Article

AMERICAN MARKETING ASSOCIATION

Journal of Marketing
2023, Vol. 87(2) 210-231

DOI: 10.1177/00222429221119977
journals.sagepub.com/home/jmx
SAGE

Complaint De-Escalation Strategies on Social Media

Dennis Herhausen, Lauren Grewal, Krista Hill Cummings, Anne L. Roggeveen, Francisco Villarroel Ordenes, and Dhruv Grewal

Abstract

To date, the literature offers multiple suggestions for how to recover from service failures, albeit without explicitly addressing customers' negative, high-arousal states evoked by the failure. The few studies that do address ways to improve negative emotions after failures focus on face-to-face interactions only. Because many customers today prefer to complain on social media, firms must learn how to effectively de-escalate negative, high-arousal emotions through text-based exchanges to achieve successful service recoveries. With three field studies using natural language processing tools and three preregistered controlled experiments, the current research identifies ways to mitigate negative arousal in text-based social media complaining, specifically, active listening and empathy. In detail, increasing active listening and empathy in the firm response evokes gratitude among customers in high-arousal states, even if the actual failure is not (yet) recovered. These findings provide a new theoretical perspective on the role of customer arousal in service failures and recoveries as well as managerially relevant implications for dealing with public social media complaints.

Keywords

social media complaints, digital service recovery, negative arousal, de-escalation

Online supplement: https://doi.org/10.1177/00222429221119977

Service failures are inevitable and regular. Despite insights gained from many years of service recovery research, many firms still struggle to offer effective responses to complaining customers (Knox and Van Oest 2014). In the latest National Customer Rage Study (Customer Care Measurement & Consulting 2020), less than one-third of respondents indicated being satisfied with service recoveries, and two-thirds expressed negative, high-arousal emotions, such as anger. The relative anonymity provided by social media has also boosted expressions of negative arousal (Williams 2019). Failing to de-escalate these expressed negative, high-arousal emotions may be a critical reason that many recovery attempts are unsuccessful (Flechas 2020).

Despite various suggestions in prior literature for how to respond to complaining customers after a failure, few studies address ways to improve negative emotions evoked by the failure, and those that do tend to focus on face-to-face interactions only (Tax, Brown, and Chandrashekaran 1998). However, approximately 89% of customers today prefer social media communication with firms over other channels (Avochato 2021). Social media communication is characterized by a single modality (i.e., text only), asynchronous interactions (i.e., time lags in the conversation), and exposure to public scrutiny (i.e., broadcasted interactions). Therefore, de-escalating negative arousal in this distinct setting is crucial to current-day recovery attempts.

Although recent research indicates that firms should address public complaints to limit detrimental effects on other customers (Herhausen et al. 2019; Herhausen 2020), it is unclear which response strategies are best suited to de-escalate highly aroused customers and evoke a feeling of gratitude in the social media complainant. We focus on customer gratitude

Dennis Herhausen is Associate Professor of Marketing, Vrije Universiteit Amsterdam, The Netherlands (email: dennis.herhausen@vu.nl). Lauren Grewal is Associate Professor of Business Administration, Tuck School of Business, Dartmouth College, USA (email: lauren.s.grewal@tuck.dartmouth.edu). Krista Hill Cummings is Assistant Professor, Babson College, USA (email: khill@babson.edu). Anne L. Roggeveen is Charles Clarke Reynolds Professor of Retailing & Marketing, Babson College, USA (email: aroggeveen@babson.edu). Francisco Villarroel Ordenes is Assistant Professor of Marketing, Department of Business and Management, LUISS Guido Carli, Italy (email: fvillarroel@luiss.it). Dhruv Grewal is Toyota Professor of Commerce and Electronic Business, Babson College, USA (email: dgrewal@babson.edu).

(i.e., the emotional appreciation for benefits received when firm actions exceed requirements) in this social media context for three reasons. First, gratitude is a crucial first step in restoring customer relationships and is associated with numerous positive, long-lasting outcomes such as satisfaction, commitment, and perceived relationship value (Bonchek 2015; Palmatier et al. 2009). Second, de-escalation is not focused on "solving the problem" but rather on addressing negative, high-arousal emotions. Gratitude is a strong indicator that the customers' high-arousal emotions have been de-escalated effectively (Xia and Kukar-Kinney 2013). Third, firms often try to move complaining customers to private channels in their public response (Golmohammadi et al. 2021), and gratitude as a potential reaction in the public customer reaction to a firm response may be the only visible outcome for readers of complaints.

Drawing from crisis negotiation literature (Vecchi, Van Hasselt, and Romano 2005), we define "de-escalation" for our research as responding to a complaint with the aim of lowering negative, high-arousal emotions, whereas "recovery" is defined more broadly as all the actions a firm can take to resolve the problems or loss caused by a service failure (Khamitov, Grégoire, and Suri 2020). Thus, de-escalation is a specific element within the broad armory of different recovery tactics (Van Vaerenbergh et al. 2019). We postulate that it is essential that de-escalation precedes or complements other recovery tactics.

We believe that there are two effective response strategies firms could adopt to de-escalate negative arousal and enhance costumer gratitude in social media: active listening and empathy. Active listening implies paying attention to what the customer says and demonstrating that attention through actions such as repeating, paraphrasing, or adapting the language to the customer. Empathy involves connecting emotionally with complaining customers by indicating understanding of their feelings, using explicit expressions of validation and affirmation. In a text-based context, active listening is related to the style of the response (i.e., linguistic style matching) and empathy is related to the content of the response (i.e., using empathetic words).

We explore the effects of active listening and empathy on customer negative arousal with three field studies and three preregistered experiments. The results affirm that high- (vs. low-) arousal states reduce complainants' gratitude for the firm response: a 1% increase in negative high-arousal words leads to a 19% lower probability of gratitude. We also find converging evidence that both active listening and empathy by the firm, independently, can de-escalate high-arousal emotions and increase complainants' gratitude, *even without providing any actual recovery*. The field data establish that increasing firm active listening (empathy) by 1% increases the probability of customer gratitude by up to 14% (90%). Thus, relative to active listening, empathy is a stronger lever to enhance relevant outcomes.

Our findings provide important theoretical and managerial contributions. First, we propose a novel theoretical perspective on handling negative emotions in text-based complaining, and we shed light on the consequences of providing active listening and empathy in firm responses. By assessing the implications of these responses for customer gratitude, we extend current debates about how firms should handle public complaints on social media. Second, while it may seem to be intuitive that firms should use active listening or empathy to de-escalate customers' arousal in face-to-face settings, this is the first empirical research that quantifies their potential benefits in asynchronous complaint settings using social media data. Third, we provide concrete recommendations for how to implement active listening and empathy in written, asynchronous communication. When responding to social media complaints, linguistic style matching of function words signals that an employee is actively listening while using empathetic words from our newly developed dictionary signals empathy. Indeed, our field studies indicate that firms are not currently taking full advantage of these strategies in social media interactions.

Conceptual Development and Hypotheses

We provide an illustrative overview of empirical research on customers' emotions in service failures in Table 1. This review highlights that research on service failures and complaints has predominantly focused on the face-to-face context and has only just started to explore the social media context. A recent opinion piece speculates that dealing with complaints in social media should largely mimic in-person strategies (Grégoire and Mattila 2020), but the social media context features some notable differences that may be relevant for dealing with the emotions evoked by failures. Thus, it is not clear to what degree findings from the face-to-face context generalize to social media complaint settings.

Social media communications largely rely on a single modality (i.e., text), whereas face-to-face interactions include multiple verbal and nonverbal cues, such as body-language expressions and vocal intonations (Singh, Marinova, and Singh 2020). The absence of such cues and the anonymity of social media can promote the use of more negative and aroused language by consumers (Coker 2020). In addition, social media communication is asynchronous, marked by time lags within the conversations. Asynchronous communication gives firms more time to reexamine what customers say (Berger and Iyengar 2013), which might support more conscious complaint handling. However, asynchrony can also be detrimental if the complaint is ambiguous and requires clarification between communication partners (Moffett, Folse, and Palmatier 2021).

Moreover, communication in social media is open to public scrutiny, so any complaints are "broadcast" to both firms and external others. This broadcasting tends to prevent customers from posting content that makes them look bad (Barasch and Berger 2014) and increases the likelihood that they blame the firm for the failure. In response, firms strive to resolve complaints as quickly as possible (Herhausen et al. 2019) or force the communication into a private complaint channel (Golmohammadi et al. 2021) to limit the negative effects in terms of other people reading the complaint exchange.

Table 1. Main Studies on Customers' Emotions in Service Failures.

Study	Method(s)	Context	Focus	Valence and Arousal	De-Escalation	Main Findings
Tax, Brown, and Chandrashekaran (1998)	Experimental	Face-to-face	Complainant	Negative valence and low arousal only (i.e., dissatisfaction)	—	Perceived empathy in the firm response increases satisfaction.
Smith and Bolton (2002)	Experimental	Face-to-face	Complainant	No empirical differentiation of emotions with negative valence	—	Customers' negative emotions after service failures influence recovery evaluations.
Bougie, Pieters, and Zeelenberg (2003)	Experimental	Face-to-face	Complainant	Differentiating high- versus low-arousal emotions (i.e., anger and dissatisfaction)	—	Customers show high- and low-arousal negative emotions after service failures.
Nguyen and McColl-Kennedy (2003)	Conceptual	Face-to-face	Complainant	Negative valence and high arousal only (i.e., anger)	Yes	Active listening in the firm response is expected to decrease negative emotions.
Kalamas, Laroche, and Makdessian (2008)	Experimental	Face-to-face	Complainant	Negative valence and high arousal only (i.e., anger and related emotional states)	—	Negative, high-arousal emotions are the dominant response to a service failure.
Gelbrich (2010)	Survey/ experimental	Face-to-face	Complainant	Negative valence and high arousal only (i.e., anger and frustration)	—	Negative, high-arousal emotions enhance vindictive word of mouth.
Strizhakova, Tsarenko, and Ruth (2012)	Experimental	Face-to-face	Complainant	Negative valence and high arousal only (i.e., anger)	—	Negative, high-arousal emotions enhance vindictive word of mouth.
Joireman et al. (2013)	Survey	Face-to-face	Complainant	Negative valence and high arousal only (i.e., anger)	—	Negative, high-arousal emotions increase revenge behavior after a firm response.
Surachartkumtonkun, McColl-Kennedy, and Patterson (2015)	Survey	Face-to-face	Complainant	Differentiating high- versus low-arousal emotions (i.e., rage, anger, and dissatisfaction)	—	Negative high- versus low-arousal emotions become dominant for unresolved failures.
Umashankar, Srinivasan, and Parker (2016)	Field data/ experimental	Face-to-face	Complainant	Negative valence and high arousal only (i.e., anger)	—	Customers' negative emotions dominate their cognitive responses to service failures.
Grégoire et al. (2018)	Survey/ experimental	Face-to-face	Complainant	Negative valence and high arousal only (i.e., anger)	—	Negative emotions increase the desire for revenge after a firm response.
Herhausen et al. (2019)	Field data	Social media	Observer	Differentiating high- versus low-arousal emotions (i.e., anger, anxiety, disgust, and sadness)	—	High- versus low-arousal emotions increase negative effects on observers.
Golmohammadi et al. (2021)	Field data	Social media	Observer	No empirical differentiation of	—	Publicly responding to customer complaints

(continued)

Table 1. (continued)

Study	Method(s)	Context	Focus	Valence and Arousal	De-Escalation	Main Findings
				emotions with negative valence		has negative effects on observers.
This study	Field data/ experimental	Social media	Complainant	Isolating high versus low arousal from negative valence (while controlling for other emotions)	Yes	High- versus low-arousal emotions reduce gratitude. Active listening and empathy in the firm response de-escalate high arousal emotions and increase gratitude. For low-arousal emotions, there are diminishing effects for active listening while the effect of empathy varies across studies.

However, it is equally important to examine how such social media response strategies are viewed from the perspective of the complaining customer.

In addition, most existing studies do not differentiate between high- versus low-arousal negative emotions. The notable exceptions that considered the role of arousal in face-to-face settings did not consider the de-escalation of negative arousal (Bougie, Pieters, and Zeelenberg 2003; Surachartkumtonkun, McColl-Kennedy, and Patterson 2015), and the only study that considered the role of arousal in social media complaining focused on observer effects only (Herhausen et al. 2019). Given that customers in negative, high-arousal states are less receptive to firm responses (Gelbrich 2010; Joireman et al. 2013), the isolation of valence (which is typically negative in complaints) and arousal (which can be higher and lower in complaints) appears important in understanding how to best respond to complaints. In addition, insights on how to implement de-escalation in social media response strategies to reduce negative arousal would greatly benefit firms.

The present research studies firm response strategies to customer complaints in social media, isolating arousal from negative valence in customer complaints and examining ways to de-escalate negative arousal. More specifically, we explore the role of high negative arousal associated with a customer complaint after a service failure and the complainant's likelihood of showing gratitude after the firm's social media response efforts (H_1), and how using de-escalation strategies in the firm response affects gratitude for high-arousal complainants (H_2 and H_3).

High Arousal and Customer Gratitude

Emotions evoked by service failures affect customers' reactions to relationship restoration efforts, which impact customer gratitude and subsequent loyalty (Umashankar, Srinivasan, and Parker 2016). Although emotions can vary along valence (positive/negative) and arousal (activated/deactivation; Russell 1980), after a service failure, customers generally experience negative emotions, varying in arousal levels (Bonifield and Cole 2007; Surachartkumtonkun, McColl-Kennedy, and Patterson 2015). For example, anger is a high-arousal emotion, whereas dissatisfaction is a medium-arousal emotion, and both would be directed at the firm that has failed a customer (e.g., an airline that canceled a flight).[1]

Even if multiple negative emotions can be experienced after a service failure (Valentini, Orsingher, and Polyakova 2020), high-arousal emotions make any restoration more difficult, due to the idiosyncratic behavioral tendencies they evoke (Bougie, Pieters, and Zeelenberg 2003). Anger, a quintessential high-arousal emotion, increases desires for revenge or vindictive word of mouth after a failure (Gelbrich 2010). Anger also elicits heightened expressive tendencies (Kalamas, Laroche, and Makdessian 2008), making it more likely that customers publicly complain. In addition, when customers experience negative high-arousal emotions, a successful relationship restoration becomes less likely, because they become unreceptive to problem solving and are less likely to see any firm response as beneficial or deserving of gratitude.

As customers experience these negative, high-arousal emotions, they are less likely to be able to emotionally appreciate the benefits (i.e., feel gratitude) that the firm is attempting to provide them. As mentioned previously, these high-arousal emotions make restoration efforts harder for the firm (and their employees). Therefore, firms need to focus on reducing these emotions to evoke greater levels of customer gratitude.

> **H_1:** When the complaining customer expresses relatively higher (vs. lower) negative arousal in a public social media context, this reduces the likelihood of gratitude for the firm response.

[1] Drawing on previous research that shows that the emotional response and degree of negative arousal is a function of the type and magnitude of the failure, we treat negative arousal as a failure-specific variable, but we also acknowledge a potential consumer-specific influence on the failure severity–higher negative arousal link.

De-Escalation of High Arousal

Given the detrimental effect postulated in H_1, de-escalating negative high arousal may be critical for successfully addressing customer complaints in social media. As the crisis negotiation literature points out, significant loss events put people in an emotional state characterized by strong negative arousal, such that they behave and think irrationally (Noesner and Webster 1997). A person in crisis thus tends to be unreceptive to solutions. To deal with such a situation, the Behavioral Change Stairway Model (Vecchi, Van Hasselt, and Romano 2005) suggests a relationship-building process in which the negotiator de-escalates the negative arousal with active listening and empathy. Other prominent models in the crisis negotiation literature also note the importance of de-escalating negative arousal to enable problem-solving (Web Appendix A).[2]

We acknowledge that service failures may be less severe than many crises, but we also note pertinent parallels between a person in crisis and an enraged customer after a failure. Angry customers experience high-arousal, negative emotions; have a strong desire to be heard; and are not very receptive to offered resolutions (Bonifield and Cole 2007; Gelbrich 2010). In Web Appendix A, we juxtapose prominent crisis negotiation models with the service recovery journey to highlight their similarities. That is, the service recovery journey includes stages such as recognizing the failure, addressing it through a service recovery, and seeking positive postrecovery behaviors, which line up with crisis negotiation models that refer to identifying a problematic situation, addressing it, and adopting various behaviors. However, crisis negotiation literature also explicitly includes a stage for addressing and de-escalating high-arousal, negative emotions, which the existing service recovery literature does not.

De-escalating high negative arousal with active listening. In crisis negotiations, active listening is a key technique for connecting with another person and reducing arousal (McMains and Lanceley 2003). Verbal techniques to demonstrate active listening include language mirroring and paraphrasing (Vecchi, Van Hasselt, and Romano 2005). Specifically, previous research suggests that communicators who match their partners' language style are actively engaged at a fundamental and structural level (Cannava and Bodie 2017; Ireland and Henderson 2014; Niederhoffer and Pennebaker 2002). Thus, in text-based service recoveries on social media, firms can use language style mirroring in its response to demonstrate active listening.

Active listening engenders feelings of affiliation and rapport (Lakin and Chartrand 2003; Min, Jung, and Ryu 2021), which can elicit low arousal and improve subsequent outcomes. Negotiation research has also found that when negotiators mimic language during online chats, it enables them to connect emotionally (Swaab, Maddux, and Sinaceur 2011). Active listening has also been proposed to reduce anger in service failure situations (Nguyen and McColl-Kennedy 2003). When customers believe that a firm's employee is actively listening, they tend to be more trusting, more satisfied, and more willing to do future business with the firm (Ramsey and Sohi 1997). By matching the linguistic style of a highly aroused, complaining customer, employees demonstrate that they are actively listening and engaging with the complaint, which should de-escalate the negative arousal of the customer and increase their gratitude. Therefore,

> **H_2:** When complaining customers express relatively high negative arousal in a social media context, providing more (vs. less) active listening increases the likelihood of gratitude for the firm response.

De-escalating high negative arousal with empathy. Empathy is another key technique identified by the crisis negotiation literature for reducing arousal (McMains and Lanceley 2003). A negotiator who exhibits genuine concern and outwardly demonstrates emotional involvement with a situation can more effectively de-escalate negative, high-arousal emotions (Van Hasselt, Romano, and Vecchi 2008). Recent marketing research conceptualizes empathy as a response to another person's situation that is marked by the ability to feel warmth, compassion, and concern for others, as well as understanding the other person's cognitive-emotional experience as if it were affecting the observer directly (Allard, Dunn, and White 2020). For example, the phrase "I can imagine how difficult that situation was" highlights that employees are putting themselves in customers' situation (Pedersen 2021). Thus, in text-based service recoveries on social media, firms can use such content in their response to demonstrate an understanding of how the customer is feeling through words of validation and affirmation in order to express empathy.

Acknowledging the emotions of a failed customer is effective in de-escalating arousal because the customer feels understood and emotionally supported (Nguyen and McColl-Kennedy 2003). Recent research explicitly notes that conversational agents can reduce negative arousal among counterparts by expressing greater empathy (Chin, Molefi, and Yi 2020). When an empathetic firm response de-escalates a customer's state of high arousal, the customer should experience gratitude toward the firm response. We expect:

> **H_3:** When the complaining customer expresses relatively high negative arousal in a social media context, providing more (vs. less) empathy increases the likelihood of gratitude for the firm response.

We test our predictions with three field studies, complemented by preregistered controlled experiments to rule out potential

[2] Before these models emerged, the focus in crisis negotiation literature tended to be on problem-solving approaches, rather than de-escalating first. Only in 1997 did the FBI Crisis Negotiation Unit recognize the importance of de-escalating negative, high-arousal emotions before problem-solving can begin (Noesner and Webster 1997).

Table 2. Overviews of Studies.

Study	Details	Results
Study 1a	*Preliminary test of all hypotheses in the field and dictionary development*: 682 service recovery interactions from the Facebook pages of 30 German service firms.	*H_1, H_2, and H_3 are supported.* Inverted U-shape effect of active listening and no effect of empathy for low-arousal complaints.
Study 1b	*Split active listening and empathy and test their independent effects on arousal reduction*: Preregistered study with a single-factor design (control vs. high active listening vs. high empathy) with 315 U.S. participants.	*De-escalating effects of both active listening and empathy.*
Study 2	*More rigorous test of all hypotheses in the field (including robustness tests)*: All 5,068 service recovery interactions from the Facebook pages of a *Fortune* 500 U.S. airline during one year.	*H_1, H_2, and H_3 are supported.* Inverted U-shape effect of active listening and no effect of empathy for low-arousal complaints.
Study 3a	*Testing the linear and nonlinear effects of active listening while keeping empathy constant*: Preregistered study with a 2 (low and high arousal) × 3 (low, moderate, and high active listening) design with 850 U.S. participants	*H_1 and H_2 are supported.* Inverted U-shape effect of active listening for low-arousal complaints.
Study 3b	*Testing the linear and nonlinear effects of empathy while keeping active listening constant*: Preregistered study with a 2 (low and high arousal) × 3 (low, moderate, and high empathy) design with 851 U.S. participants.	*H_1 and H_3 are supported.* Positive effect of empathy for low-arousal complaints.
Study 4	*Generalization of observed effects to product failures and different social media channel*: 564 service recovery interactions from the Twitter account of a leading U.K. retailer.	*H_1, H_2, and H_3 are supported.* Inverted U-shape effect of active listening and positive effect of empathy for low-arousal complaints.

endogeneity issues due to omitted variables, strategic behavior, and unobserved heterogeneity. Table 2 provides an overview of studies and hypothesis testing.

Study 1a: Field Study of Text-Based Service Recoveries

Data Collection

Study 1a uses field data from public service interactions to examine our hypotheses. Four research assistants manually collected complaints after service failures from the Facebook pages of 30 German service firms, representing five industries (couriers, hospitality, insurance, telecommunications, and transportation). They were instructed to extract the data from every fifth customer post if it met three criteria: (1) was a complaint directed at the firm, (2) to which the firm responded, after which (3) the complaining customer responded. Thus, each of the 682 service interactions extracted consisted of a sequence of at least three posts.[3] In addition to extracting the text, the research assistants noted the response time and number of reactions from other users (likes, comments, or shares). Most interactions were in German (97%) and were translated into English to enable the use of established linguistic dictionaries. Our measurement approach is in Figure 1, and a summary of the measures are reported in Table 3.

Construct Measures

Customer complaint. To determine the negative arousal expressed by a complaining customer, we use the dictionaries created by Villarroel Ordenes et al. (2017), which refer to four categories of words differentiated by valence (positive vs. negative) and activation (high vs. low). We capture negative high arousal as the percentage of negative, relatively high-arousal words in the customer complaint (e.g., "furious," "outrage").[4]

Firm response. We rely on written features to operationalize active listening and empathy. To discriminate the two constructs, we relate active listening to the writing style and empathy to the content of the firm response. Previous research used linguistic style matching to measure how actively engaged conversation partners are with one another in an interaction (Niederhoffer and Pennebaker 2002). Thus, we captured active listening with the linguistic style matching (LSM) measure from Gonzales, Hancock, and Pennebaker (2010), as detailed in Table 4.

In line with Herhausen et al. (2019), we derived the degree of LSM between the firm response fr at time 2 with the customer complaint cc at time 1 in three steps. First, we mined the use intensity of each of the nine function word categories FWj separately in the complaint and the firm response. Second, the

[3] If a longer exchange occurs between the customer and firm, we only consider the first firm response, reasoning that if the initial response is unsuccessful, other consumers may support the complaint or be negatively affected, with negative consequences for the firm (Herhausen et al. 2019; Hogreve and Hoerner 2019).

[4] Given the interchangeable use of terms "arousal" (Russell 1980) and "activation" (Russell and Barret 1999) in the circumplex model, we deem activation a reasonable proxy for arousal. The four dictionaries have been validated with the dictionary of affect from Whissell (2009). In a different approach to measure arousal and valence, Warriner, Kuperman, and Brysbaert (2013) establish norms for valence and arousal across 13,915 English words. A total of 607 words are in common with our dictionaries. The 272 negative high-arousal words received a mean arousal rating of 5.02 and a mean valence rating of 3.22; the 95 negative low-arousal words received a mean arousal rating of 4.21 and a mean valence rating of 3.25 (nine-point scales). Thus, we examine relatively higher and lower arousal nested around the midpoint of the arousal scale.

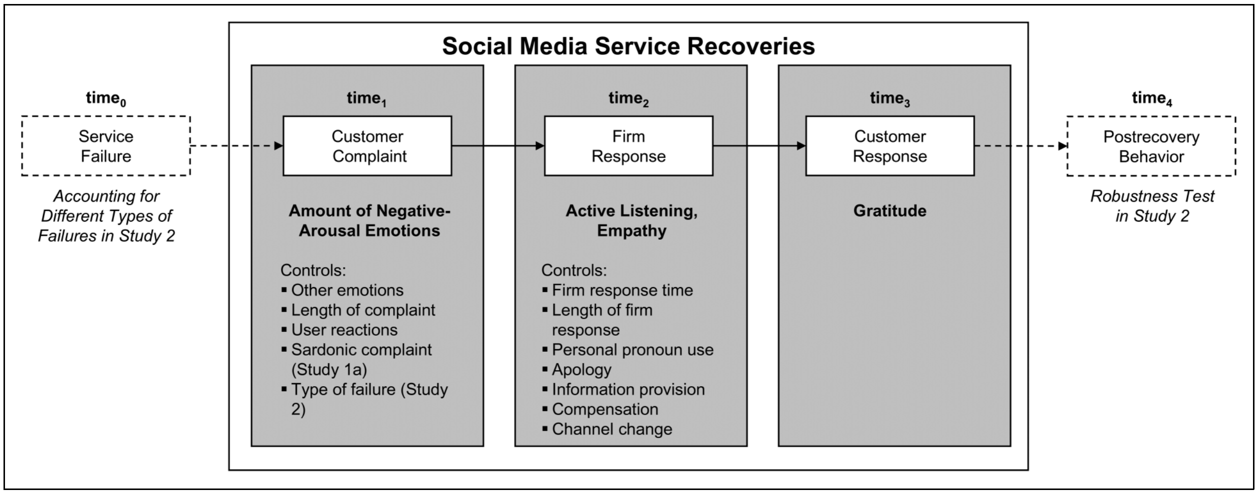

Figure 1. Measurement Approach in the Field Studies.
Notes: We are unable to observe the service failure that prompted customers to file a complaint on the social media page. However, we use an LDA to account for different types of failures in Study 2 (see Web Appendix H).

degree of LSM of each function word category j in the firm response comes from the formula:

$$\text{LSMj}_{\text{fr}} = 1 - \left[\frac{|\text{FWj}_{\text{cc}} - \text{FWj}_{\text{fr}}|}{\text{FWj}_{\text{cc}} + \text{FWj}_{\text{fr}} + .0001}\right]. \quad (1)$$

Third, by aggregating all nine LSM scores with equal weights, we obtain an overall LSM score bound between 0 and 1. A score closer to 1 reflects a greater degree of active listening in the firm response. For example, consider the following stylized complaint with a high percentage of personal pronouns that signal informality (Tausczik and Pennebaker 2010) and a low percentage of negations that signal simplicity (Chung and Pennebaker 2007) of the complaining customer.

> Complaint*:* **I** am angry about **your** service. **You** changed the departure time of **my** flight on short notice which caused **me** to miss **my** next flight. Now **I** am stranded at the airport. Please get back to **me**. **We** need to find a solution for **my** problem!

To signal active listening, the service employee can match the linguistic style of the customer by also using personal pronouns and avoiding negations (LSM score = .66):[5]

> Response 1: **I** am sorry about **your** experience. It is unfortunate that [Airline] had to change the departure time and missed to inform **you** on time. Please send **me** a private message and provide **me** with more detail about **your** situation. **I** will find a solution for **your** problem.

In contrast, using a similar number of words and providing the same content, the following response does not signal active listening because the linguistic style of the customer is not matched (i.e., more impersonal pronouns and more negations; LSM score = .36):

> Response 2: [Airline] is sorry about **this** experience. It is **not** ok that the departure time had to be changed and [Airline] did **not** send **this** information on time. Please send a private message and provide more detail about **this** situation. [Airline] will find a solution for **the** problem.

We validate our operationalization of active listening with LSM in Study 1b, where participants perceive responses with higher LSM as more indicative of active listening.

For empathy, we developed a new dictionary that measures all words that indicate an employee's effort to connect emotionally with the complainant by expressing an understanding of their emotional state, using words of validation and affirmation. We drew terms from previous studies of empathy in communication research (Alam, Danieli, and Riccardi 2018), as well as from the firm responses in our data sets, to create a preliminary word list. We added synonyms, avoided homonyms and words that are too general, and ensured context specificity to social media interactions. The initial word lists consisted of 144 words. Three expert judges evaluated this list to check the relevance of each word to the construct definitions and noted additional words to be included. We then assessed interjudge consistency, discussed diverging opinions, and only retained words consistently evaluated as relevant to reflect empathy. We checked for internal and external validity, as summarized in Table 5. The final dictionary consists of 112 words/word combinations/word stems and is also displayed in Table 5.

Customer response. Gratitude is defined as expressing emotional appreciation for the firm response. We measure whether customers show gratitude or not (1/0) in their responses, using a new dictionary that relies on terms identified in previous communication studies of gratitude. We added synonyms, avoided homonyms

[5] Response 1 and Response 2 score similarly on empathy (both 2.13). Details are provided in Table 4.

Table 3. Measurements in the Field Studies.

Variable	Operationalization	Source
Customer Complaint		
Negative high arousal	Dictionary capturing relatively high arousal negative words (percentage of matching words in the complaint, e.g., "furious," "outrage"). Validation with perceptual ratings in Study 1b.	Villarroel Ordenes et al. (2017)
Firm Response		
Active listening	Linguistic style matching in nine categories of function words of the firm response with the customer complaint: • articles (e.g., a, an, the) • auxiliary verbs (e.g., am, be, will) • conjunctions (e.g., and, but, whereas) • high-frequency adverbs (e.g., rather, really, very) • impersonal pronouns (e.g., it, that, those) • negations (e.g., no, not, never) • personal pronouns (e.g., I, you, we) • prepositions (e.g., at, in, into) • quantifiers (e.g., few, much, some) Validation with perceptual ratings in Study 1b.	Gonzales, Hancock, and Pennebaker (2010)
Empathy	Dictionary capturing words that reflect empathy (percentage of matching words in the firm response). Validation with manual coding in Study 1a and perceptual ratings in Study 1b.	New measure
Customer Response		
Gratitude	Dictionary capturing words that reflect gratitude in the customer response (dummy coded). Validation with manual coding in Study 1a and a machine learning classifier in Study 2.	New measure
Control Variables		
Negative low arousal	Dictionary capturing relatively low-arousal negative words (percentage of matching words in the complaint; e.g., "disappointed," "unhappy")	Villarroel Ordenes et al. (2017)
Positive high arousal	Dictionary capturing relatively high arousal positive (percentage of matching words in the complaint; e.g., "energy," "outstanding")	Villarroel Ordenes et al. (2017)
Positive low arousal	Dictionary capturing relatively low-arousal positive words (percentage of matching words in the complaint; e.g., "content," "nice")	Villarroel Ordenes et al. (2017)
Length of complaint	Number of characters in the customer complaint	Berger and Milkman (2012)
User reactions	Proxy for general support of other users (number of likes, comments, and shares on the complaint)	Schaefers and Schamari (2016)
Sardonic complaint	Two coders rated whether the customer used irony or sarcasm in the complaint (dummy coded)	Johnen and Schnittka (2019)
Firm response time	Time stamp of customer complaint minus time stamp of firm response (converted to hours)	Homburg, Ehm, and Artz (2015)
Length of firm response	Number of characters in the firm response	Berger and Milkman (2012)
Personal pronoun use	Use of "I" versus "we" in the firm response	Packard, Moore, and McFerran (2018)
Apology	Dictionary capturing words that reflect an apology in the firm response (dummy coded)	Herhausen et al. (2019)
Information provision	Dictionary capturing "cognitive processes" in the firm response (percentage of matching words)	Herhausen et al. (2019)
Compensation	Dictionary capturing words that reflect a compensation in the firm response (dummy coded)	Herhausen et al. (2019)
Channel change	Dictionary capturing words that reflect evoking a channel change in the firm response (dummy coded)	Herhausen et al. (2019)

Table 4. Measurement Details on Active Listening.

Functional Words and Their Linguistic Meaning	
Our operationalization of active listening relates to the matching of functional words. In contrast with nonfunction words (e.g., nouns, verbs, adjectives), which convey content, function words reflect linguistic style, set the tone for social interactions, and are key to understanding the relationship among communication partners. Previous literature has related the nine function word categories to different communication styles, such as formal versus informal language (i.e., articles, auxiliary verbs, impersonal pronouns, personal pronouns, prepositions), complex versus simple language (i.e., conjunctions, negations), and concrete versus abstract language (i.e., high-frequency adverbs, quantifiers), as summarized below.	
Functional Words	**Linguistic Meaning**
Articles	Formal/informal: more articles = more formal (Boyd, Blackburn, and Pennebaker 2020)
Auxiliary verbs	Formal/informal: more auxiliary verbs = more informal (Boyd, Blackburn, and Pennebaker 2020)
Conjunctions	Complexity/simplicity: more conjunctions = more complex (Chung and Pennebaker 2007)
High-frequency adverbs	Concrete/abstract: More adverbs = more concrete (Packard and Berger 2021)
Impersonal pronouns	Informal/formal: more impersonal pronouns = more formal (Tausczik and Pennebaker 2010)
Negations	Complexity/simplicity: more negations = more complex (Chung and Pennebaker 2007)
Personal pronouns	Informal/formal: more personal pronouns = more informal (Boyd, Blackburn, and Pennebaker 2020)
Prepositions	Formal/informal: More prepositions = more formal (Boyd, Blackburn, and Pennebaker 2020)
Quantifiers	Concrete/abstract: more quantifiers = more concrete (Qiu et al. 2012)

Measurement of Active Listening in the Stylized Examples

Complaint: **I am** angry **about your** service. **You** changed **the** departure time **of my** flight **on** short notice **which** caused **me to** miss **my** next flight. **Now I am** stranded **at the** airport. Please get **back to me**. **We** need **to** find **a** solution **for my** problem!

Response 1: **I am** sorry **about your** experience. **It is** unfortunate **that** [Airline] **had to** change **the** departure time **and** missed **to** inform **you on** time. Please send **me a** private message **and** provide **me** with **more** detail **about your** situation. **I will** find **a** solution **for your** problem.

Response 2: [Airline] **is** sorry **about this** experience. **It is not** ok **that the** departure time **had to be** changed **and** [Airline] **did not** send **this** information **on** time. Please send **a** private message **and** provide **more** detail **about this** situation. [Airline] **will** find **a** solution **for the** problem.

	Complaint	**Response 1**	**Response 2**
Word count	46	47	47
Articles (e.g., a, the)	6.52	6.38	8.51
Auxiliary verbs (e.g., am, are, be, did, had, is, will)	4.35	8.51	12.77
Conjunctions (e.g., and)	.00	4.26	4.26
High-frequency adverbs (e.g., about, back, now)	6.52	4.26	4.26
Impersonal pronouns (e.g., it, that, this, which)	2.17	4.26	10.64
Negations (e.g., not)	.00	.00	4.26
Personal pronouns (e.g., I, me, my, us, we, you, your)	21.74	17.02	.00
Prepositions (e.g., about, at, for, of, on, to, with)	17.39	14.89	10.64
Quantifiers (e.g., more)	.00	2.13	2.13
Active listening	—	.66	.36
Empathy	—	2.13	2.13

Notes: The word "about" is in the LIWC dictionaries of both "high-frequency adverbs" and "prepositions."

and general terms, and ensured context specificity to social media interactions. The initial list of 13 gratitude words, as well as emoticons, were evaluated by three expert judges, and we checked the extended list for internal and external validity. The final dictionary in Table 5 consists of 20 words/word combinations/word stems/emoticons. We validated the new dictionary for gratitude with manual coding in Study 1a and a machine learning classifier in Study 2 (see Web Appendix B).

Table 5. New Dictionaries for Empathy and Gratitude.

Dictionary Validation

Type of Validity	**Validation Procedure**
Construct validity: Does the text represent the theoretical concepts?	Following the "empirically guided" approach, we created initial word lists reflecting the construct definitions directly from our data sets.
Concurrent and convergent validity: Does the researcher's measurement of the constructs relate to other measurements?	Our measurement indicates concurrency with human ratings of empathy (intercoder agreement = .87, r_{Study_1a} = .56), human ratings of gratitude (intercoder agreement = .95, r_{Study_1a} = .79), and machine learning classification of gratitude (r_{Study_2} = .67).
Causal validity: Is the construct in the data set causally related to other constructs?	We have natural time lags between variables and include several control variables in the model that address rival hypotheses (e.g., different service recovery strategies that may lead to gratitude).
Predictive validity: Does the construct have the expected effects on a meaningful variable?	Across multiple data sets and different measurement approaches, we confirm the theoretically derived relationship between empathy and gratitude in the field studies.
Generalizability: Are results based on multiple data sets?	The measurement and results are replicated with three independent samples (i.e., service interactions from 30 German service firms, service interactions from a *Fortune* 500 U.S. service firm, and interaction from a leading U.K. retailer).
Robustness: Is more than one method used?	We replicate the focal relationships in a controlled experimental setting where we manipulate empathy and measure gratitude (Study 3b).

New Dictionaries

Empathy

Empathy is defined as connecting emotionally with complaining customers by indicating understanding of their feelings, using explicit expressions of validation and affirmation.

Word List: admir*, affection*, appreciat*, assur*, better, care, careful*, caring, challenging, comfort*, commitment*, confiden*, considerate, contact*, contented*, courag*, determin*, devot*, difficult, discourag*, divin*, eager*, encourag*, engag*, entertain*, enthus*, excel*, excit*, experience, faith*, favor*, favour*, feedback, feel*, fix*, forgiv*, frustrat*, gentle*, gently, glad, gladly, gratef*, grati*, happen*, hear, hearing, heartwarm*, help*, honest, honest*, honor*, honour*, hope*, hoping, imagine, improve*, improving, keen*, kind*, know*, look, "make it better," "makes me really sad," "makes me sad," "my mistake," notify*, open*, openness, "our mistake," patience, peace*, perfect*, personally, pleas*, precious*, promis*, relief, reliev*, resolv*, respect, safe*, same, satisf*, save, sense, share*, sharing, similar, sincer*, sound*, support*, sympath*, "tell me," thank*, thoughtful*, touch, true*, truly, trust*, understand*, upset, useful*, valuabl*, value*, valuing, welcom*, wish, worthwhile, "you are right," "you are totally right," "you're right," "your position"

Gratitude

Gratitude is defined as expressing emotional appreciation for the firm response to the complaint.

Word List: acknowledg*, appreciat*, awesome, glad, grateful, gratitude, happy, *heart emoticon*, helpful*, kind, like, "means a lot to me," sincerely, *smiley emoticon*, super, thank*, thanx, *wink emoticon*, "you are awesome," "you made my day"

Controls. We account for a potential nonlinear effect of negative high-arousal words because previous research indicated an inverted U-shaped relationship between expressed emotional arousal and reader perceptions (Yin, Bond, and Zhang 2017). Moreover, we control for negative low-arousal words (e.g., "disappointed," "unhappy"), positive high-arousal words (e.g., "energy," "outstanding"), and positive low-arousal words (e.g., "content," "nice") and for other potential determinants of gratitude, such as the length of the complaint (which may signal more engagement; Berger and Milkman 2012), the number of reactions to the complaint (which may affect recoveries; Schaefers and Schamari 2016), and whether a complaint is sardonic (which often signals no interest in a recovery; Johnen and Schnittka 2019).

We also captured firm response time (faster responses are perceived more favorably; Homburg, Ehm, and Artz 2015), the length of the response (which may signal more effort by the firm; Berger and Milkman 2012), and personal pronouns in the firm response, because using "I" rather than "we" may increase perceptions that the employee feels and acts on the behalf of the customer (Packard, Moore, and McFerran 2018). We further controlled for different recovery strategies (i.e., apology, compensation, information provision, and channel change; Herhausen et al. 2019). We address unobserved firm heterogeneity that might arise due to different social media guidelines with fixed effects (e.g., some firms may ask employees to use scripted answers). The descriptive statistics and correlations for Study 1a are in Web Appendix C.

Methodological Approach

We analyzed the data using binary logistic regression with gratitude as our dependent variable—that is, a successful or

Table 6. Predicting Gratitude in Study 1a.

	Model 1: Main Effects			Model 2: Full Model		
	OR	SE	95% CI	OR	SE	95% CI
Controls						
Negative low arousal	1.201**	.110	1.003, 1.438	1.210**	.113	1.008, 1.452
Positive high arousal	1.001	.057	.896, 1.118	1.004	.057	.898, 1.122
Positive low arousal	1.032	.051	.936, 1.138	1.031	.052	.934, 1.138
Length of complaint	1.000	.001	.997, 1.003	1.000	.001	.997, 1.003
User reactions	.978	.020	.939, 1.019	.983	.023	.939, 1.028
Sardonic complaint	.546	.244	.227, 1.311	.588	.268	.241, 1.434
Firm response time	1.000	.000	1.000, 1.000	1.000	.000	1.000, 1.000
Length of firm response	.992**	.004	.985, .999	.992**	.004	.984, .999
Personal pronoun: "I"	.986	.045	.902, 1.078	.974	.045	.888, 1.067
Personal pronoun: "We"	.938**	.030	.881, .998	.932**	.031	.874, .994
Channel change	.810	.171	.536, 1.224	.797	.178	.514, 1.235
Compensation	2.052**	.646	1.107, 3.802	1.983	.632	1.062, 3.704
Information provision	.988	.020	.949, 1.028	.985	.021	.945, 1.026
Apology	1.377	.306	.892, 2.127	1.460	.337	.928, 2.296
Customer Complaint						
Negative high arousal (NHA)	.904*	.050	.810, 1.008	.811**	.075	.671, .965
NHA squared				1.011*	.007	.996, 1.026
Firm Response						
Active listening	1.013**	.007	1.000, 1.027	1.012*	.008	.998, 1.027
Active listening squared				.999**	.000	.998, 1.000
Empathy	.971	.027	.919, 1.025	.980	.036	.927, 1.069
Empathy squared				.997	.005	.987, 1.007
Interaction Effects						
NHA × Active listening				1.013**	.005	1.003, 1.023
NHA × Empathy				1.035*	.021	.995, 1.077
Active listening × Empathy				.997	.002	.993, 1.001
Fixed Effects						
Firms		included			included	
Log-likelihood	−394.090			−386.989		
Number of observations		682			682	

*$p < .10$.
**$p < .05$.
***$p < .01$.
Significance is based on two-tailed tests.
Notes: CI = confidence interval. We use robust standard errors in our estimation to account for clustered observations. The results of a robustness test with manual coded empathy are available from the authors. The odds ratios refer to a 1% change in negative high arousal, active listening, and empathy, to ease interpretation.

unsuccessful firm response to a complaint. The model includes linear and squared main effects of active listening and empathy, as well as interaction effects of active listening and empathy with negative high arousal. Moreover, we account for a potential interaction between active listening and empathy:

$$\begin{aligned} GRA_{t3} = {} & \beta_0 + \beta_1 NHA_{t1} + \beta_2\, AL_{t2} + \beta_3 EMP_{t2} + \beta_4 NHA_{t1}^2 \\ & + \beta_5 AL_{t2}^2 + \beta_6 EMP_{t2}^2 + \ \beta_7 NHA_{t1} \times AL_{t2} \\ & + \ \beta_8 NHA_{t1} d \times EMP_{t2} + \ \beta_9 AL_{t2} \times EMP_{t2} \\ & + \beta_{10-15} CCC_{t1} + \beta_{16-23} FRC_{t2} + \beta_{24-52} FFE + \varepsilon, \end{aligned} \tag{2}$$

where GRA_{t3} is gratitude in the customer response; NHA_{t1} is negative high arousal in the customer complaint; AL_{t2} is active listening and EMP_{t2} is empathy, exhibited in the firm response; CCC_{t1} is a vector of customer complaint controls; FRC_{t2} is a vector of firm response controls; FFE is a vector of firm fixed effects; and ϵ is the error term.

Results and Discussion

Table 6 contains the results for Study 1a. The odds ratios (OR) for negative high arousal in the full model indicate a negative linear effect (OR = .811, $p < .05$), in support of H_1. In addition, we find a positive quadratic effect (OR = 1.011, $p < .05$) on the probability of gratitude after the firm response. The turning

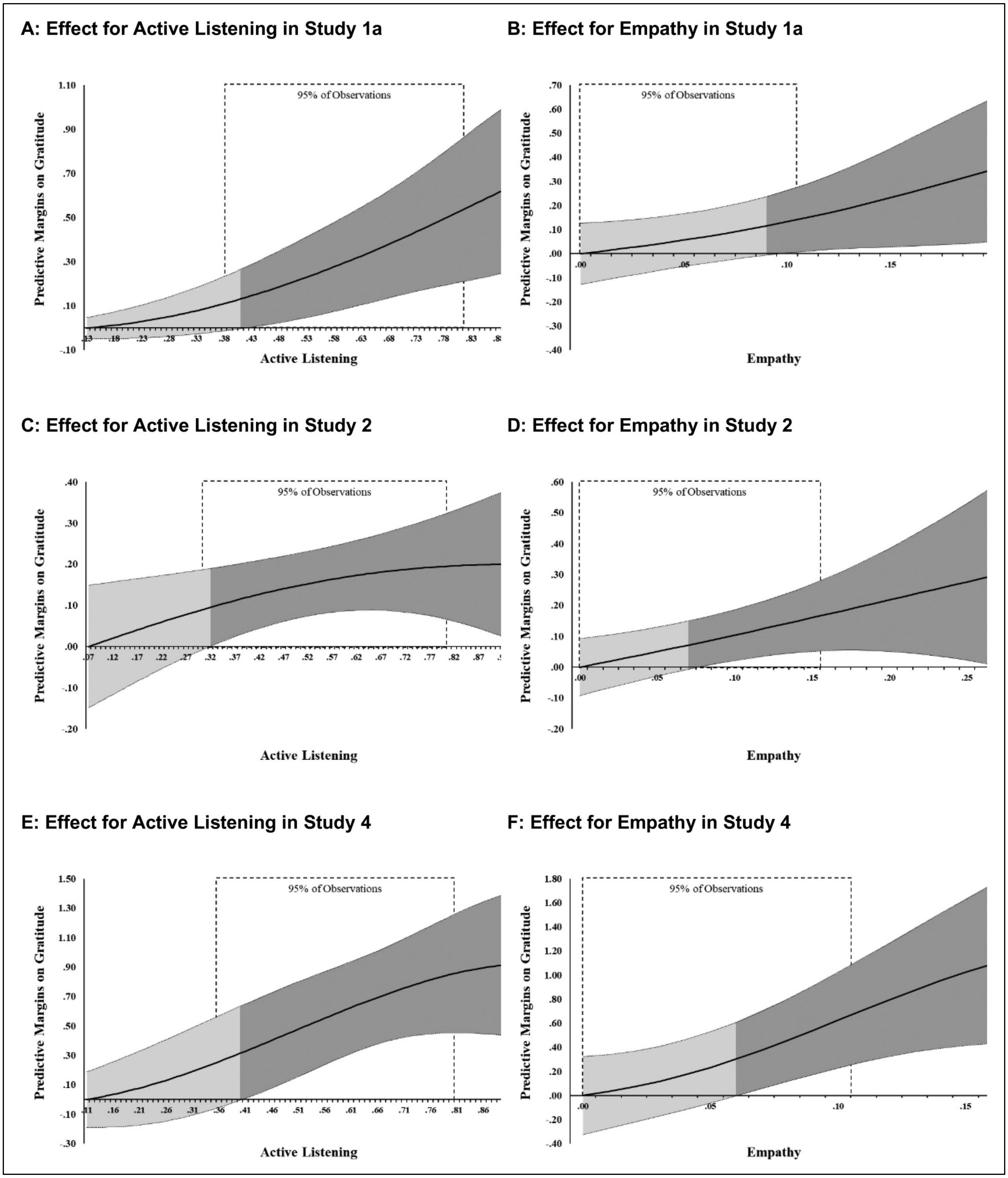

Figure 2. Effects of Active Listening and Empathy for High-Arousal Complaints.
Notes: We display predicted marginal effects of active listening and empathy on gratitude with 95% confidence intervals for high-arousal complaints (5% of negative high-arousal words). Dark gray area = effect is significant at $p < .05$ (two-tailed); light gray area = effect is not significant. Results for low-arousal complaints (0% of negative high-arousal words) are displayed in Web Appendix E.

point of this decreasing relationship occurs at 10% negative high-arousal words, which represents more than 99.3% of the observations in our data set. Furthermore, we find a positive and marginally significant effect of active listening (OR = 1.011, $p < .10$), a negative effect of active listening squared (OR = .999, $p < .05$), and positive interaction effects of both active listening (OR = 1.013, $p < .05$) and empathy (OR = 1.035, $p < .10$) with negative high arousal on the probability of gratitude.

We focus the predicted margins of active listening and empathy on 5% negative high-arousal words, because this signals high arousal, and on 0% negative high-arousal words, because this signals the absence of negative high-arousal emotions (i.e., low-arousal complaints).[6] We display these effects in Figure 2 and Web Appendix E. For high arousal, we find positive effects of active listening (OR = 1.170, $p < .01$) and empathy (OR = 1.382, $p < .01$), in support of H_2 and H_3. With a 1% increase in active listening (empathy), we observe up to a 17% (38%) increase in the probability of gratitude.[7] Thus, providing empathy is more effective than offering active listening. For low arousal, we find an inverted U-shaped relationship between active listening and the probability of gratitude, with a maximum when active listening takes a value of .66. We find that 67% of firm responses are below, 1% match, and 32% exceed this turning point. We find no significant effect of empathy for low-arousal complaints.

In summary, we find support for H_1–H_3. However, the field data only provide the effects of de-escalation strategies on gratitude, and we were unable to measure the actual de-escalation evoked by active listening and empathy (i.e., decrease in customer arousal before and after a firm intervention). Using a controlled experiment in Study 1b, we explicitly measure de-escalation.

Study 1b: De-Escalation of Negative High Arousal

Can providing active listening or empathy without a service recovery really de-escalate high negative arousal? We explicitly test this assumption in Study 1b, in which we independently manipulate the degree of active listening or empathy an employee provides after a complaint while not providing any actual service recovery. Thus, we measure negative arousal twice—after the failure but before the firm response, and after the firm response—to isolate the de-escalating effects of active listening and empathy. To enhance realism, we use a frequent social media response strategy: directing a complaining customer to a private conversation without recovering the failure. While such a strategy might minimize the negative perceptions other consumers and investors from reading the exchange, it runs the risk of further upsetting an already angry customer. Thus, responding with a forced channel change provides a good context to explore the de-escalating effects of active listening and empathy without providing an actual recovery.

Design and Procedure

Study 1b was a preregistered study on Amazon Mechanical Turk (MTurk) run in conjunction with CloudResearch (https://aspredicted.org/yf8bg.pdf). In line with the preregistration, eight participants were excluded because they wrote nonsense in response to open-ended questions about the scenarios and/or because they did not pass manipulation and attention checks. Thus, we have a final sample of 315 U.S. participants (M_{age} = 42.0 years, 51% women). All participants were exposed to the same negative, high-arousal scenario, then were randomly assigned to one of three employee response conditions in a single-factor between-subjects design (control vs. high active listening vs. high empathy). None of these conditions features any service recovery. Web Appendix G provides the experimental stimuli.

Participants first read about a customer's experience with an airline, and the scenario prompted them to imagine they had experienced it, such that they had to deal with several incidents, including a delayed flight and a missing suitcase. Participants then responded to an open question by detailing how they would feel after these service failures. Then in randomized order, they indicated their valence (1 = "very unpleasant," and 7 = "very pleasant"), arousal (1 = "very mellow," and 7 = "very fired up"), and anger ("I would feel angry with the airline"; 1 = "strongly disagree," and 7 = "strongly agree") related to the airline experience.

In the second part of the survey, participants read a high-arousal complaint they might have written on the airline's Facebook site, based on their experience, which contained 5% negative, high-arousal words to match the high-arousal slope in Study 1a. On the following page, they read that after about 10 minutes, they received an employee response: a control response, a high-active-listening response, or a high-empathy response. We designed all responses in line with the Study 1a measures (control: active listening = .72, empathy = 1.74; high active listening: active listening = .86, empathy = 1.74; high empathy: active listening = .73, empathy = 5.22).[8]

Participants then responded to the same questions from the first part of the survey (i.e., open question and valence, arousal, and anger measures). After completing attention and manipulation checks, participants reported whether they believed the airline response was written by a human or bot (1 = "human," and 7 = "bot"), their frequency of travel on airlines in the past five years (1 = "not at all," and 5 = "a great deal"), how often they experience service failures (1 = "not at all," and 7 = "a great deal"), and demographics such as age, gender, and language.[9]

Results and Discussion

Results. We employed a mixed design, such that participants answered dependent variables twice (once after the high-arousal scenario, prior to the employee response, and once after the employee response). To test whether both active listening and empathy effectively attenuate high negative arousal in text-

[6] We display response surfaces that span the whole range of negative high-arousal words in Web Appendix D.

[7] We report a robustness test related to the matching of personal pronouns in Web Appendix F.

[8] We confirmed the differences in a pretest (N = 143). The active listening condition was higher (M = 6.24, SD = .87) on active listening than both the high-empathy (M = 5.30, SD = 1.55; $p = .002$) and control (M = 3.14, SD = 1.73; $p < .001$) conditions, and the high-empathy condition was higher on empathy (M = 6.38, SD = .610) than the high-active-listening (M = 4.22, SD = 1.97) or control (M = 2.88, SD = 1.78; both $p < .001$) conditions.

[9] None of these covariates affected arousal (all $ps > .14$).

based recoveries, we compare the difference score between the measurements of arousal at time 1 and time 2 (i.e., a pre- and postemployee response).

A between-subjects analysis of variance (ANOVA) on the difference score of arousal is significant ($F(2, 312) = 3.78$, $p = .024$). When the employee engages in high active listening, participants indicate a significant reduction in arousal from time 1 to time 2 ($M_{\Delta\_arousal} = -1.13$, $SD = 1.60$) compared with when the employee uses a control response ($M_{\Delta\_arousal} = -.726$, $SD = 1.46$, $p = .048$). A similar pattern arises when the employee exhibits high empathy (cf. control response; $M_{\Delta\_arousal} = -1.27$, $SD = 1.40$, $p = .009$). We find no significant difference between active listening and empathy ($p = .55$); both are effective strategies for reducing arousal, independent of each other and compared with more neutral responses.[10]

Study 1b provides evidence of the de-escalating effects of both active listening and empathy after text-based complaining. Specifically, providing either high active listening or high empathy (vs. a more neutral control response), even without providing an actual recovery, lowers the negative arousal of complaining customers.

Study 2: Field Study of a *Fortune* 500 Firm

Study 1a was based on a sample of customers of German service firms and Study 1b was a controlled experiment to examine high-arousal de-escalation. To further our understandings and findings from these data in the field, in Study 2, we analyze text-based service interactions between a *Fortune* 500 airline and its customers, following a rigorous sampling process and accounting for potential selection biases and complaint heterogeneity. We also capture behavioral outcomes among a subsample of these complaining customers.

Data Collection and Measurement

We collected all publicly available digital service interactions from the international Facebook page of a *Fortune* 500 airline for a period of 12 months, from August 2015 until July 2016. During this time, 18,576 complaints appeared on the site,[11] and the airline responded to 9,642 of these. In 5,068 cases, the complaining customer reacted to the firm response, and this set represents our final sample. We used the same measures as in Study 1a. The descriptive statistics and correlations are reported in Web Appendix C.

Self-Selection of Firm and Customer Responses

Because we need information about the customer complaint, the firm response, and the customer response, the sample is limited to 5,068 cases with complete service interactions. Our estimates thus might be biased by two self-selection processes. First, social media employees of the firm decide to respond publicly to a complaint or not. Second, the customer decides to react publicly to the firm's response or not. The potential selection biases might stem from observable factors (e.g., textual features of the complaint) or unobservable factors (e.g., employee workload, availability of the customer). Therefore, we employed two-stage Heckman selection models, in which we estimated the availability of (1) the firm response and (2) the customer response as binary dependent variables in the first stage. Then we computed inverse Mills ratios (IMRs) to account for potential selection biases in the second stage. To avoid the possibility that the IMRs reflect a linear combination of the regressors in the main analysis, we also need variables that satisfy the requirements of relevance and exogeneity for the Heckman correction. We use the daily average response rate to all posts other than the focal complaint (ranging from 0% to 100% in our data). It might be influenced by the availability of frontline employees or technical issues, so it should influence the firm's choice to respond to the focal post but not gratitude as a reaction to that response. We confirm this notion by finding a nonsignificant relationship between the daily average response rate and gratitude ($r = .01$, n.s.). To address the customer's choice, we use the frequency of complaining (i.e., number of times a customer posts to the site during the observation period), which indicates a general tendency to interact with the firm, regardless of a sense of gratitude. In this case too, we find a nonsignificant relationship between frequency of complaining and gratitude ($r = -.02$, n.s.; see Web Appendix H).

Heterogeneity of Service Failures

The type of failure might influence a customer's gratitude to the firm's response. For example, an angry customer might be harder to calm down with active listening and empathy if the issue is inappropriate employee behavior (Belanche et al. 2020). We use latent Dirichlet allocation (LDA) to account for different failure types (Ludwig et al. 2022), as detailed in Web Appendix H. Thus, we measure word occurrences across complaints, identify latent topics based on these words, calculate the probability that each word appears in a given topic, and assign all complaints to their most likely topic (i.e., different types of failures). The LDA indicates the best fit for eight different failure types in our data ("payment," "boarding," "upgrade," "family issues," "missed flight," "luggage," "employees," and "refund"), which we include as controls.

Results and Discussion

Table 7 displays the results. Negative high arousal exerts a negative linear effect ($OR = .897$, $p < .01$) and a positive quadratic effect ($OR = 1.002$, $p < .05$). The turning point of this decreasing relationship is at 28% negative, high-arousal words, which accounts for more than 99.9% of the observations in our data set, in support of H_1. In this airline context, greater negative

[10] We report the analyses for anger and valence in Web Appendix G.

[11] We consider only posts with more negative than positive words while accounting for negations.

Table 7. Predicting Gratitude in Study 2.

	Model 3: Main Effects			Model 4: Full Model		
	OR	SE	95% CI	OR	SE	95% CI
Controls						
Negative low arousal	1.022	.023	.978, 1.068	1.024	.023	.980, 1.071
Positive high arousal	1.046**	.019	1.009, 1.084	1.045**	.019	1.008, 1.083
Positive low arousal	1.026	.019	.989, 1.064	1.027	.019	.991, 1.065
Length of complaint	1.000	.000	1.000, 1.001	1.000	.000	1.000, 1.001
User reactions	1.000	.000	.999, 1.000	1.000	.000	.999, 1.000
Firm response time	1.000	.000	1.000, 1.000	1.000	.000	1.000, 1.000
Length of firm response	1.011	.013	.986, 1.038	1.009	.013	.983, 1.036
Personal pronoun: "I"	1.021**	.008	1.005, 1.038	1.022**	.008	1.005, 1.038
Personal pronoun: "We"	.999	.002	.995, 1.003	1.000	.002	.996, 1.004
Channel change	1.320**	.178	1.013, 1.720	1.314**	.178	1.008, 1.714
Compensation	1.341**	.148	1.08, 1.664	1.340**	.148	1.079, 1.664
Information provision	1.000	.006	.988, 1.013	1.000	.006	.987, 1.012
Apology	1.210**	.098	1.032, 1.419	1.220**	.100	1.04, 1.432
Customer Complaint						
Negative high arousal (NHA)	.942**	.022	.898, .983	.897***	.023	.853, .944
NHA squared				1.002**	.001	1.001, 1.004
Firm Response						
Active listening	1.002	.003	.997, 1.007	1.001	.003	.995, 1.006
Active listening squared				.999*	.000	.999, 1.000
Empathy	1.010	.007	.992, 1.019	1.010	.008	.995, 1.025
Empathy squared				1.000	.001	.998, 1.002
Interaction Effects						
NHA × Active listening				1.003**	.001	1.000, 1.005
NHA × Empathy				1.010***	.003	1.003, 1.016
Active listening × Empathy				1.000	.000	.999, 1.001
Sample Selection Controls						
$IMR_{\text{firm response}}$	1.260	.313	.775, 2.051	1.280	.317	.787, 2.081
$IMR_{\text{customer response}}$	4.999**	3.529	1.253, 19.94	5.299**	3.783	1.307, 11.47
Heterogeneity Controls						
Payment-related failure	1.254**	.147	.997, 1.578	1.267**	.149	1.006, 1.595
Boarding-related failure	1.204*	.132	.970, 1.493	1.191*	.131	.960, 1.479
Upgrade-related failure	1.291**	.139	1.047, 1.594	1.283**	.138	1.039, 1.585
Family-related failure	1.080	.167	.797, 1.463	1.082	.168	.798, 1.468
Luggage-related failure	1.267**	.146	1.011, 1.588	1.255**	.145	1.001, 1.574
Employee-related failure	1.026	.120	.816, 1.291	1.040	.122	.826, 1.309
Refund-related failure	1.430***	.161	1.147, 1.782	1.413***	.159	1.133, 1.763
Log-likelihood		−3,157.320			−3,146.795	
Number of observations		5,068			5,068	

*$p < .1$.
**$p < .05$.
***$p < .01$.
Significance is based on two-tailed tests.
Notes: We use robust standard errors in our estimation to account for clustered observations. CI = confidence interval; IMR = inverse Mills ratio from the sample selection models. We use "missed flight-related failure" as the base category. The odds ratios refer to a 1% change in negative high arousal, active listening, and empathy, to ease interpretation.

arousal is more common than in Study 1a. We also find a negative effect of active listening squared (OR = .999, $p < .10$) and positive interaction effects of both active listening (OR = 1.003, $p < .05$) and empathy (OR = 1.010, $p < .01$) with negative high arousal.

We display marginal effects in Figure 2 and Web Appendix E. For high arousal, we find linear positive effects of active listening (OR = 1.033, $p < .05$) and empathy (OR = 1.109, $p < .01$), in support of H_2 and H_3. With a 1% increase in active listening (empathy), we observe a 3% (11%) increase in the probability of gratitude. Here again, empathy is more effective than active listening. For low arousal, we find an inverted U-shaped effect of active listening with a maximum at .58. The firm offered 58%

responses below, 2% matching, and 40% above this critical point. We find no significant effects for empathy.

Robustness Test

To validate our gratitude measure, we consider actual behavior after the service interaction. A research assistant personally messaged 500 users whose reactions to the firm response appeared on Facebook. In a short message, we asked if they had used the airline again after the complaint. For the 127 responses we received, we performed a chi-square test that confirmed that our measure of gratitude relates to the actual behavior ($\chi^2 = 29.92$, $p < .01$): 82% of those who expressed gratitude used the airline again, but 66% of those who indicated no gratitude did not.

Study 3: Experimental Replications

Even with endogeneity corrections, nonexperimental research is not well suited to make causal inferences. We thus conducted two complementary experimental studies to test for the effects of providing active listening (Study 3a) or empathy (Study 3b) on gratitude for both high- and low-arousal complaints. We report the stimuli in Web Appendix G.

Study 3a: Active Listening in a Controlled Setting

For this preregistered study (https://aspredicted.org/kz4cf.pdf), participants recruited through MTurk using CloudResearch features (N = 900) completed a 2 (customer arousal: high, low) × 3 (employee response: control, medium active listening, high active listening) between-subjects design. After preregistered exclusions, we obtained a final sample of 850 U.S. participants ($M_{age} = 41.30$ years, 50.5% women), who read about a negative airline experience. The high-arousal scenario was like Study 1b, and the low-arousal scenario included less intense versions of the same issues (e.g., delay of 15 minutes instead of 2.5 hours, dirty vs. missing suitcase). These two scenarios are identical in length and significantly differ in evoked arousal, according to a pretest (N = 136; $M_{high} = 6.26$ vs. $M_{low} = 4.71$; $F(1, 96) = 35.94$, $p < .001$).

Participants imagined writing either a high- or low-arousal complaint on the airline's Facebook page (i.e., either 5% negative, high-arousal words or 0% negative, high-arousal words, in line with the slopes in the field studies). Then, participants read one of three employee responses: control response (active listening = .71, empathy = 1.74), medium-active-listening response (active listening = .79, empathy = 1.74), or high-active-listening response (active listening = .85, empathy = 1.74).[12] None of these responses recovered the failure; they all asked the customer to send a private message. Participants then indicated their gratitude, considering the scenario and subsequent response they received (1 = "very ungrateful," 7 = "very grateful"). We also included some exploratory items, attention and manipulation checks, possible covariates/controls, and demographic items, as in Study 1b (see Web Appendix G for all additional analyses with these items).

A 2 × 3 ANOVA for gratitude produced a significant main effect of arousal ($M_{high} = 2.95$, $M_{low} = 4.13$; $F(1, 848) = 127.53$, $p < .001$, in support of H_1), a significant main effect of active listening ($M_{high} = 3.80$, $M_{medium} = 3.63$, $M_{low} = 3.18$; $F(2, 848) = 21.32$, $p < .001$), and a significant interaction effect ($F(2, 848) = 2.35$, $p = .096$). Planned contrasts for the high-arousal complaint revealed that participants indicated higher gratitude after the firm response for high active listening ($M_{high} = 3.25$) compared with medium active listening ($M_{medium} = 2.89$; $F(1, 843) = 3.94$, $p = .048$) and the control condition ($M_{control} = 2.74$; $F(1, 843) = 8.34$, $p = .004$). The medium-active-listening and control conditions did not significantly differ for the high-arousal complaint ($F(1, 843) = .74$, $p = .390$). These findings support H_2.

Planned contrasts for the low-arousal complaint revealed that participants indicated higher gratitude for high active listening ($M_{high} = 4.27$; $F(1, 843) = 9.63$, $p = .002$) and medium active listening ($M_{medium} = 4.36$; $F(1, 843) = 12.42$, $p < .001$) compared with the control condition ($M_{control} = 3.72$). High and medium active listening did not significantly differ for the low-arousal complaint ($F(1, 843) = .63$, $p = .594$). Given the significant interaction effect, we further explore the low-arousal pattern with the quadratic trend, as suggested by Rosenthal and Rosnow (1985). The trend analysis supports an inverted U-shaped effect ($F(1, 423) = 5.21$, $p = .023$).

Study 3b: Empathy in a Controlled Setting

We recruited U.S. participants through MTurk using CloudResearch features (N = 899) for this preregistered study (https://aspredicted.org/ap7ej.pdf) and asked them to complete a 2 (customer arousal: high, low) × 3 (employee response: control, medium empathy, high empathy) between-subjects design. Participants who did not pass the preregistered manipulation or attention checks were excluded, leaving a final sample of 851 respondents ($M_{age} = 42.04$ years, 57.2% women). This experiment was identical to Study 3a, except that the employee responses feature a control (active listening = .71, empathy = 1.74), medium-empathy (active listening = .69, empathy = 5.22), or high-empathy (active listening = .67, empathy = 6.96) version.[13] Again, none of the responses recovered the service failure (see additional analyses in Web Appendix G).

[12] In pretests, the responses (N = 136) significantly differ in perceptions regarding active listening (but not on empathy; $p = .39$). The medium-active-listening condition (M = 4.13, SD = 1.87) is perceived as significantly higher on active listening than the control condition (M = 2.56, SD = 1.67; $p < .001$), and high active listening (M = 5.98, SD = 1.54) is perceived as significantly higher than both medium active listening ($p < .01$) and control (both ps < .001).

[13] The pretest shows that the responses (N = 166) differ significantly in perceptions regarding empathy but not active listening ($p = .76$). Specifically, the medium-empathy condition is significantly higher on empathy (M = 5.22, SD = 1.50) than the control (M = 2.82, SD = 1.67; $p < .001$), and high empathy (M = 6.16, SD = 1.05) is significantly higher than both medium empathy ($p = .003$) and control ($p < .001$).

A 2 × 3 ANOVA for gratitude produced a significant main effect of arousal ($M_{high} = 3.49$, $M_{low} = 4.56$; $F(1, 850) = 113.00$, $p < .001$, in support of H_1) and a significant main effect of empathy ($M_{high} = 4.50$, $M_{medium} = 4.19$, $M_{low} = 3.32$; $F(2, 850) = 48.27$, $p < .001$) but no significant interaction effect ($F(2, 850) = .64$, $p = .526$). Planned contrasts for the high-arousal complaint revealed that participants indicated higher gratitude after the firm response for high empathy ($M_{high} = 4.04$) compared with medium empathy ($M_{medium} = 3.61$; $F(1, 845) = 6.22$, $p = .013$) and the control condition ($M_{control} = 2.76$; $F(2, 845) = 53.30$, $p < .001$) as well as for medium empathy compared with the control condition ($F(2, 845) = 23.70$, $p < .001$). These findings support H_3.

Planned contrasts for the low-arousal complaint revealed that participants indicated higher gratitude for high empathy ($M_{high} = 4.95$; $F(2, 845) = 37.38$, $p < .001$) and medium empathy ($M_{medium} = 4.79$; $F(2, 845) = 26.72$, $p < .001$) compared with the control condition ($M_{control} = 3.88$). High and medium empathy did not significantly differ for the low-arousal complaint ($F(1, 845) = .93$, $p = .335$). Taken together, across two controlled experiments, involving active listening (Study 3a) and empathy (Study 3b), we find causal evidence in line with our field data.

Study 4: Generalization to Product Failures

In our data thus far, we focus on service failures. Considering similarities between service and product failures (Khamitov, Grégoire, and Suri 2020), we examine whether active listening and empathy exert similar effects in a product-focused context in Study 4. We examine the customer complaints of a U.K. retailer on its Twitter profile. Using a third-party data analytics tool, we scraped this Twitter account to gather all complaints over the course of one month. As in Study 1a and Study 2, we only considered complaints that prompt both a firm and a subsequent customer response. Accordingly, we identified 564 customer-initiated complaint interactions. The measures are as in Studies 1a and 2 (see Table 3 and Web Appendix C).

Table 8 displays the results for Study 4. Negative high arousal has a negative linear effect in the main effects model (OR = .864, $p < .01$), in support of H_1. We find a significant negative effect of active listening squared (OR = .999, $p < .05$), a positive effect of empathy (OR = 1.079, $p < .05$), and positive interaction effects of both active listening (OR = 1.009, $p < .05$) and empathy (OR = 1.044, $p < .05$) with negative high arousal.

We display marginal effects in Figure 2 and Web Appendix E. For relatively high-arousal customers, the linear positive effects of active listening (OR = 1.137, $p < .05$) and empathy (OR = 1.896, $p < .01$) support H_2 and H_3. Assuming a 1% increase in active listening (empathy), we observe a 14% (87%) increase in the probability of gratitude. Thus, empathy is much more effective than active listening. For low arousal, we find an inverted U-shaped effect of active listening with a maximum at .65 (65% of firm responses are below, 3% match, and 32% exceed this critical point) and a linear positive effect of empathy (OR = 1.173, $p < .05$).

General Discussion

Across three field studies and three complementary, preregistered experiments, we find detrimental effects of customer high arousal that reduces complainants' gratitude after a firm response. The converging evidence comes from data that represent different platforms (Twitter, Facebook) and markets (Germany, United Kingdom, mixed countries) as well as experimental studies. Active listening and empathy by the firm, independently of each other, de-escalate high arousal in customers and increase gratitude, even without any actual recovery. While active listening is related to the style of the response, empathy is related to the content of the response. Study 1b sheds light on de-escalating effects of active listening and empathy compared with a control response. While the focus of this research is on de-escalating high arousal in customers, we also explore the impact of de-escalation for customers in low-arousal states. For low-arousal customers, we find diminishing effects for the use of active listening, such that its initial positive effect disappears as active listening increases. The effects of empathy on gratitude for low-arousal customers varies across studies. We discuss why intercultural aspects could inform these diverging results in the "Limitations and Further Research Directions" section.

Contributions to the Literature

Our findings provide important contributions to the literature on text-based complaint handling. First, we propose a novel perspective on de-escalating high-arousal emotions in text-based complaining by providing responses that signal active listening and empathy. These insights can help resolve the divergent perspectives regarding whether asynchronous text-based communication, which is characterized by higher arousal levels than other communication channels (Williams 2019), harms or helps firms' recovery efforts (Berger and Iyengar 2013; Moffett, Folse, and Palmatier 2021). When considering high- versus low-arousal negative emotions expressed in complaints, we find a lower likelihood of gratitude after the recovery, calling into question common managerial approaches that only measure valence with sentiment. Drawing from crisis negotiation literature, we examine two options for firms to de-escalate negative high arousal and reduce its negative consequences: active listening and empathy.

To the best of our knowledge, this study is the first to consider this important link, and we posit that de-escalation may be a missing link that can help explain why so many recovery attempts are ineffective. A recent meta-analysis of face-to-face complaining proposes that "negative emotions can be attenuated only through monetary compensation, such as discounts, cashback, and refunds" (Valentini, Orsingher, and Polyakova 2020, p. 212). However, our results across different methodologies and study contexts suggest that after a high-arousal complaint both active listening and empathy by the firm can reduce customer arousal and increase gratitude, at least partially replacing those more cost-intensive approaches. Notably, even when the

Table 8. Predicting Gratitude in Study 4.

	Model 5: Main Effects			Model 6: Full Model		
	OR	SE	95% CI	OR	SE	95% CI
Controls						
Negative low arousal	1.012	.041	.936, 1.095	1.019	.042	.939, 1.105
Positive high arousal	1.159**	.071	1.027, 1.308	1.200**	.089	1.038, 1.387
Positive low arousal	1.184***	.055	1.082, 1.296	1.187***	.054	1.086, 1.296
Length of complaint	1.017	.013	.993, 1.043	1.016	.013	.992, 1.042
User reactions	.996	.161	.726, 1.367	1.007	.163	.733, 1.383
Firm response time	1.000	.000	.999, 1.001	1.000	.000	.999, 1.001
Length of firm response	1.005*	.003	.999, 1.010	1.005*	.003	.999, 1.011
Personal pronoun: "I"	.967	.034	.902, 1.037	.969	.035	.903, 1.041
Personal pronoun: "We"	.964	.022	.922, 1.009	.964	.023	.920, 1.010
Channel change	.756	.219	.429, 1.333	.755	.221	.425, 1.340
Compensation	1.217	.264	.795, 1.863	1.172	.265	.752, 1.826
Information provision	1.015**	.007	1.001, 1.029	1.015**	.007	1.000, 1.029
Apology	.852	.268	.460, 1.577	.851	.277	.450, 1.609
Customer Complaint						
Negative high arousal (NHA)	.864***	.033	.802, .931	.908	.078	.767, 1.075
NHA squared				.976	.015	.946, 1.007
Firm Response						
Active listening	1.010	.008	.995, 1.026	1.012	.008	.996, 1.029
Active listening squared				.999*	.000	.999, 1.000
Empathy	1.069**	.037	.999, 1.145	1.079**	.041	1.002, 1.161
Empathy squared				1.003	.010	.983, 1.022
Interaction Effects						
NHA × Active listening				1.009**	.005	1.001, 1.018
NHA × Empathy				1.044**	.018	1.008, 1.080
Active listening × Empathy				.996	.003	.991, 1.002
Log likelihood		−356.216			−349.029	
Number of observations		564			564	

*$p < .10$.
**$p < .05$.
***$p < .01$.
Significance is based on two-tailed tests.
Notes: CI = confidence interval. We use robust standard errors in our estimation to account for clustered observations. The odds ratios refer to a 1% change in negative high arousal, active listening, and empathy, to ease interpretation.

actual recovery was not evident on social media (because customers were asked to converse in a private channel), the provision of active listening and empathy by the firm showed clear evidence of customers feeling increased gratitude. Moreover, our findings indicate that empathy is more effective than active listening when dealing with high-arousal customers. These findings are in line with recent discussions of the importance of empathy-based marketing (Pedersen 2021) as well as with studies that examine how to de-escalate hate speech in social media (Hangartner et al. 2021). Taken together, by assessing the implications of active listening and empathy in firm responses, we extend current debates about how firms should handle public complaints on social media.

Second, while it may seem to be intuitive that firms should use active listening or empathy to de-escalate customers' complaints in face-to-face settings (Flechas 2020), we are not aware of any empirical efforts to quantify their potential benefits in asynchronous settings using social media data. In our field studies, a 1% increase in negative, high-arousal words lowers the probability that customers express gratitude for the text-based service recoveries up to 19%, mirroring similar evidence from face-to-face service recoveries (Gelbrich 2010). We also quantify how active listening and empathy separately de-escalate high-arousal complainants' and increase gratitude, even before they receive any actual recovery. In our field data, increasing active listening by 1% increases the probability of gratitude by up to 14%, and increasing empathy by 1% increases this probability of gratitude even more, by up to 90%. Thus, empathy is the stronger lever to enhance recovery outcomes.

Third, we provide concrete recommendations for how to measure and implement active listening and empathy in social media channels. The valence-arousal dictionaries enable the measurement of distinct emotions in real time and allow for planning of appropriate response strategies to de-escalate negative high arousal with active listening and/or empathy. In face-to-face and voice-to-voice contexts, nonverbal cues such as facial expressions and tone can indicate how closely the person is listening (Agrawal and Schmidt 2003), but in digital settings, customers must rely only on written cues. Building on linguistic literature, we argue and demonstrate that active listening can be accomplished online through linguistic style matching of function words (Table 4). Because function words are processed rather unconsciously, our operationalization of linguistic style matching avoids the potential for the superficial appearance of coordination due to simple repetition of content words (Niederhoffer and Pennebaker 2002). Note that we are not suggesting that simply using more function words is necessarily better. Rather, we suggest that the function words that complainants use reflect their linguistic style, and this, in turn, needs to be matched by the service employee. This may involve more of certain function word categories and less of others. Empathy in social media conversations can be demonstrated by identifying the emotions that the customers are feeling and responding with verbal cues of genuine concern (Alam, Danieli, and Riccardi 2018). Moreover, gratitude can be measured with words that express a feeling of appreciation. We have developed and validated new dictionaries provided in Table 5 that are ready for use by researchers and firms.

Managerial Implications

Customers frequently turn to social media to voice complaints when they experience a service or product failure. The 2020 Consumer Rage Study (Customer Care Measurement & Consulting 2020) highlighted that these complaints often feature negative, high-arousal emotions. Consistent with this observation, almost two-third of the complaints in our field studies used negative, high-arousal emotion words. We offer several recommendations for firms to address such complaints.

Identifying expressed arousal in text-based complaints. Firms must consider not just valence but also arousal in text-based complaining. For each additional percentage point increase in negative, high-arousal words or phrases (e.g., "I am outraged!"), the probability of gratitude with a firm response decreases by up to 19%. The dictionaries we provide enable firms to measure distinct emotions in real time, then provide appropriate response strategies to de-escalate if needed. Once arousal has been identified in text-based complaints, firms should adopt at least one of the two de-escalation strategies that we propose and test: active listening and empathy.

Responding with active listening in text-based complaints. Firm responses should signal active listening by adapting their style to the style of the complaining customer. However, our data suggest that on average, the firm responses in the field studies reflect only a medium level of active listening (Study 1a = .60, Study 2 = .55, Study 4 = .59, on a measurement scale ranging from 0 to 1). While the simple repetition of content words may evoke a superficial appearance, using linguistic style matching of function words has the advantage that these are processed rather unconsciously by the receiver. Given that matching function words is a rather complex endeavor, we recommend using an algorithm based on our measure to capture the nine function word categories in a complaint and provide automatic suggestions on how to improve the responses style (e.g., an approach like Grammarly [www.grammarly.com]). Any automatic response algorithms should be crafted to maximize linguistic style matching for high-arousal customers. In situations where support by algorithms is not possible, service employees could follow the heuristics of formal versus informal, complex versus simple, and concrete versus abstract language summarized in Table 4. Here, firms should train employees on detecting the style of the complainant and on responding accordingly (e.g., a rather informal complaint should receive a response with many personal pronouns, whereas a rather formal complaint should receive a response with many impersonal pronouns).

Responding with empathy in text-based complaints. In addition to adapting the style of their response, firm responses should also adapt the content by including empathy words from our dictionary. For example, phrases like "I can imagine how you feel" or "you have my sympathy" can reduce customer arousal and increase gratitude. Any automatic response to a high-arousal complaint should be crafted to include empathy words from our dictionary, and service employees should be sensitized to use such words. So far, the importance of empathy when responding is not reflected in our data. On average, the firm responses use only a moderate amount of empathy words (Study 1a = 4% of total words, Study 2 = 6% of total words, Study 4 = 5% of total words). Both active listening and empathy reduce arousal and increase gratitude independently, even when there has not been a service failure recovery. However, our findings indicate that empathy is more effective than active listening for dealing with high-arousal customers. Providing 1% more empathy increases gratitude by 90% relative to the 14% increase seen for providing 1% more active listening. Therefore, empathy is the stronger lever to enhance the gratitude associated with the recovery efforts, and firms should prioritize the inclusion of empathy words (e.g., "you are right," "our mistake") in their responses to high-arousal complaints.

Limitations and Further Research Directions

Although we obtain consistent results across multiple studies and broad support for the proposed effects of active listening and empathy, we acknowledge some limitations that we hope stimulate further research. Complaining customers who reach out to firms always experience a certain level of negative arousal (Gelbrich 2010), but on social media, customers who engage in broadcasting (Barasch and Berger 2014) may phrase their complaints to avoid making themselves look bad, which might trigger more neutral arousal states. In this sense, we study a bounded range of arousal, and we examine relatively lower and higher arousal in our field studies, such that our empirical tests are rather conservative.

While we complement the evidence from the field with randomized experiments, we nevertheless need to acknowledge potential sources of endogeneity in the field data. We were able to measure expressed arousal but not failure severity, and such omitted variables may have affected the field results. In addition, a firm's response may change based on the expressed arousal of the customer, and we are not able to account for such strategic behavior of firms, Finally, the decision to provide low versus high active listening and empathy may vary by service employee, posing a source of unobserved heterogeneity. We urge future research to keep these challenges in mind when exploring complaint handling in text-based channels.

Although the focus of our research was on de-escalating high-arousal complainants, we did also explore the impact of varying levels of active listening and empathy for low-arousal complainants. We find diminishing returns for the use of active listening in response to low-arousal complaints. The initial positive effect disappears, with turning points in the field studies that range between values of .58 and .66 for active listening. The effects of empathy on gratitude for low-arousal customers vary across studies, including nonsignificant effects (Study 1a: German sample, Study 2: international sample) and linear positive effects (Study 3b: U.S. sample, Study 4: U.K. sample). We posit that these varying effects may occur due to cultural factors and expectations at both the firm and customer levels, as culture has been shown to influence how customers respond to firms' expressions of empathy. Cultures that embrace expressed politeness as a norm might prompt customers to display greater gratitude for expressions of empathy. Thus, additional research should investigate which levels of empathy are optimal in different cultures.

Whereas previous research often focused on solving the problem and our research focuses on the de-escalation, a combination of both is a very intriguing direction for future research. There could be conditions where the order matters, and sometimes it might be warranted to first solve the problem and then de-escalate the negative, high-arousal emotions. It has also become more common for the actual service recoveries to happen with direct messages, and the emotions expressed in private versus public interactions could vary widely. Thus, a major limitation in the literature is the lack of research on complaint handling and service recoveries in private text-based service chats where an one-on-one interaction between employees and customers takes place that offer more possibilities for de-escalation and relationship building.

Finally, social media interactions in general are often driven by high-arousal, negative emotions. Thus, it would be interesting to examine how well our text-based operationalizations of active listening and empathy can de-escalate arousal in discussions among consumers, thereby making interaction partners more receptive to each other's perspective and improving social media conversations.

Acknowledgments

The authors are grateful to Jochen Wulf for support in the data collection for Study 2 and to Stephan Ludwig as well as participants of research seminars at Vrije Universiteit Amsterdam, IESEG School of Management, and University of Coimbra for helpful feedback on the manuscript. The data collection for Studies 1a and 2 was supervised by the first author; the data collection for Studies 1b, 3a, and 3b was supervised by the second author; and the data collection for Study 4 was supervised by the fifth author. The first author analyzed the data for Studies 1a, 2, and 4, and the first and second author independently analyzed the data for Studies 1b, 3a, and 3b.

Associate Editor

Robert Meyer

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Dennis Herhausen https://orcid.org/0000-0002-4335-1703

References

Agrawal, M.L. and Marcus Schmidt (2003), "Listening Quality of the Point of Service Personnel as Impulse Trigger in Service Purchase: A Research Framework," *Journal of Services Research*, 3 (1), 29–43.

Alam, Firoj, Morena Danieli, and Giuseppe Riccardi (2018), "Annotating and Modeling Empathy in Spoken Conversations," *Computer Speech & Language*, 50, 40–61.

Allard, Thomas, Lea H. Dunn, and Katherine White (2020), "Negative Reviews, Positive Impact: Consumer Empathetic Responding to Unfair Word of Mouth," *Journal of Marketing*, 84 (4), 86–108.

Avochato (2021), "54 Text Messaging Stats Every Business Should Know," (July 2), https://blog.avochato.com/54-text-messaging-stats-every-business-should-know-avochato.

Barasch, Alixandra and Jonah Berger (2014), "Broadcasting and Narrowcasting: How Audience Size Affects What People Share," *Journal of Marketing Research*, 51 (3), 286–99.

Belanche, Daniel, Luis V. Casaló, Carlos Flavián, and Jeroen Schepers (2020), "Robots or Frontline Employees? Exploring Customers' Attributions of Responsibility and Stability after Service Failure or Success," *Journal of Service Management*, 31 (2), 267–89.

Berger, Jonah and Raghuram Iyengar (2013), "Communication Channels and Word of Mouth: How the Medium Shapes the Message," *Journal of Consumer Research*, 40 (3), 567–79.

Berger, Jonah and Katherine L. Milkman (2012), "What Makes Online Content Viral?" *Journal of Marketing Research*, 49, 192–205.

Bonchek, Mark (2015). "Why Customer Gratitude Trumps Loyalty," *Harvard Business Review* (October 19), https://hbr.org/2015/10/why-customer-gratitude-trumps-loyalty.

Bonifield, Carolyn and Catherine Cole (2007), "Affective Responses to Service Failure: Anger, Regret, and Retaliatory Versus Conciliatory Responses," *Marketing Letters*, 18 (1), 85–99.

Bougie, Roger, Rik Pieters, and Marcel Zeelenberg (2003), "Angry Customers Don't Come Back, They Get Back: The Experience and Implications of Anger and Dissatisfaction in Services," *Journal of the Academy of Marketing Science*, 31, 377–93.

Boyd, Ryan L., Kate G. Blackburn, and James W. Pennebaker (2020), "The Narrative Arc: Revealing Core Narrative Structures Through Text Analysis," *Science Advances*, 6 (32), eaba2196.

Cannava, Kaitlin and Graham D. Bodie (2017), "Language Use and Style Matching in Supportive Conversations Between Strangers and Friends," *Journal of Social and Personal Relationships*, 34 (4), 467–85.

Chin, Hyojin, Lebogang Wame Molefi, and Mun Yong Yi. (2020), "Empathy Is All You Need: How a Conversational Agent Should Respond to Verbal Abuse," in Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, 1–13.

Chung, Cindy and James W. Pennebaker (2007), "The Psychological Functions of Function Words," in *Social Communication*, 1st ed., Klaus Fiedler, ed. Psychological Press, 343–59.

Coker, Brent (2020), "Arousal Enhances Herding Tendencies When Decision Making," *Journal of Consumer Behaviour*, 19 (3), 229–39.

Customer Care Measurement & Consulting (2020), "National Customer Rage Study," (accessed April 12, 2022), https://www.customercaremc.com/insights/national-customer-rage-study/2020-national-customer-rage-study/.

Flechas, Sophie Rodriguez (2020), "Four Tips on How to De-Escalate in Customer Service," OMQ (March 29), https://omq.ai/blog/4-tips-deescalation/.

Gelbrich, Katja (2010), "Anger, Frustration, and Helplessness After Service Failure: Coping Strategies and Effective Informational Support," *Journal of the Academy of Marketing Science*, 38 (5), 567–85.

Golmohammadi, Alireza, Taha Havakhor, Dinesh K. Gauri, and Joseph J. Comprix (2021), "Complaint Publicization in Social Media," *Journal of Marketing*, 85 (6), 1–23.

Gonzales, Amy L., Jeffrey T. Hancock, and James W. Pennebaker (2010), "Language Style Matching as a Predictor of Social Dynamics in Small Groups," *Communication Research*, 37 (1), 3–19.

Grégoire, Yany, Fateme Ghadami, Sandra Laporte, Sylvain Sénécal, and Denis Larocque (2018), "How Can Firms Stop Customer Revenge? The Effects of Direct and Indirect Revenge on Post-Complaint Responses," *Journal of the Academy of Marketing Science*, 46 (6), 1052–71.

Grégoire, Yany and Anna S. Mattila (2020), "Service Failure and Recovery at the Crossroads: Recommendations to Revitalize the Field and its Influence," *Journal of Service Research*, 24 (3), 323–28.

Hangartner, Dominik, Gloria Gennaro, Sary Alasiri, Nicholas Bahrich, Alexandra Bornhoft, Joseph Boucher, et al. (2021), "Empathy-Based Counterspeech Can Reduce Racist Hate Speech in a Social Media Field Experiment," *Proceedings of the National Academy of Sciences*, 118 (50), e2116310118.

Herhausen, Dennis (2020), "How to Keep Complaints from Spreading," *Harvard Business Review*, 99 (3), 19–23.

Herhausen, Dennis, Stephan Ludwig, Dhruv Grewal, Jochen Wulf, and Marcus Schoegel (2019), "Detecting, Preventing, and Mitigating Online Firestorms in Brand Communities," *Journal of Marketing*, 83 (3), 1–21.

Hogreve, Jens and Kathrin Hoerner (2019), "Service Recovery on Stage: Effects of Social Media Recovery on Virtually Present Others," *Journal of Service Research*, 22 (4), 421–39.

Homburg, Christian, Laura Ehm, and Martin Artz (2015), "Measuring and Managing Consumer Sentiment in an Online Community Environment," *Journal of Marketing Research*, 52 (5), 629–41.

Ireland, Molly E. and Marlone D. Henderson (2014), "Language Style Matching, Engagement, and Impasse in Negotiations," *Negotiation and Conflict Management Research*, 7 (1), 1–16.

Johnen, Marius and Oliver Schnittka (2019), "When Pushing Back Is Good: The Effectiveness of Brand Responses to Social Media Complaints," *Journal of the Academy of Marketing Science*, 47 (5), 858–78.

Joireman, Jeff, Yany Grégoire, Berna Devezer, and Thomas M. Tripp (2013), "When Do Customers Offer Firms a 'Second Chance' Following a Double Deviation? The Impact of Inferred Firm Motives on Customer Revenge and Reconciliation," *Journal of Retailing*, 89 (3), 315–37.

Kalamas, Maria, Michel Laroche, and Lucy Makdessian (2008), "Reaching the Boiling Point: Consumers' Negative Affective Reactions to Firm-Attributed Service Failures," *Journal of Business Research*, 61 (8), 813–24.

Khamitov, Mansur, Yany Grégoire, and Anshu Suri (2020), "A Systematic Review of Brand Transgression, Service Failure Recovery and Product-Harm Crisis: Integration and Guiding Insights," *Journal of the Academy of Marketing Science*, 48 (3), 519–42.

Knox, George and Rutger Van Oest (2014), "Customer Complaints and Recovery Effectiveness: A Customer Base Approach," *Journal of Marketing*, 78 (5), 42–57.

Lakin, Jessica L. and Tanya L. Chartrand (2003), "Using Nonconscious Behavioral Mimicry to Create Affiliation and Rapport," *Psychological Science*, 14 (4), 334–39.

Ludwig, Stephan, Dennis Herhausen, Dhruv Grewal, Liliana Bove, Sabine Benoit, Ko de Ruyter, et al. (2022), "Communication in the Gig Economy: Buying and Selling in Online Freelance Marketplaces," *Journal of Marketing*, 86 (4), 141–61.

McMains, Michael J. and Frederick J. Lanceley (2003), "The Use of Crisis Intervention Principles by Police Negotiators," *Journal of Police Crisis Negotiations*, 3 (1), 3–30.

Min, Kyeong Sam, Jae Min Jung, and Kisang Ryu (2021), "Listen to Their Heart: Why Does Active Listening Enhance Customer Satisfaction After a Service Failure?" *International Journal of Hospitality Management*, 96, 102956.

Moffett, Jordan W., Judith Anne Garretson Folse, and Robert W. Palmatier (2021), "A Theory of Multiformat Communication: Mechanisms, Dynamics, and Strategies," *Journal of the Academy of Marketing Science*, 49 (3), 441–61.

Nguyen, Doan T. and Janet R. McColl-Kennedy (2003), "Diffusing Customer Anger in Service Recovery: A Conceptual Framework," *Australasian Marketing Journal*, 11 (2), 46–55.

Niederhoffer, Kate G. and James W. Pennebaker (2002), "Linguistic Style Matching in Social Interaction," *Journal of Language and Social Psychology*, 21 (4), 337–60.

Noesner, Gary W. and Mike Webster (1997), "Crisis Intervention: Using Active Listening Skills in Negotiations," *FBI Law Enforcement Bulletin*, 66, 13.

Packard, Grant and Jonah Berger (2021), "How Concrete Language Shapes Customer Satisfaction," *Journal of Consumer Research*, 47 (5), 787–806.

Packard, Grant, Sarah G. Moore, and Brent McFerran (2018), “(I’m) Happy to Help (You): The Impact of Personal Pronoun Use in Customer–Firm Interactions,” *Journal of Marketing Research*, 55 (4), 541–55.

Palmatier, Robert W., C.B. Jarvis, J.R. Bechkoff, and Frank R. Kardes (2009), “The Role of Customer Gratitude in Relationship Marketing,” *Journal of Marketing*, 73 (5), 1–18.

Pedersen, Carsten L (2021), “Empathy-Based Marketing,” *Psychology & Marketing*, 38 (3), 470–80.

Qiu, Lin, Han Lin, Jonathan Ramsay, and Fang Yang (2012), “You Are What You Tweet: Personality Expression and Perception on Twitter,” *Journal of Research in Personality*, 46 (6), 710–8.

Ramsey, Rosemary P. and Ravipreet S. Sohi (1997), “Listening to Your Customers: The Impact of Perceived Salesperson Listening Behavior on Relationship Outcomes,” *Journal of the Academy of Marketing Science*, 25 (2), 127.

Rosenthal, Robert and Ralph A. Rosnow (1985), *Contrast Analysis: Focused Comparisons in the Analysis of Variance*. Cambridge University Press.

Russell, James A. (1980), “A Circumplex Model of Affect,” *Journal of Personality and Social Psychology*, 39 (6), 1161.

Russell, James A. and L.F. Barret (1999), “Core Affect, Prototypical Emotional Episodes, and Other Things Called Emotion: Dissecting the Elephant,” *Journal of Personality and Social Psychology*, 76 (5), 805–19.

Schaefers, Tobias and Julia Schamari (2016), “Service Recovery via Social Media: The Social Influence Effects of Virtual Presence,” *Journal of Service Research*, 19, 1–17.

Singh, Sunil K., Detelina Marinova, and Jagdip Singh (2020), “Business-to-Business E-Negotiations and Influence Tactics,” *Journal of Marketing*, 84 (2), 47–68.

Smith, Amy K. and Ruth N. Bolton (2002), “The Effect of Customers’ Emotional Responses to Service Failures on Their Recovery Effort Evaluations and Satisfaction Judgments,” *Journal of the Academy of Marketing Science*, 30 (1), 5–23.

Strizhakova, Yuliya, Yelena Tsarenko, and Julie A. Ruth (2012), “I’m Mad and I Can’t Get That Service Failure off My Mind: Coping and Rumination as Mediators of Anger Effects on Customer Intentions,” *Journal of Service Research*, 15 (4), 414–29.

Surachartkumtonkun, Jirapon, Janet R. McColl-Kennedy, and Paul G. Patterson (2015), “Unpacking Customer Rage Elicitation a Dynamic Model,” *Journal of Service Research*, 18 (2), 177–92.

Swaab, Roderick I., William W. Maddux, and Marwan Sinaceur (2011), “Early Words That Work: When and How Virtual Linguistic Mimicry Facilitates Negotiation Outcomes,” *Journal of Experimental Social Psychology*, 47 (3), 616–21.

Tausczik, Yla R. and James W. Pennebaker (2010), “The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods,” *Journal of Language and Social Psychology*, 29 (1), 24–54.

Tax, Stephen S., Stephen W. Brown, and Murali Chandrashekaran (1998), “Customer Evaluations of Service Complaint Experiences: Implications for Relationship Marketing,” *Journal of Marketing*, 62 (2), 60–76.

Umashankar, Nita, Raji Srinivasan, and Jeffrey R. Parker (2016), “Cross-Buying After Product Failure Recovery? Depends on How You Feel About It,” *Journal of Marketing Theory and Practice*, 24 (1), 1–22.

Valentini, Sara, Chiara Orsingher, and Alexandra Polyakova (2020), “Customers’ Emotions in Service Failure and Recovery: A Meta-Analysis,” *Marketing Letters*, 31 (2), 199–216.

Van Hasselt, Vincent B., Stephen J. Romano, and Gregory M. Vecchi (2008), “Role Playing: Applications in Hostage and Crisis Negotiation Skills Training,” *Behavior Modification*, 32 (2), 248–63.

Van Vaerenbergh, Yves, Dorottya Varga, Arne De Keyser, and Chiara Orsingher (2019), “The Service Recovery Journey: Conceptualization, Integration, and Directions for Future Research,” *Journal of Service Research*, 22 (2), 103–19.

Vecchi, Gregory M., Vincent B. Van Hasselt, and Stephen J. Romano (2005), “Crisis (Hostage) Negotiation: Current Strategies and Issues in High-Risk Conflict Resolution,” *Aggression and Violent Behavior*, 10 (5), 533–51.

Villarroel Ordenes, Francisco, Stephan Ludwig, Ko De Ruyter, Dhruv Grewal, and Martin Wetzels (2017), “Unveiling What Is Written in the Stars: Analyzing Explicit, Implicit, and Discourse Patterns of Sentiment in Social Media,” *Journal of Consumer Research*, 43 (6), 875–94.

Warriner, Amy Beth, Victor Kuperman, and Marc Brysbaert (2013), “Norms of Valence, Arousal, and Dominance for 13,915 English Lemmas,” *Behavior Research Methods*, 45 (4), 1191–1207.

Whissell, Cynthia (2009), “Using the Revised Dictionary of Affect in Language to Quantify the Emotional Undertones of Samples of Natural Language,” *Psychological Reports*, 105 (2), 509–21.

Williams, Leah K. (2019), “How to De-Escalate an Angry Customer over Social,” BSquared (accessed April 12, 2022), https://bsquared.media/angry-customer-social/.

Xia, Lan and Monika Kukar-Kinney (2013), “Examining the Penalty Resolution Process: Building Loyalty Through Gratitude and Fairness,” *Journal of Service Research*, 16 (4), 518–32

Yin, Dezhi, Samuel D. Bond, and Han Zhang (2017), “Keep Your Cool or Let It Out: Nonlinear Effects of Expressed Arousal on Perceptions of Consumer Reviews,” *Journal of Marketing Research*, 54 (3), 447–63.