations varies widely, from less than a minute up to 39 minutes (Mean=11.5 minutes, SD=9
minutes). These types of data provide important insight into the interpersonally demanding nature of
customer service work, and these data are not based on or biased by agents’ memory.

New insights from analyses of digital traces of emotions in customer service conversations

The unique data we described above can attempt to address a large range of questions. Our
goal here is to illustrate the insights that such data can provide about issues and aspects of emotion in
customer service that have not been previously explored. We do not report on hypothesis testing data,
but rather an overview of novel quantitative insights about the phenomenon that is the focus of this
volume – emotion in service delivery. We specifically report next (a) what emotions do customers \(\text{really}\)
express in service conversations? (b) What customer emotions do service agents \(\text{really}\) encounter as
they perform service delivery work? (c) What customer emotions do service agents encounter in the
course of their work shifts? (d) How do customer emotions relate to customer evaluations of the service
delivery? (e) What are some differences in emotions expressed by customers in different industries?

a. What emotions do customers \(\text{really}\) express?

A first, somewhat surprising finding in our chat-service data regards the lack of emotion
expressed by customers in digital service. Our first, simplistic look, was at the emotion expressed in
individual customer messages; we found that most customer messages do not contain any emotions
(see Figure 1). This finding is notable since there is a common perception of emotion being central to
service delivery. Further surprising is that in instances where customers do express emotion, this
emotion is far more positive than negative! As evident in Figure 2, only 5% of customer messages
express negative emotion, while 20% of messages express positive emotion. It seems that contrary to
common belief the more dominant customer behavior is displaying positive not negative emotion. This
is an important point because we have substantial research about the extent and impact of customer
expressions of negative emotions toward service agents, and little research about the effect of
expressions of positive emotions (e.g., Foulk, Woolum, & Erez, 2016; Grandey, Rafaeli, Ravid, Wirtz, &
Steiner, 2010; Groth & Grandey, 2012; Rafaeli et al., 2012).

\(^4\) Current trends develop customer service delivered through texting (SMS) rather than chat. We have begun
analyzing such data, and find that in texting service agents converse with up to 20 customers at a given point in
time.
Figure 1. Frequencies of expressed positive and negative emotion in customer messages [Retail data, n=3,659,053]  

Figure 1 depicts a view would be excruciatingly difficult to obtain without analyses of digital traces data because without this technology discerning the emotions expressed in individual messages would be extremely labor-intensive endeavor.

A possible expansion of the analysis in Figure 1 is to look at the emotion expressed throughout full customer conversations rather than individual messages. Individual messages may not convey emotions, while full conversations may exude an emotional overtone. The benefit of the analysis of digital-traces data is the relative simplicity of aggregating from individual expressions to full conversations. Indeed, in this research line we continued beyond customer message-level emotion scores to capture the emotion overtone of full conversations. We defined conversations as having **Positive Emotion** if they included at least one positive message and no negative messages, as having **Negative Emotion** if they included at least one negative message and no positive messages, and as having **Multiple Emotions**, if they included at least one positive and one negative message. When emotion was not identified in any of the messages in a conversation a conversation was defined as having **No Emotion**.

The view into expression of emotion over full conversations -- as shown in Figure 2 -- shows that more than 80% of customer conversations do contain some expression of emotion. Interestingly here too it seems that negative emotions are not as frequent or dominant as available research seems to imply. Rather, Figure 2 shows that conversations are far less likely to have negative emotions, than positive emotions, multiple emotions, or no emotions. Only a small proportion of customers (5%) express only negative emotions, probably reflecting the 5% of customer messages that include negative emotions. In contrast, positive emotions appear in 20% of customer messages, and a dominant 57% of

---

5 Since our goal in this chapter is to illustrate insights that can be gleaned using the new tools and data we describe here, we report in different figures results for different samples from different industries. Our goal is not to describe any specific sample, organization, or industry. Rather, our goal is to show the diverse ways that data can be extracted and described in order to trigger future research ideas. Only where our goal is to draw comparisons between samples, we provide the comparative data between the samples.
customer conversations have customers expressing only positive emotions. This portrays customers in a new light, as being more pleasant and amenable to service agents than our literature on emotion in service delivery previously presumed. And opens a host of new questions for research.

Figure 2. Frequencies of emotion in service conversations [Retail, n =439,585]

These automated analyses of digital traces of organic data provide important insight into the multiplicity of customer emotions. Available research frequently speaks about the presence and effects of negative customer emotion. Service agents implicitly imply that expressions of negative emotions are a frequent and critical element in their work. The idea that negative emotions may occur also in the presence of other—positive—emotions has not been addressed in previous research. The general idea that people express and experience mixed emotions has received limited research attention (e.g., Berrios, Totterdell, & Kellett, 2015). What Figure 2 shows is that a substantial portion (21%) of customers express negative emotions together with positive emotions, within the same conversation. In other words, customers express multiple emotions, both negative and positive.

Recognizing that customers can convey multiple emotions within the same conversation raises a question about the emotional tone of a full conversation. The analysis in Figure 2 created a taxonomy of different types of conversations. But since we can also develop a continuous score indicating the amount of positive and negative emotion expressed in each conversation, we can compute the sum of negative and positive emotion scores in a conversation to provide a summary score for the emotional balance of each conversation. Such summing up is a crude estimate of the balance of positive and negative emotion in conversations, because it implicitly assumes that positive emotions cancel out the damaging effects of negative emotions expressed by the same customer. This of course is an empirical question that deserves research attention, but not a totally unlikely proposition. Customer expressions of positive emotions can be viewed as forms of social support or resource that customer imbue to service agents, while expressions of negative emotions are viewed as work demands and causes of depletion.

Notwithstanding, the bottom line is that such a summary score affords a way to examine the profile of service conversations in terms of the emotions that customers express. Conversations with a
summary positive score suggest a conversation with an overall balance of positive emotions, while those
with a summary negative score indicate a conversation that has a balance of negative emotions. The
results of this computation – as depicted in Figure 3 – continue to show more overly positive emotion
conversations than overly negative emotion conversations. The tail of positive emotion conversations
in Figure 3 is much longer than the tail of negative emotion conversations. Given this we suggest that
there might be more variation in the types of positive emotion compared to negative emotion in a
conversation. The important message in Figure 3 is that more than 70% of conversations where a
customer expresses some emotion, have an overall positive emotion.

![Figure 3. Distribution of sum of positive and negative customer emotions in full conversations [Airline, n=
21,122]](image)

b. What customer emotions do service agents really encounter in their service delivery work?

By looking at the emotions that customers express in conversations, we obtain a more refined
understanding of customers’ emotions in service. For example, we can ask whether customer emotions
vary with the time of day or day of the week a service conversation occurs. A complementary view that
adds more detailed insight is obtained by looking at the evolution of emotion customers express within
a conversation. For example, we can look at the emotions expressed early on (at the opening of)
conversations or later, toward the end of the conversations. In Figure 4 we show just this, an analysis
that focuses on emotions that customers express at the beginning and the end of their conversations.
Figure 4 shows that the emotions customers express vary (and improve) from the first to the last
message. In aggregate, customers appear to start off conversations with very mild negative emotions
(sentiment scores < 0), and end conversations with expressions of mildly positive emotions (sentiment scores around 0.7). This pattern is not related to time of the day a conversation occurs.

Figure 4. Customer emotion in first message and last message of a conversation with a service agent [Airline, n= 25,668] ⁶

A more refined look at the emotion customers express within conversations is depicted in Figure 5, which shows the emotions typical to multiple stages or sections within conversations. This type of analysis portrays service encounters as a sequence of emotional displays, extending beyond the focus on peak emotions, as suggested by Verhoef, Antonides, and de Hoog, (2004), for example. Since customer conversations vary in the number messages they comprise, we must first create a standardized metric that allows comparisons of different conversations. We obtain such standardization by splitting all conversations into 10 roughly equal sections; this standardization means that sections in different interactions may comprise a different number of messages, but all conversations comprise exactly 10 sections (or 10 quartiles). Using such standardization we were able to average the emotion scores of all customer messages in each section and obtain a metric of the valence of customer emotion per section. Each conversation is thus defined as comprising 10 sections, and 10 emotion scores. The result of this standardization allows us to depict the flow of emotion over the course of multiple interactions, as shown in Figure 5⁷.

Figure 5 suggests that conversations have a standard structure comprising three within conversation stages —opening, middle (or main), and closing. Figure 6 shows negligible variations between replications of the analysis of emotion by section at different hours, and on different days of the week. Similar to Figure 5, Figure 6 again suggests that conversations open with minor negative

---

⁶ Figure 4 seems to suggest a minor effect of a drop in positive customer emotion in the end of conversations emotion in the first and last hour of the day: slightly higher positive emotion is expressed at 08:00 than at 09:00, and slightly lower positive emotion is expressed at 21:00 than at 22:00. These are not big differences and our analyses could not identify why they occur. Further and future research might be able to unravel these dynamics.

⁷ The data in Figure 6 describe only the subset of conversations that included 10 or more customer messages. We repeated this analysis on all the data, including shorter interactions, as a robustness check, and found a similar pattern. In this repeated analysis, we stretched short interactions that include less than 10 customer messages, by duplicating missing quantiles. For example, for an interaction with length 5: 1,2,3,4,5, the 10 points were 1,1,2,2,3,3,4,4,5,5.
emotions, and end with more positive emotions. The main and middle of conversations shows customers as being mostly neutral (non-emotional).

(b) By day of week

Figure 5. Customer emotion expressed in different sections of service conversations, at different hours of the day and different days of the week [Telecommunication, n= 390,438]

It is not surprising that customers express negative emotion at the starts of their conversations, since initiating a service conversation usually means the customer has a problem. The more positive emotion expressed toward the end of conversations presumably suggests that the conversation helped resolve the customer’s problem. The middle sections, where there seems to be little expression of emotion by customers, likely focus on the technical issues relevant to the customer problem or its solution. In further analyses—reported below- we examine the trajectory of the patterns of improvement of emotion, as they relate to the quality of resolution of the customer problem. We show this by connecting the pattern of improvement in emotion to post-conversation surveys of customer satisfaction.

c. What customer emotions do service agents encounter over the course of a work shift?

Agents often report encountering a lot of hostile customer emotions, but this depiction of service work hasn’t been objectively verified. Our data allow addressing this question in two ways. First, as depicted in Figure 6, we tracked the overall customer emotion (resulting from 124 conversations) encountered by a random employee on a random workday.
Looking at the emotions encountered by that one agent carries the risk of a sampling bias. Perhaps we randomly selected a particularly problematic agent? Thus, we also compute an aggregation of emotions that all customers convey over the course of a full workday across all agents. This depiction removes the concern of a sampling error, an outlier or special case agent. Figure 7 thus shows the cumulative positive and negative emotions expressed by all customers over the course of a full workday. The picture depicted by Figure 7 shows a rise and fall of expressions of positive and negative emotions. The general picture is of more expressions of positive emotions than negative emotions, with some outlying points. We propose that the transitions in customer emotions evident in Figure 6 and Figure 7 are the most difficult part of service agents’ work, similar to the depleting and debilitating social influence that Rafaeli and Sutton (1991) described in encounters with emotionally contrasting social expressions by conversation partners.
Relating customer expression of emotions to customer evaluations of the service conversation.

Our analyses thus far have been purely descriptive, showing the emotions that customers express in the context of customer service conversations. An additional perspective that research relying on digital traces data can provide is the relationship of customer emotions expressed during service conversations to customer evaluations of the service agent and service conversation after the service ended. For this perspective we integrate the analyses of customer expressed emotion with data we have regarding customer evaluations of service quality. The firm we work with, like many service providers, follows up on service conversations with a text message asking customers to respond to a short survey assessing their satisfaction with the service they received. The survey asks basic questions, such as “How satisfied are you with the performance of the service agent,” allowing responses of 1 through 5, with 1 indicating high dissatisfaction, 5 indicating high satisfaction. Responses to such post-service surveys are voluntary and hence will always only represent partial data for all the customers in our samples. Notwithstanding, the sample sizes of our analyses here are still substantially larger than most sample sizes in previous emotion-in-customer-service research.

Figure 8 depicts the pattern of emotion expressed by customers within their service conversations, broken down by the customers’ response regarding their level of satisfaction with the performance of the service agent. As noted, the evaluations of the agent performance were given after the service conversation had ended. In contrast, the assessment of the expressed emotion was conducted during the conversation.
The frame of reference in Figure 8—the bold line—depicts the mean emotion throughout the conversation in the population of customers who did not respond to a post-service survey. Figure 8 shows that the emotions expressed by customers who were extremely satisfied with the agent performance (rating of 5) climbed higher, to include expressions of more positive emotions. In contrast, emotions expressed by customers who were dissatisfied with their agent’s performance (rating of 1) remained low throughout the conversations. The figure also shows is that these customers started out with the same level of negative emotions as most other customers, negating the possibility that these customers who started out with more negative emotions.

Figure 8. Customers’ expressed emotions during service conversation by level of customer satisfaction with the performance of the service agent after the service conversation had ended [Airline, n= 6,973]

A different, but related measure of customer satisfaction—known as Net Promoter Score (NPS)—tells a similar story. NPS is the response of customers to the question “would you recommend this service provider to your friends or family members.” Figure 9 shows that, relative to the general population of non-responding customers, customers who are highly satisfied, and eager to recommend the service provider to friends or acquaintances, show a steeper improvement in the positive emotions they express. Moreover, customers who end up highly satisfied with the service they received (NPS=5) start expressing higher positive emotions earlier in the conversation than customers who were mildly satisfied (NPS=3), and customers who were not at all satisfied (NPS=1) do not express any positive emotions during the conversation.
Finally, in Figure 10, we show the evolving customer emotions during conversations as they relate to customer evaluations of the extent to which their service issue or problem were resolved by the conversation. This criterion — referred to in the service industry as First Call Resolution (FCR) — is an index of efficiency of handling customer service issues. The picture depicted in Figure 10 suggests that service conversations more effective in resolving customer needs also evoke positive customer emotions earlier in the service conversation.

**Figure 9.** Customers expressed emotions during service conversation by level of customer overall satisfaction with the service agent as measured by the Net Promoter Score (NPS) [retail, n= 123,554]

**Figure 10.** Customer emotion in sections of interaction by customer evaluations of whether their service issue had been resolved [Telecommunication, n=390,438].

e. Customer emotions in different industries?

The methods we describe for automated analyses of emotions of digital traces of expressions of emotion by customers are unique in providing quick and relatively easy—automated rather than manual—analyses of large amounts of data. A new and interesting opportunity such data afford is examination and comparisons of emotions expressed by customer of different companies or industries. The data we describe here included service conversations from three firms, intentionally selected as belonging to three different industries: telecommunication services, retail services and airline services.
Figure 11 shows the comparison of the customer emotions in each of these firms, essentially replicating the distribution shown in Figure 2 for the three industries.

![Customer emotion in service conversations in firms from three industries](image)

There are some similar patterns of customer emotion across the different service firms as shown in Figure 11: in all firms both less than 20% of conversations are devoid of emotion and positive emotion has the largest presence. There appears to be a substantial difference between the telecommunication firm (having substantially fewer conversations with only positive emotions) and the other two firms (38% compared to 52% in airline and 57% in retail). The amount of conversations having negative emotions also appears to be higher in telecommunications than in the other two firms (10%, vs. 7% in airline and 5% in retail). Such comparisons add to our understanding of the specific difficulties of handling customer support services in different firms and industries.

Differences in customer emotion may relate to differences in other aspects of service delivery, which of which our data also allows comparisons. For example, that the average number of customer messages in a full conversation is longer in the telecommunication firm (14.8 (SD=13.2)), than in the retail (8.32 (SD=7.80)), and the airline (8.12 (SD=7.78)) firms. The conversations in the telecommunication firm are, on average, almost double in length. Moreover, employees of the telecommunication firm interact with 13 to 69 (Mean=35, SD=17) customers during a shift, and their conversations vary in duration between 3 minutes and 49 minutes (Mean=19, SD=15). In comparison, customer service agents in the retail industry firm interact with 11 to 82 (Mean=46, SD=20) customers during a shift, and their conversations vary in duration between 3 minutes and 37 minutes (Mean=14; SD=12). It could be, for example, that length of conversations, or the number of customers with whom an agent typically converses, relate to the emotions elicited in customers. We could not conduct such analyses in the scope of this chapter, but they do suggest new directions for future research.

f. Relating customer expression of emotion to service agents’ behaviors.
A particularly intriguing vantage point that our analyses can provide regards the relationship between the emotions expressed by customers and the behaviors of service agents. This question was the foundation early work on emotional labor, which presumed that the emotions displayed by service agents is critical to the satisfaction of customers (Sutton & Rafaeli, 1988). More recent analyses, however, posited reverse causality, showing that expressions of anger by customers can hamper employees’ cognitive abilities (Rafaeli et al., 2012). In this vein, we sought to examine the relationship between customer expression of emotion and work behaviors of service agents. This line of our work is reported in greater detail in Altman et al. (2019). A key and intriguing graphic depiction of our findings is shown in Figure 12.

![Figure 12. Relationship between emotion in customer messages and employee response time (RT) in subsequent message. (Airline, n= 1,447,070 customer messages in 208,210 service conversations)\(^8\)](image)

Figure 12 shows firstly that the lowest response time—which is presumably reflects the best agent performance if we assume customers do not like to wait—occurs when customers do not express any emotion. Second, we see that positive customer emotions may add to the overall agent response time, but this is a marginal addition. Third, and perhaps most significantly, the presence of negative customer is related to an increase in agent response time. Fourth, the gray areas in Figure 12, which represent the variance (95% confidence interval) around the mean response times show individual differences among agents. We see that the variance increases with more extreme positive and negative customer emotions. A possible interpretation is that this variance is due to differences in agent sensitivity to emotion expressed by others, perhaps reflecting the extent to which different agents internalize the emotions of others, or differences in agent emotional intelligence. Clearly this pattern calls for additional research.

\(^8\) Figure 12 is based on emotion detection done by SentiStrength (Thelwall, 2013). Values above 0 indicate expression of positive emotion, values below zero are expressions of negative emotion.
Discussion

This chapter describes a new approach for studying customer emotions in service interactions. The approach capitalizes on digital traces that modern day service delivery comprises (Rafaeli et al., 2019). To the best of our knowledge this is the first objective and detailed depiction of the actual emotional encounters that customers express, and that customer-service work comprises. Previous research reported primarily what agents recall, relying on self-report data of agents about their work experience. Since self-reports are subjected to various biases, the analyses of digital traces data that we propose and describe here provide a more nuanced, objective and fine-grained perspective on the actual nature of service work. The contribution of this chapter is therefore threefold: (a) New methods: we propose automated emotion analysis as a useful tool for research and management of emotion in customer service; (b) Greater breadth: we document dynamics of customer emotions in large-scale samples of real customer-service interactions; (c) Higher granularity: we document emotion dynamics within and throughout service conversations, and within and throughout work shifts of service agents.

The automated sentiment analysis tools that our analyses rely upon can be implemented in real-time and provide insights into customer expressed emotions as a conversation unfolds. Such implementation may offer a way to assess the effectiveness of customer service before a service conversation has ended and without the additional costs and issues associated with customer self-report.

The digital data and newly developed tools for sentiment analyses allow exploration of emotions in large samples of genuine customer-service interactions. The new types of data and methods that we describe and illustrate offer substantial benefits: the research provides objective, unobtrusive views of customer emotions that draw directly from customer expressions, with no self-report intervention and biases (Webb, Campbell, Schwartz, & Sechrest, 1966; Xu et al., 2019). This chapter thus provides a lens into the dynamics of emotions in service that could not be obtained using traditional research methods. The methods and findings we depicted cannot replace traditional research in psychology and organizational behavior, and do not test significance or causality. Rather, the findings unravel new dynamics that should be followed up with more research, both research using traditional experimental methods, and digital traces research that allows inferences of causality. For example, one project we are still working on explores whether customer sentiment influences which customer an agent chooses to respond to. In this effort we view a response to a customer as a form of attention and thus a reward, which helps us continue to examine the complicated question of whether customer anger is rewarded (Glikson, Rees, Wirtz, Kopelman, & Rafaeli, 2019). A second project examines the relationships between customer emotions and agent response times (Altman et al., 2019).

Toward Future Research

Our analyses only scratch the surface of the types of issues and questions that future research can address utilizing the opportunities that we highlight. The amount of relevant data, its velocity (the pace with which it is accruing and increasing), and its variety (relevant data comes in many forms, and from many sources both within and outside of the organization) coalesce to a rich and exciting research agenda.

Our findings show that, in contrast to common belief, negative emotions are expressed by only a small proportion of customers (< 10%) and appear in less than 5% of customer messages. Positive customer emotions are much more common than was previously recognized; positive emotions are also
expressed in higher intensity than negative emotions. Some customer conversations include both positive and negative emotions (21%), but even this occurs in fewer conversations than conversations with only positive customer emotions (57%). The bottom line is that there is a lot of positive emotion in service delivery, yet we have little research on the triggers, patterns or effects of positive customer emotions. This is therefore the first call for future research arising out of this chapter.

The co-occurrence of positive and negative emotions is shown in our data to be somewhat randomly spread over service agents’ work shifts. This suggests that service delivery work is an emotional roller coaster, saturated with affective events (Weiss & Cropanzano, 1996). Service agents move back and forth between encounters with negative and positive customer emotion. The contrast between encountering positive and negative customer emotions may be responsible for the high burnout typical to service agents, rather than the sole effect of negative customer emotions as is commonly believed. As Rafaeli and Sutton (1991) showed, encounters with contrasting emotions weaken and exhaust people. The co-occurrence and wavering between positive and negative customer emotions may also undelay the tendency of service agents to report a high presence of customer expressions of negative emotions. Contrasting effects influence perception and memory (Manstead, Wagner, & MacDonald, 1983), and most available research on the extent of negative emotions has so far relied on self-report methods, which are notoriously susceptible to memory effects. Our second call for future research regards finer testing and verification of these influences.

The evolving emotional footprint within service conversations is an additional new lens our analyses offer; our data show a consistent pattern wherein customer emotions start out negative in almost all conversations, segue into a middle segment which is a plateau with little or no customer expression of emotion, and finally with an increase toward positive emotion at the end of conversations. Our analyses also suggest that the trajectory of this curve—which varies between customers—may be meaningful. Our observations of the data plots of large numbers of customers lead us to offer a proposition that the trajectory—the angle of the curve, and / or the point from which customer emotions begin to improve into positive expressions—can predict the extent to which a customer will be content with the service agent, the overall service, and the resolution of the customer service issue after the conversation has ended. Our third call for future research is therefore a validation of this prediction.

Validation research is particularly important because it can show the merit of automated emotion assessment conducted in real-time during a service interaction. Once the relationship between customer emotions and customer satisfaction is empirically documented, emotion assessments can be exploited not only to predict customer satisfaction, but also to identify and perhaps preempt service failures (Tax & Brown, 1998). Current research on service failures relies on customer reports and responses to surveys, which come at a delay, and with limited response rates (e.g., Casidy & Shin, 2015; Joireman, Grégoire, Devezer, & Tripp, 2013; Smith & Bolton, 2002). With further empirical research, post-hoc customer satisfaction might be replaced by real-time monitoring of emotion, which can provide useful tools for managerial decisions or supervisor interventions.

The data, tools and insights shown in this chapter can also provide a practical contribution. It suggests new ways for service organizations to identify customer emotions and can assist service managers in leveraging analyses of customer emotions into their operational procedures. Automated tools can allow a wide range of analyses on all sorts and forms of big-data and connect these data to emotions. Altman et al. (2019), for example, use quantitative models of similar data to examine the influences of customer emotions on employee efficiency. Ashtar et al. (2018) used similar data to relate customer emotions to unscheduled micro breaks taken by employees. Implementations of these and
other findings can promote management of operations (George, Osinga, Lavie, & Scott, 2016), human resources (McAbee, Landis, & Burke, 2017), and service delivery (Rafaeli et al., 2017).

In short, we hope to convey in this chapter is that a new digital age approach to the study of affect in service offers exciting opportunities. Our data highlights the type of novel questions that can be asked, and accordingly the novel answers that the Digital Age approach to emotion research can offer. We are certain that our own research is only the tip of a huge iceberg of potential research. We hope our review will stimulate other researchers to join us in this new research adventure.

References


Emotion in Digital Service – Page 26


Jones, N. M., & Silver, R. C. (2019). This is not a drill: Anxiety on Twitter following the 2018 Hawaii false missile alert. *American Psychologist*.


