Using Speech Acts to Elicit Positive Emotions for Complainants on Social Media

Young Anna Argyris [a,*] & Kafui Monu [b] & Yongsuk Kim [c] & Yilu Zhou [d] & Zuhui Wang [e] & Zhaozheng Yin [f]

[a] *Department of Media and Information, College of Communication Arts and Sciences, Michigan State University, 404 Wilson Road, East Lansing, MI 48824, USA*
[b] *Management Information Systems Area, School of Business, University of Northern British Columbia, Prince George, BC, Canada*
[c] *Management Information Systems Area, School of Business, Sungkyunkwan University, Seoul, South Korea*
[d] *Information Systems Area, Gabelli School of Business, Fordham University, New York, NY, USA*
[e] *Department of Computer Science, The State University of New York at Stony Brook, Stony Brook, NY, USA*
[f] *Artificial Intelligence Institute, Department of Biomedical Informatics, Department of Computer Science, Department of Applied Mathematics & Statistics (Affiliated), The State University of New York at Stony Brook, Stony Brook, NY, USA*

Abstract

A carefully tailored tone in response to a complaint on social media can create positive emotions for an upset customer. However, very few studies have identified what response tones, based on an established theory, would be most effective for complaint management. This study conceptualizes a service agent's response tones based on Ballmer and Brennenstuhl's (1981) classification of speech acts and examines how an agent's use of speech acts elicit positive emotions for the complainant. Ballmer and Brennenstuhl classify speech acts within the dimensions of conventionality and dialogicality, and they suggest the two dimensions interact. Thus, we examine the impact of each dimension of speech acts and the interactions between the two dimensions on the elicitation of positive emotions for complainants. We collected over 100,000 tweets and classified firm agents' speech acts and complainants' emotions by designing deep learning architectures (i.e., bi-directional recurrent neural networks). Our fixed-effect regression results show that a low level of each speech act leads to the elicitation of customers' positive emotions but that the combination of the two erodes the individual advantages. This study expands Ballmer and Brennenstuhl's (1981) speech act classification from a speaker's perspectives to a listener's perspectives by contextualizing it in an analysis of service agents' tones and their roles in eliciting positive emotions among complainants.

Keywords: Complaint management; Social media; Deep learning algorithm; Speech act theory

Introduction

An increasing number of customers leverage social media to complain about products and services they received from companies (Gunarathne, Rui, & Seidmann, 2017, 2018; Honey & Herring, 2009). These complaints, amplified by the ubiquity and reach of social media, can propagate negative word of mouth (WOM) at an unprecedented scale and lead to damaged company reputations and financial losses (Hansen, Kupfer, & Hennig-Thurau, 2018; He, Rui, & Whinston, 2017; Herhausen, Ludwig, Grewal, Wulf, & Schoegel, 2019). Accordingly, many firms have built customer service teams dedicated to handling complaints through social media to prevent an uncontrollable outbreak of hostile publicity (Herhausen et al., 2019). Of the various social media platforms, twitter is the choice of many companies and customers because of its instant posting capabilities (Honey & Herring, 2009) and, thus, was chosen as the main focus in this study.

Abbreviation: B&B, Ballmer and Brennenstuhl; SA, speech acts

* Corresponding author.
E-mail address: argyris@msu.edu (Y.A. Argyris).

Recent studies have started attending to complaint handling on social media (e.g., Hu & Liu, 2004; Ma et al., 2015; Gunarathne et al., 2017). Yet, much remains unknown about the response strategies, particularly response tones, that firms can employ to elicit customers' positive emotions in conflict negotiation contexts. Although the importance of a service agent's response tone, apart from the linguistic meanings of the response itself, has been noted for decades in the service recovery literature (Beebe & Waring, 2002; Davidow, 2000; Davidow, 2003; Morris, 1988), few studies have identified response tones that are effective for addressing complaints.

The primary goal of this study is to address this gap in the literature by identifying a response tone that service agents can use on social media to elicit positive emotions among complainants to prevent an outburst of negative word of mouth (WOM). In this study, we define a response tone as the semantic style of a service agent's response that conveys their intentions and attitudes toward the customer, apart from the content of the response (Beebe & Waring, 2002; Beebe & Waring, 2004). Response tones matter for two reasons. First, an appropriately tailored tone can be more effective for eliciting positive emotions among complainants than the content of a response. This is because for some complainants, being treated with dignity is more important than receiving redress and compensation (Davidow, 2003; Morris, 1988). Second, training employees to use an optimal response tone reduces the financial burden of resolving complaints, while also reducing the stress levels of service agents (Davidow, 2000).

In identifying effective response tones, we draw from Ballmer & Brennenstuhl (1981) speech act classification. Speech act theory was devised in philosophy in the 1960s and has been used in a wide array of disciplines, including linguistics, psychology, legal studies, public policies, and recently, artificial intelligence (AI). Speech acts refer to the semantics of a conversation devised to deliver the speaker's intentions with the aim of influencing the listener's interpretation of the speech (Searle, 1999). Speech acts are not merely uttered but carefully crafted to adjust the listener's interpretations of the linguistic and direct meaning of sentences: hence, the term speech “acts” (Austin, 1962). Speech acts concern not “what is said, but how it is said” (Green, 2017). Among a few available classifications of speech acts, B&B's classification was chosen for its unique focus on speech acts used in bi-directional negotiations (Chang & Woo, 1994). Due to their focus, B&B's classification is increasingly adopted for the development of AI-based client-facing systems designed to address customer requests (Goldstein & Sabin, 2006; Jiang & Huhns, 2005; Steiner, 2019; Vieira, Moreira, Wooldridge, & Bordini, 2007). Given B&B speech acts' close fit with our definition of a response tone and its' broad applicability, we conceptualize the response tones based on B&B's speech acts in this study.

Another advantage of B&B's classification is that it considers the multidimensions of speech acts and the interactions between them, thereby allowing for in-depth analyses of response tones. In particular, B&B's classification identifies speech acts along the dimensions of conventionality and dialogicality, and it suggests interactions between the two. Conventionality refers to the extent to which speech is prescripted or personalized on the spot for the listener. A high level of conventionality is called a “socially entrenched speech act,” while a low level is called a “private speech act.” Dialogicality denotes the extent to which either a speaker invites a listener's response or concludes the conversation. A high level of dialogicality is called “dialogue,” and a low level is called “monologue.” We investigate which level of each speech act and what combination(s) of the two speech acts would be effective for creating positive emotions for complainants.

To study this phenomenon, we identified all the companies that had twitter customer care accounts on the American Customer Satisfaction Index (ACSI), which is the national cross-industry measure of customer satisfaction in the US. Of these 34 accounts, we retrieved 102,407 tweets, from which we extracted 34,709 interactions made up of 86,744 tweets. An interaction is defined herein as a pair of tweets exchanged between an agent and a customer involved in the resolution of the customer's issue. According to this definition, an interaction consists of two tweets—(i) the agent's tweet to address the customer's complaint, and (ii) the customer's response to the agent's tweet. We analyzed how the agent's choice of speech acts in (i) elicits positive emotions for the customer in (ii). To identify speech acts in agents' tweets, we developed two recurrent neural networks (RNNs) (Lipton, Berkowitz, & Elkan, 2015) for conventionality and dialogicality. To recognize the kind of emotion (positive, neutral, and negative) in complainants' tweets, another multi-class RNN was proposed. Finally, employing a regression estimation with dyadic interaction-level fixed effects, we tested the effects of the two dimensions of speech acts and their interactions on creating positive emotions for complainants.

Literature Review

Existing Literature on Social Media Complaint Management

Several prior studies have attended to complaint management on social media, attesting to its rapid growth in practice, although much remains unknown about corporate complaint handling strategies. The most common topic in these studies is the strategies that complainants employ to capture an agent's attention to solve their problems. For example, customers who use polite tones in their complaints receive faster responses from firms, and customers' social media standing (e.g., the number of followers) helps augment the effects of their politeness (Hu, Tafti, & Gal, 2019). Similarly, complainants in the presence of other social media users receive faster responses (Gunarathne et al., 2018).

The second topic deals with factors that facilitate complaining behaviors on social media. Past successes in obtaining redress from a firm increases the chances of the customer complaining again (Ma, Sun, and Kekre, 2015). Repeat complainants with concerns about the complaint handling processes are less likely to feel better even after expressing their concerns (Gunarathne et al., 2017).

The third topic in past studies deals with response strategies—in terms of content and tones—that firms can employ to handle complainants' concerns, and this strand of research has received less attention than the prior two. Among a variety of response contents including handoff, apology, explanation, gratitude, direct messaging (DM), gratitude is found to create positive emotions for complainants, while proposing a handoff worsens them (Gunarathne et al., 2017). Another study suggests that a firm's defensive strategy, as opposed to accommodative strategy, in which a firm denies responsibility for an issue grows social media observers' purchase intentions for utilitarian products (Johnen & Schnittka, 2019). A defensive strategy allows observers to refute negative claims about the firm to which they are loyal, thereby alleviating their cognitive dissonance. Notably, the kind of response tone (i.e., formal or informal) changes or reverses observers' reactions to different response strategies. When the firm uses defensive responses, the response tone must be formal to show the firm's seriousness. We credit Johnen and Schnittka's (2019) for separating response tones from response content and demonstrating that response tones significantly affect observers' reactions to the firm's handling of complaints beyond the extent that response content explains. Nonetheless, Johnen and Schnittka (2019) did not include the multidimensionality of tones or interactions between them, nor did they show which response tone creates positive emotions among the customer who raised the issue.

In this study, we aim to fill this gap by identifying a response tone that an agent can use to create positive emotions for complainants without resorting to the excessive or premature provisions of redress. We ground our investigation in a firmly established theory to provide a systematic suggestion for a tone and to increase its generalizability for the growing areas of research and practice (e.g., AI) relevant to complaint management.

Theoretical Framework and Hypotheses Development

Conceptualization of Response Tones Based on Speech Act Theory

For identifying a proper response tone that a firm can employ on social media, we can employ the theory of speech acts (Austin, 1962; Searle & Vanderveken, 1985). Speech acts concern the semantics of a conversation, not the linguistic, direct meaning of a conversation (Searle, 1999). Recognition of the significance of speech acts has underscored "the ability of language to do other things than deliver the content" (Green, 2017). Apart from delivering the content, a speaker can do a variety of things with speech acts, such as making requests, asking questions, giving orders, and making promises (Searle & Vanderveken, 1985). In this sense, speech acts are used interchangeably with illocutionary force (Austin, 1962; Austin, 1970). The utterance of a meaningful sentence such as "You'll be more punctual in the future" may mean that "the speaker is making a prediction or issuing a command or even a threat" (Green, 2017). In this example, the speaker uses the illocutionary force to influence the listener's interpretation of the meaning of those words. Since the 1960s, speech act theory has become influential in various fields, including philosophy, linguistics, psychology, legal theory, and lately, AI in the design of speech for automated agents (Green, 2017).

Ballmer and Brennenstuhl's Classification of Speech Acts

Among several classifications of speech acts, B&B's speech acts are relevant to analyzing negotiations between two parties (Chang & Woo, 1994), and, hence, are relevant to analyzing an agent's response to a complainant. Other classifications exist, such as Searle's (1999), yet they are suitable for analyzing unidirectional speech (Chang & Woo, 1994). In addition, the B&B classification was built according to a bottom-up model and, thus, fits with the participatory communicative patterns of social media (Chang & Woo, 1994).

Given this unique focus on dyadic negotiations, B&B's classification has been applied to developing intelligent and interactive customer-facing systems (Goldstein & Sabin, 2006; Jiang & Huhns, 2005; Steiner, 2019; Vieira et al., 2007). For instance, Woo and Chang (1992) used B&B's classification to develop negotiations between intelligent agent-based systems. They used B&B's classification to "map" a conversation and identify what the automated agent speaker should say in response to the other party in a negotiation. For instance, an agent must decide proper speech acts for an incoming claim (e.g., "I should get a 100% refund") in order to exercise bargaining power over the complainant. The capability of B&B's classification to devise counterarguments in a negotiation is relevant to the field of interactive marketing, especially the subfields concerning AI-assisted customer engagement. Examples include a chatbot created to interact with consumers (Gnewuch, Morana, & Maedche, 2017) and self-service technologies devised to negotiate sales terms with a firm (Su et al., 2001).

Due to these advantage and relevance of B&B's classification, we chose B&B's speech act classification over other classifications in our pursuit of a proper response tone.

Expansion of B&B's Speech Act Classification

Although B&B devised the classification with consideration of a hearer in a negotiation, their focus was on the speaker who tries to exert influence on the hearer, without considering the impact of the speech acts on the hearer's emotions. With an example, "A 100% refund is not possible," B&B's focus would be to increase the speaker's bargaining power by influencing the hearer's interpretation of the speaker's speech act. Many questions can arise surrounding the hearer's emotional reactions to the speech acts, such as, would the hearer feel intimidated? B&B have not yet explained how a choice of speech acts affects the listener's emotions.

Complainants' emotions are critical for the success of twitter complaint management because complainants' positive emotions increase their loyalty (DeWitt, Nguyen, & Marshall, 2008), and negative emotions result in the discontinuation of their patronage (Chebat & Slusarczyk, 2005) as well as in the viral spread of negative WOM (Hansen et al., 2018; Herhausen et al., 2019). Thus, how a complainant emotionally responds to a speech act reflects the success of complaint management on social media and is worthy of further attention. Thus, we expand B&B's classification to include the hearer's emotions.

Multidimensionality of B&B's Speech Acts and Interactions Among Them

According to B&B, two dimensions of speech acts exist that concern the semantics of a sentence: conventionality and dialogicality. Each of these has two levels. Conventionality refers to the intention of the speaker to either follow a predefined script or to engage in improvised speech. A high-level conventional speech act—i.e., a socially entrenched speech act—follows a pre-established script independent of the speaker or the listener. A low-level conventional speech act—i.e., a private speech act—is customized for the speaker and the listener. A socially entrenched speech act is institutionalized and controlled; a private speech act is personalized and original. Socially entrenched speech acts apply to all members of society, while private speech acts are customized for the listener. These speech acts can be contextualized in complaint management settings as follows. An agent can say "Please do not hesitate to contact us if you have further questions" in a socially entrenched speech act, or they can say "Please let me know if your vehicle has another problem with the transmission" in a private speech act. The former is socially entrenched because the sentence can be applied to anyone, while the latter is private because it applies specifically to the complainant. In this sense, a private speech act is customized for the listener and, thus, applies to this complainant only.

Dialogicality identifies the speaker's intention to invite a response from the listener. At the high level of dialogicality, called dialogue, two voices engage with each other from different perspectives; at the low level of dialogicality, called monologue, only one person speaks. While dialogue invites a multiplicity of speakers and a variety of perspectives, monologue is conclusive. Dialogue is associated with debates and conversations; monologue is associated with statements. Examples of these speech acts are "Let me know what you think of our offer" in dialogue versus "This is what we can offer for you" in a monologue.

In addition, these two dimensions interact with each other, creating the four speech acts (Fig. 1). Each speech act conveys the speakers' distinct attitudes and intentions toward the hearer. The combination of a private speech act and a dialogue speech act renders the speech act known as "Interaction." In Interaction, the speaker tries to gain control over the listener, and so does the listener over the speaker. Thus, Interaction is a basis for a verbal struggle or heated discussion. Interaction also means reciprocal cooperation in which both parties work toward finding mutually agreeable solutions. Although each word in a sentence should not be used out of the context to identify the speech act of a sentence (Green, 2017), some of B&B's common examples for Interaction include "reassure," and "appease."

The combination of a private speech act and a monologue produces the speech act known as "Expression." In Expression, the speaker uses a private speech act without inviting the listener to join in the conversation. Thus, this speech act sounds like an unfiltered revelation of the speaker's emotional state to the listener rather than logical speech. Again, with caution against out-of-context interpretations, the examples that B&B provide are "express sympathy" and "address imperiously."

The combination of a socially entrenched speech act and a monologue results in the speech act known as "Appeal." In Appeal, the speaker tries to gain control over the listener using conclusive, pre-scripted speeches. The examples that B&B provide are "call attention to," "officially ask," and "demand." Lastly, the combination of a socially entrenched speech act and a dialogue speech act produces the speech act known as "Discourse." Discourse includes many mutual appeals and expressions and is lengthy and eloquent. Examples are "acknowledge," and "enter into a discussion."

Hypotheses on Conventionality and Dialogicality

To our knowledge, no prior study has provided assertions as to how B&B's speech acts manifest in an agent's response tone, creating positive emotions for the complainant. Due to the lack of prior studies, we referred to the literature on crisis

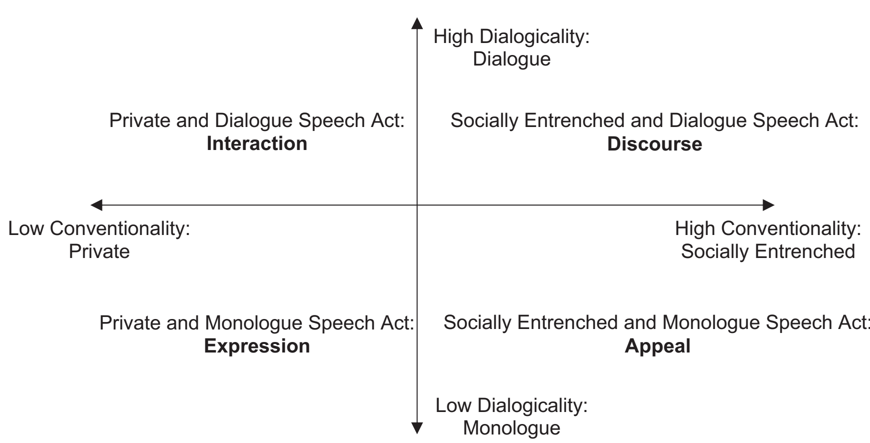

Fig. 1. Interaction between dialogicality and conventionality (Modified from table 4.20 in Ballmer and Brennenstuhl (1981), p. 31).

communication on social media to search for studies relevant to formulating hypotheses on the impact of B&B's speech acts on complainants' emotions. According to Kelleher and Miller (2006), customers prefer personalized tones on corporate social media sites. A personalized tone has positive outcomes in online reputation damage control (van Noort, Voorveld, & van Reijmersdal, 2012). Customers prefer a private tone on social media because communication over social media lacks the visual and social cues that exist in face-to-face or phone conversations (Keeling, McGoldrick, & Beatty, 2010). Customers want their conversations with companies to be personable (Yang, Kang, & Johnson, 2010). Twitter increases the expectation of a personalized voice to reply to complaints or inquiries rather than a formal, generic tone (McCorkindale, 2010).

The preference for a private speech act, as shown in the corporate crisis management field, can be explained by the limited affordances of twitter. Twitter interactions are asynchronous (Einwiller & Steilen, 2015) and do not guarantee a response from a company (Hu et al., 2019). Twitter interactions do not provide exclusive and individual attention to the parties involved in the interactions because the majority (80%) of twitter users are on mobile devices (Brandt, 2019), and thus are likely engaged in multiple tasks (Lee & Benbasat, 2003). To compensate for these limitations of twitter conversations, complainants may prefer a private speech act so that they can replicate intimate and exclusive face-to-face or phone communication in which they are given individual attention. As such, we expect that complainants on social media would prefer a private speech act to a socially entrenched speech act.

H1. In a given interaction, a private speech act in the agent's message will elicit positive emotions for the complainant to a greater degree than a socially entrenched speech act will.

Next, we explore which level of dialogicality alleviates complainants' emotions on social media. A monologue speech act is considered more desirable for corporate crisis communication on social media presumably because 80% of twitter users are on mobile devices (Brandt, 2019) and, thus, prefer short responses to elaborate ones. On twitter, a firm's replies ought to be precise and conclusive as in monologue rather than iterative as in dialogue (McCorkindale, 2010).

We found three studies that suggest the possibility that a monologue speech act would be preferable to a dialogue speech act on social media, although the studies do not empirically compare the two levels. Kim, Kim, and Nam (2014) analyzed how firms use twitter, and they found that firms tend to provide short answers quickly. Ballmer and Brennenstuhl (1981) have shown that asking for further information on twitter decreases complainants' satisfaction with a firm's remedy. Niu and Fan (2016) have claimed that agent responses that do not require customers to take further action have positive effects on the alleviation of customers' negative emotions. Complainants perceive the agent's conclusive and short answers as understanding of the urgency of their problems (Niu & Fan, 2016). Based on these studies, we expect that a monologue speech act will lead to more positive emotions for complainants than a dialogue speech act. Therefore, we hypothesize the following.

H2. In a given interaction, a monologue speech act in the agent's message will elicit positive emotions for the complainant to a greater degree than a dialogue speech act will.

Hypotheses on Interactions Between Conventionality and Dialogicality

As noted above, B&B have suggested that these two dimensions interact but did not posit how these interactions affect the listener's emotions. As described in the developments of the two prior hypotheses (H1 and H2), a combination of a private and monologue speech act would be preferable to any other combination given that a private speech act would be more effective than a socially entrenched one (H1) and that a monologue speech act would be more effective than a dialogue speech act (H2). Here, however, we argue otherwise, based on B&B's descriptions of the four speech acts (Fig. 1). The combination of a private speech act and a monologue speech act (as in Expression) will erode the individual benefits of each.

A private speech act combined with monologue to create Expression does not sound like an approach that considers the sensibilities of the listener but that focuses on the speaker's emotional state (Ballmer & Brennenstuhl, 1981). Expression can, therefore, sound inattentive.

In contrast, a private combined with dialogue speech act can meet the social media user's expectations of a response from an organization (Kelleher & Miller, 2006). According to B&B, although Interaction can mean a “hot” quarrel between an agent and a complainant, it also means that both parties pay attention to each other, laying a foundation for better behavior and mutual cooperation. Such exchanges of heated comments and responses are expected on social media platforms (Lee & Chau, 2018).

Furthermore, a private speech act can cancel out the negative effects of a dialogue speech act when combined in a single response. Recall that disadvantages of a dialogue speech act on social media include lengthy exchanges of posts, making the agent appear nonempathetic to the urgency of the complainant's problems (Niu & Fan, 2016). Despite this weakness, if a firm's agent uses both a private speech act and a dialogue speech act, the complainant may not find the agent nonempathetic because the use of a private speech act implies individual attention. Consider the following conversation excerpts as an example: “There is nothing I can do for you” (Expression) versus “There is nothing I can do for you, but can I do anything else for you to make you feel better?” (Interaction). As shown in this example, we posit that the positive impact of a private speech act will increase when it is combined with a dialogue speech act (Interaction).

A socially entrenched speech act, on the contrary, does not cancel out the shortcomings of a dialogue. A socially entrenched speech act alone is rhetorical and involves standard phrases. When it is combined with a dialogue speech act that involves invitations to a conversation, creating Discourse, this results in many redundant phrases. Recall B&B's examples of Discourse, mentioned above: “acknowledge,” and “enter into a discussion.” These phrases are likely to create redundant phrases when used on twitter. Consider the following tweeted example of Discourse and compare it to the above example of Interaction: “There is nothing I can do for you, but can I do anything else for you to make you feel better?” (Interaction) versus “Unfortunately, we are unable to provide further assistance to you. Please do not hesitate to contact us if you have any other questions or comments” (Discourse). The latter is lengthy, redundant, and does not convey the agent's empathy regarding the urgency of the complainant's problem. Thus, we do not anticipate that socially entrenched speech acts will augment the impact of dialogue. In sum, we hypothesize the following:

H3. There will be interaction effects between the conventionality and dialogicality of a speech act in such a way that, in a given interaction, private speech act combined with a dialogue speech act (Interaction) in the agent's message will elicit positive emotions for the complainant to a greater degree than a private speech act combined with a monologue speech act (Expression).

Method

Data Collection

Data collection began with our reviewing the companies included in the ACSI list, which was created by the University of Michigan and is currently the only national cross-industry measure of customer satisfaction in the US (www.theacsi.org). From this ACSI list, we selected all the companies that had twitter accounts dedicated to handling complaints. We omitted all-purpose twitter accounts from our analysis. For example, Citi Group has two twitter accounts: @AskCiti to handle customer complaints and @Citi to share information with the public. In this case, we included only @AskCiti in our dataset. As a result, 34 companies were selected (Appendix A), which was, to the best of our knowledge, the complete coverage of the companies that have engaged with twitter customer complaint management at the time of our data collection. The 34 companies cover a wide range of industries, including accommodation and food services, arts and entertainment, finance and insurance, information, manufacturing, retail trade, and transportation and warehousing.

For each customer care twitter account, we collected the most recent tweets from the three years prior to the date we began our data collection. Twitter's API allows for the retrieval of an account's last 3,200 tweets. We circumvented this limit by using an automatic scroll-down function (Selenium with Python). As a result, the number of retrieved tweets per company varied depending on the firm's activity level. Of the 34 companies in our sample, 14 companies had more than 3,200 collected tweets. We crawled a total of 102,407 tweets from 34 companies. Appendix A presents the breakdown of the final sample by industry.

Extraction of Interactions for Complaint Handling

As we are interested in how an agent's speech acts influence the complainant's emotions in the immediate aftermath of a customer's response to the agent's message, we needed to extract an “interaction.” Previously, we conceptually defined an interaction as a pair of messages exchanged between an agent and a customer involved in the resolution of the customer's issue. An interaction is operationalized as a pair comprised of an agent's tweet and the subsequent complainant's tweet, which forms part of the thread (or conversation) involving the complainant's issue. Based on this definition and the accompanying operationalization, we took the following steps to extract interactions from the numerous and complex multiple-party and multi-topical tweets we had collected.

First, we needed to screen out tweets that were unrelated to complaint handling from our dataset. Although we downloaded tweets from customer care twitter accounts only, these tweets still included generic conversations (e.g., “Hello everyone!”). In order to eliminate irrelevant tweets, we limited our data to “conversations,” defined as “a series of interactions between an agent and the corresponding customer on a single complaint sorted in chronological order.” By this definition, conversations must meet: (1) be initiated by a complainant, (2) have at least one agent response tweet, and (3) include at least two complainant tweets (the complainant's initiating tweet as well as their response to the agent's tweet).

Once we extracted conversations, we broke each conversation into a series of interactions (as defined above) in chronological order. It was necessary to remove outliers related to: (1) time lag between the interactions within a conversation and (2) the number of interactions in a conversation. The former involves the agent's delayed responses, and the latter suggests complicated complaints escalating the complainant's frustration (Orthaber & Márquez-Reiter, 2011). To avoid confounding our analysis with these extraneous variables, we applied Miller's (1991) contention that values greater than three standard deviations from the mean can be interpreted as outliers and thus can be removed from analysis. As such, we removed those conversations that had longer than 42.95-day time lag (versus mean 1.75 day) between the interactions or had over 14 interactions (versus mean 2.98 interactions) per conversation.

As a result, of the 102,407 initially crawled tweets, we extracted 13,446 conversations and 34,709 interactions comprising of 86,744 tweets. Table 1 summarizes the final sample.

Independent Variables: A Firm's Speech Acts

Development of Automatic Speech Act Classification Models

To classify firms' speech acts, we proposed two RNNs (Lipton et al., 2015) for the two dimensions (i.e., conventionality and dialogicality), respectively. The proposed model is illustrated in Fig. 2. Before being processed by the deep learning models, tweets were transformed by an embedding matrix to obtain corresponding word vectors. For this study, a pre-trained GloVe embedding matrix (Pennington, Socher, & Manning, 2014) was employed to generate 200-dimensional

Table 1
Description of sample.

Types of data	Counts
Interactions (turns) consisting of tweets	34,709
Conversations consisting of interactions	13,446
Complainant emotion	34,709
Negative (−1)	14,034 (40%)
Neutral (0)	10,610 (31%)
Positive (+1)	10,065 (29%)
Agent speech act	34,709
Conventionality: Private speech act (1)	24,050 (69%)
Socially entrenched speech act (0)	10,659 (31%)
Dialogicality: Monologue speech act (1)	8,520 (25%)
Dialogue speech act (0)	26,189 (75%)

word vectors. The GloVe embedding matrix was pre-trained on a twitter dataset that contained 1.2 million vocabulary terms. The word vectors were then passed through a bi-directional recurrent layer with 128 gated recurrent units (Bi-GRU) (Chung, Gulcehre, Cho, & Bengio, 2014) to obtain corresponding bi-directional hidden representations. To help the model pay attention to important words, a word-level attention layer was added on top of the previous bi-directional recurrent layer. After that, the weighted hidden features were passed through two fully connected layers with 128 and 64 hidden units, respectively. Each hidden layer was followed by a batch normalization layer (Szegedy, Ioffe, Vanhoucke, & Alemi, 2017) and a drop-out layer (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014). The batch normalization layer improved the model training speed, improved performance, and made the model training stable. The drop-out layer prevented the overfitting issue during model training. The final output of the proposed model was a continuous value between 0 and 1. For the conventionality dimension, if output values were greater than 0.5, the output results would be predicted as private. Otherwise, the results would be considered as socially entrenched. For the dialogicality dimension, if output values were greater than 0.5, the output results would be considered as monologue, and if values were less than 0.5, the results would be considered as dialogue.

To train the proposed models for the conventionality dimension, 2,000 tweets by agents were randomly selected and labeled by three experts as either private or socially entrenched. The final label was determined by the majority-voting method to reduce each human annotator's errors and subjectivity. The voting method is a common technique that uses multiple models to obtain better performance than could be obtained from any individual model (Onan, Korukoğlu, and Bulut 2016a; Onan, Korukoğlu, and Bulut 2016b). For the dialogicality dimension, another set of 2,000 tweets by agents were randomly selected and labeled by the three experts, and the final label was determined by the majority-voting method, in the same manner as for conventionality. Examples of the labels are presented in Table 2.

During model training, for each task (i.e., classifying as either conventionality or dialogicality), 1,800 samples were evenly selected for the model training, and 200 (the rest of the tweets) were used to validate the performance of the two models. After the models were trained for 50 epochs, the proposed models achieved 84.50% testing accuracy for classification of conventionality and 93.00% testing accuracy for classification of dialogicality (Table 3). In addition, we tried several relevant deep learning-based models to compare the performance. These models were: (1) a two-layer multilayer perceptron model (España-Boquera, Castro-Bleda, Gorbe-Moya, & Zamora-Martinez, 2011) with 128 and 64 hidden units, respectively; (2) a one-dimensional convolutional neural network (Krizhevsky, Sutskever, & Hinton, 2017) with one convolutional layer of 128 filters and a kernel size of 1 × 3; and (3) a recurrent neural network with 128 gated recurrent units. The testing results for these methods are summarized in Table 3, from which we can observe that the proposed model (Bi-GRU-RNN) outperforms the other models. Therefore, we chose these two proposed models to complete the classification of the remaining agent tweets.

Calculation of a Speech Act in An Interaction

We assigned the value of 1 if it had a private speech act and 0 if it had a socially entrenched speech act. Similarly, we assigned 1 if the tweet had a monologue speech act and 0 if it had a dialogue speech act.

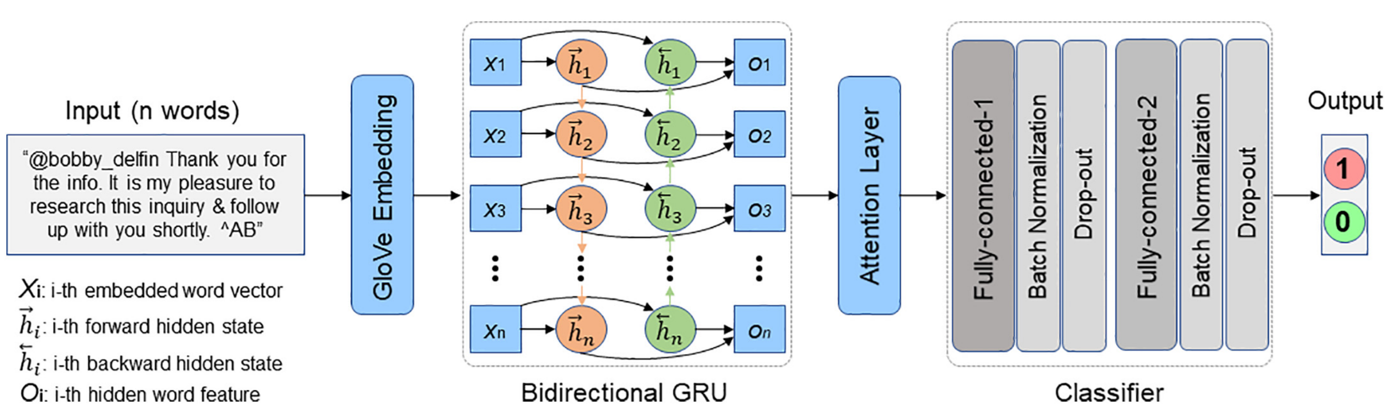

Fig. 2. An illustration of the proposed bidirectional recurrent neural network-based model architecture.

Table 2
Examples of labels for conventionality and dialogicality.

Private	***Socially entrenched***
• Karen Radley Acura's Service Dept can be reached at 866–979-9,863. I can call them tomorrow & request they follow up. • There are several #Acura dealers in your area: Woodbridge, Falls Church & Chantilly. Which location is convenient for you? • Great photo of your Acura, thanks for sharing William! Which dealership is your vehicle being serviced at?	• I recommend speaking with Insurance Company about the extent of damage & they can confirm if involving Acura is necessary • Please note an #Acura dealership is the only authorized facility for warranty repairs. We hope that helps. • It would be our pleasure to help. Please direct message your full name, address, and phone number.
Monologue	***Dialogue***
• Certainly Jennifer we have replied via direct message with the contact info as requested. • I certainly understand the frustration & we've noted ur feedback. I recommend checking back on our site for future updates. • I appreciate you asking, but we are unable to recommend modifications to your vehicle. I apologize for the inconvenience.	• Greetings Rajeev, were you going to be towing the vehicle to the #Acura dealership? Is there something I can help with? • Hi Susan, I regret to learn of your negative experience. Please share more details regarding your concern. • We didn't hear back from you Missy. It would be our pleasure to assist you. Let us know if our involvement is needed.

Dependent Variable: A Complainant's Emotion Elicited by the Agent's SA

Classification of a Complainant's Emotion in Each Tweet

To recognize the kind of emotion (negative, neutral, or positive) in complainants' tweets, a multi-class RNN was proposed. The proposed model is like the model architecture illustrated in Fig. 2 except for the final output layer with 3 units for the emotion classification task. The complainants' tweets were processed by the pre-trained GloVe embedding matrix to obtain corresponding 200-dimensional word vectors. After that, the word vectors were passed through a bi-directional recurrent layer containing 128 gated recurrent units (Bi-GRU) to obtain bi-directional context information. One word-level attention layer was appended to help the model focus on important words during the model training and generate weighted hidden features. These hidden features were then passed through two fully connected layers with 64 and 32 hidden units, respectively. Each fully connected layer was followed by a batch normalization layer and a drop-out layer. The final output was a three-dimensional vector that indicated the predicted emotion category of the input tweet. For example, one output vector with values of [1, 0, 0] denoted the input tweet was classified as a negative emotion, following a "one-hot" coding scheme of [negative, neutral, positive] for our task.

To train the proposed model, 1,320 tweets were randomly selected from the dataset, and they were labeled by three experts. The final label was determined by the majority-voting method, yielding a dataset consisting of 474 negative tweets, 439 neutral tweets, and 407 positive tweets. Examples of these labels are presented in Table 4. Then, 1,100 tweets were evenly selected as the training set, and 220 (the rest of the tweets) were used to evaluate the model performance. After the model was trained for 100 epochs, the proposed model achieved 80.90% testing accuracy. In addition, we tested three other deep learning models for the complainants' emotion classification task. These methods included the two-layer multilayer perceptron model, the one-dimensional convolutional neural network, and the recurrent neural network. The corresponding test results of these methods are exhibited in Table 5. The final testing results of the proposed model (i.e., Bi-GRU-RNN) outperforms the other deep learning models. Therefore, we decided to use the proposed model to predict the remaining complainants' emotions.

Table 3
Accuracy testing results of speech act classification algorithm.

	Conventionality	Dialogicality
Deep learning models	Embedding (GloVe-Twitter)	Embedding (GloVe-Twitter)
2-layer-MLP	72.00%	77.50%
1D-CNN	79.50%	86.00%
GRU-RNN	82.50%	89.00%
Bi-GRU-RNN (proposed)	84.50%	93.00%

Calculation of a Complainant's Emotion in An Interaction

The results from the above classification of a complainant's emotion were used to measure the dependent variable. We assigned one of the following values to the complainant's emotion: −1 (negative), 0 (neutral), or 1 (positive).

Control Variables

As shown in Table 6, we controlled for 19 time-varying factors found in prior studies to affect a complainant's emotion over the course of a conversation (Gunarathne et al., 2017). Each of our 19 control variables can be put into one of five categories —content, complainant, company agent, interaction, and temporal (calendar) variables. By controlling these variables that potentially affect a complainant's emotions, we can better tease out the effects of speech acts on the complainant's emotion.

Notably, we controlled for the most common types of content that appeared in an agent's response. The most commonly found content of a response includes an apology, handoff, gratitude, or request for a direct message (DM, Gunarathne et al., 2017). In extracting the contents, three coders identified common linguistic

Table 4
Examples of labels for complainants' emotions.

Negative emotion	Neutral emotion	Positive emotion
• I guess, I have tried all those steps more than 20 times in last couple of months! Result is same! • And now my luggage isn't on the flight. How does this all in one day? • I don't detest bank of America it is their lack of integrity in how they treat their employees is what I loathe that & when they steal money	• Is there any difference if i pre-order WoD or if i buy it when it comes out? • When do you open to assist passengers at BSB Brasilia • I contacted them and filled in every thing what should I do next??	• I love Aetna and Aetna staff. I am sure there must be thousands of tech staff behind this gr8 company. • Also, huge shoutout to @AcuraClientCare for regular outside-the-box thinking/ideas/support. #Acura brand-loyal in big part to them. #cheers • Thanks! That is very helpful:)

Table 5
Accuracy testing results of sentiment analysis used for classifying complainants' emotions.

Method	2-layer-MLP	1D-CNN	GRU-RNN	Bi-GRU-RNN (proposed)
Accuracy	69.55%	73.18%	78.64%	**80.90%**

patterns that appeared in the minimum of 200 tweets per content type. Then, a template-based pattern matching program was developed to select the tweets that contain those patterns using regular expressions in Python. The results were validated by the same coders on the 200 randomly selected tweets per content. This process was repeated until the three coders reached an agreement that the results were 100% accurate and sufficiently extracted all the linguistic patterns of each content type. Appendix B provides the correlation table listing these variables.

In addition, we controlled for message concreteness, which is the overall average concreteness score of all the words included in a tweet (=the sum of the concreteness score of each word in a tweet/the total number of words in the tweet). Concreteness of a word refers to whether a word is specific and definite rather than abstract and generic (Brysbaert, Warriner, & Kuperman, 2014). Message concreteness was controlled because some customers may prefer concrete responses to general ones. In addition, we controlled for complainants' emotions in prior interactions because this could affect their emotions in subsequent interactions.

Table 6
Descriptions of control variables.

Category	Variable	Operationalization
Content	Apology	Whether the agent apologized (1 = Yes, 0 = No)
	Handoff	Whether the agent handed the focal interaction to another party (1 = Yes, 0 = No)
	Gratitude	Whether the agent expressed gratitude (1 = Yes, 0 = No)
	DM	Whether the agent mentioned direct messaging (1 = Yes, 0 = No)
Complainant	Private speech act (complainant)	Whether the complainant used a private speech act (1 = Yes, 0 = No)
	Monologue speech act (complainant)	Whether the complainant used a monologue speech act (1 = Yes, 0 = No)
	Message length (complainant)	The number of characters used in the complainant tweet
	Previous emotion	The sentiment of the customer in the previous interaction
Company agent	Message concreteness	The extent to which the agent tweet was concrete (ranging from 0 to 1) determined by the overall concreteness score (= sum of the concreteness score for each word in a Tweet / the total number of words in the Tweet)
	Message length (agent)	The number of characters used in the agent tweet
Interaction	Turn order	The number of previous interactions (turns) before the focal interaction in the given conversation
	Consecutive agent tweets	Whether consecutive agent tweets exist (1 = Yes, 0 = No)
	Consecutive complainant tweets	Whether the focal interaction contains a complainant tweet only –i.e., whether complainant tweets were made in a row (1 = Yes, 0 = No)
	Interaction length (in minutes)	Time elapsed since the last interaction (i.e., time between the previous customer tweet and the focal previous tweet)
	Number of retweets	The total number of retweets the focal interaction generated
	Number of likes	The total number of likes the focal interaction generated
Calendar	Weekend interaction	Whether the interaction took place on weekend (Sat or Sun) (1 = Yes, 0 = No)
	Year dummy	A dummy variable for the calendar year when the customer tweet was made in the current interaction
	Month dummy	A dummy variable for the calendar month during which the interaction occurred

Model Specification

As our interest centers on the effects of an agent's speech acts on the complainant's emotion in the immediate response, our unit of analysis is an interaction (as defined in the section, *Extraction of Interactions for Complaint Handling*). For the analysis, we employed (interaction-level) fixed-effect regression estimation with robust standard errors. This fixed-effects estimation approach permits us to compare only interactions belonging to a conversation to one another; thus, statistically significant results indicate that consistent patterns exist in interactions across conversations. This fixed-effect regression estimation provides at least two advantages to answering our research question.

The first advantage is that this fixed-effect regression estimation allows us to observe dynamic changes in agents' choices of speech acts and the accompanying emotions in complainants' responses. An agent's choices of speech acts and the complainant's subsequent emotions are likely to fluctuate over the course of a conversation. A fixed-effects estimation model attributes the variance found in the complainant's emotions to the variance in the agent's speech acts. Thus, a significant result obtained from this estimation model demonstrates that an agent changes their use of speech acts during the course of a conversation (i.e., from an interaction to another) and the customer's emotion changes accordingly.

The other advantage is that the fixed-effect estimation model can eliminate complainant-specific unobserved heterogeneity such as the complainant's demographics, network popularity, and personality. The estimation can also control for time-invariant complaint-, agent-, and company-specific factors because they are, by definition of our interaction, kept constant in our analysis. Examples of complaint-, company- and agent-specific unobserved heterogeneity are types of complaints, company culture, policies regarding tweeting, the company's products or services, the industry in which the company operates, and the agent's demographics.

As a result of these strong controls implemented, this model allows us to make a causal inference regarding the effect of an agent's speech acts on a complainant's emotion. Specifically, we model complainant *i*'s emotion in interaction *k* as a function of agent *j*'s speech acts in that interaction as shown below:

$$\begin{aligned} \textit{Complainant Emotion}_{ijk} = \beta_0 &+ \beta_1 \textit{Private SA}_{ijk} \\ &+ \beta_2 \textit{Monologue SA}_{ijk} \\ &+ \beta_3 \textit{Private SA}_{ijk} \\ &* \textit{Monologue SA}_{ijk} + \gamma_1 X_{ijk} \\ &+ \alpha_k + \varepsilon_{ijk} \end{aligned} \tag{5}$$

where SA refers to Speech Act, X_{ijk} denotes the control variables and α_k represents interaction *k*'s fixed effects.

Results

Table 7 shows the descriptive statistics of the major variables. Before the analysis, a natural-log transformation was applied to highly skewed variables, such as the number of retweets, the number of likes, and the time lag between interactions. The variance inflation factors for all variables were below 2.0, so multicollinearity was not a concern.

Table 7
Descriptive statistics.

Variable	Mean	Std. dev.	Min	Max
Private speech act (complainant)	0.55	0.50	0	1
Monologue speech act (complainant)	0.79	0.41	0	1
Interaction length (in minutes)	1,042.41	4,185.48	0	49,981
Message length (agent)	120.85	51.25	11	438
Message length (complainant)	90.30	40.72	6	324
Message concreteness	0.19	0.50	0	2.53
# of retweets	0.31	6.79	0	337
# of likes	0.55	10.12	0	564
Turn order	4.49	3.12	1	13
Consecutive agent tweets	0.03	0.17	0	1
Consecutive complainant tweets	0.58	0.49	0	1
Weekend Interaction	0.15	0.35	0	1
Apology	0.08	0.28	0	1
Handoff	0.03	0.18	0	1
DM	0.01	0.09	0	1
Gratitude (agent)	0.17	0.38	0	1
Previous emotion	−0.22	0.82	−1	1
Private speech act (agent)	0.31	0.46	0	1
Monologue speech act (agent)	0.75	0.43	0	1
Complainant emotion	−0.11	0.83	−1	1

Notes: N = 34,709 (interactions).

Table 8 reports the results of the regression analysis. Model 1 presents the base model with the control variables only. Model 2 adds the main effects of the speech acts to the base model. Model 3 presents the full model, including the interaction term. Further, *R*-squared values show the strength of the relationship between our model and the dependent variable. We also checked Akaike's information criterion (AIC), which is a measure of goodness-of-fit based on the tradeoff between the complexity and precision of a model (Akaike, 1974). The smaller the AIC value, the better the model is. Compared to the other models (Model 1: 62557.53 and Model 2: 62547.81), Model 3 has the lowest AIC value (62543.02), suggesting that the full model best fits the data.

Hypothesis 1 states that an agent's private speech act will create positive emotions for the complainant to a greater degree than a socially entrenched speech act will. The coefficient of a private speech act is positive and significant ($\beta = 0.028$, $p < .05$ in Model 2), thus supporting Hypothesis 1.

Hypothesis 2 states that an agent's monologue speech act will create positive emotions for the complainant to a greater degree than a dialogue speech act will. The coefficient of a monologue speech act is positive and significant ($\beta = 0.032$, $p < .05$ in Model 2), thus supporting Hypothesis 2.

Hypothesis 3 states that when a private speech act is combined with a dialogue speech act, the complaints' emotions are likely to be more positive than when it is combined with a monologue speech. We tested the interaction effect between a private speech act and a monologue speech act. Model 3 shows that the coefficient of the interaction term (private * monologue) is negative and significant ($\beta = -0.069$, $p < .05$), suggesting a substitution effect between the two speech acts.

Table 8
Results from the regression.

Variables	(1) Controls	(2) Main effects	(3) Interaction
Constant	−0.757***(0.087)	−0.747***(0.087)	−0.752***(0.087)
Private SA (complainant)	0.048***(0.011)	0.049***(0.011)	0.049***(0.011)
Monologue SA (complainant)	−0.142***(0.014)	−0.143***(0.014)	−0.142***(0.014)
Interaction length (in minutes) (ln)	−0.002(0.001)	−0.002(0.001)	−0.002(0.001)
Message length (agent)	0.026(0.019)	0.033(0.019)	0.033(0.019)
Message length (complainant)	−0.327***(0.011)	−0.327***(0.011)	−0.327***(0.011)
Message concreteness	−0.021(0.040)	−0.018(0.040)	−0.018(0.040)
# of retweets (ln)	0.000(0.013)	0.000(0.013)	0.000(0.013)
# of likes (ln)	−0.007(0.009)	−0.007(0.009)	−0.007(0.009)
Turn order	0.015***(0.002)	0.015***(0.002)	0.015***(0.002)
Consecutive agent tweets	0.035(0.040)	0.034(0.040)	0.035(0.040)
Consecutive complainant tweets	−0.028(0.028)	−0.028(0.028)	−0.028(0.028)
Weekend Interaction	0.019(0.025)	0.020(0.025)	0.021(0.025)
Apology	−0.045(0.027)	−0.047(0.027)	−0.047(0.027)
Handoff	0.053(0.036)	0.049(0.036)	0.050(0.036)
DM	0.042(0.076)	0.045(0.076)	0.045(0.076)
Gratitude (agent)	−0.037*(0.018)	−0.040*(0.018)	−0.039*(0.019)
Previous sentiment	−0.212***(0.007)	−0.213***(0.007)	−0.213***(0.007)
Private SA		**0.028*(0.014)**	0.029*(0.015)
Monologue SA		**0.032*(0.015)**	0.032*(0.015)
Private*Monologue			**−0.069*(0.034)**
Interactions	34,709	34,709	34,709
Conversations	13,446	13,446	13,446
Company FE	Yes	Yes	Yes
Year/Month dummies	Yes	Yes	Yes
R-squared (within)	0.0996	0.1001	0.0999
AIC	62,557.53	62,547.81	62,543.02

Notes: SA = Speech Acts; Robust standard errors are shown in parentheses; ln = Log transformed; FE = fixed effects, * $p < .05$, ** $p < .01$, *** $p < .001$.

This substitution effect coincides with the assertion underlying H3. The positive effect of a private speech act on a complainant's emotion decreases when combined with a monologue speech act. A simple slope test (Aiken & West, 1991) reveals that the difference between private speech acts combined with dialogues and private speech acts combined with monologues is significant ($t = 2.08$, $p < .05$). Taken together, the empirical evidence supports H3.

Robustness Checks

To test the robustness of our findings, we conducted robustness tests. The results are in Appendix C. First, we used an alternative operationalization of the dependent variable, measuring a change in the complainant's emotion based on the difference between the complainant's emotion in the previous interaction and that of the current interaction. This alternative dependent variable allowed us to examine how the agent's current use of a specific speech acts makes the complainant feels better or worse (compared to the previous interaction) while controlling for the customer's previous emotional state. The result supports that both private and monologue speech acts significantly improve the complainants' emotions in the current interaction compared to the prior one ($\beta = 0.028$, $p = .054$; $\beta = 0.035$, $p < .05$, respectively; Model 1.2). Additionally, the result supports that private and monologue speech acts substitute each other when used jointly ($\beta = -0.084$, $p < .05$, Model 1.3).

Second, we tested the possible interaction effects between the speech acts of a complainant and those of an agent to rule out an alternate hypothesis that a complainant's emotion results from the interaction between the agent's speech acts and the complainant's own speech acts, instead of resulting solely from the agent's use of speech acts. Model 2 in Table C reports the results of the four possible interaction effects. None of the interactions were significant. Notably, the addition of the four interactions did not affect the significance of our main results ($\beta = -0.084$, $p < .05$), attesting to the robustness of the findings.

Discussion and Conclusion

Summary of the Findings

We began this study with the goal of identifying the response tone that an agent can employ on social media to handle complaint management processes without resorting to excessive or premature provisions of redress. To this end, we employed the theory of speech acts, which concerns the semantics of speech, due to its fit with our definition of a response tone and our study goal. We chose B&B's speech act, which provides the two dimensions of speech acts and the interactions between the two dimensions, contextualizing these two dimensions as well as the interactions in line with our conceptualization of response tones. Despite these strengths, B&B's classification is unclear regarding how the listener in a negotiation would feel when a speech act is delivered to them. Filling this gap is important in social media complaint management because complainants' negative emotions can create the viral spread of negative WOM (Kramer, Guillory, & Hancock, 2014), while complainants' positive emotions are associated with increased brand loyalty (DeWitt et al., 2008). Thus, we expand B&B's classification to cover the impacts of speech acts on a listener's emotions. Our data (>100,000 tweets)—collected from a complete coverage (34) of the firms on the ACSI list with twitter complaint handles, three RNNs, and fixed-effect regression results—demonstrate that each speech act and their combinations vary in their capacity to create positive emotions for the complainant. More specifically, both private speech acts and monologue speech acts foster positive emotions for the complainant compared to socially entrenched and dialogue speech, respectively. In addition, the response to a private speech act combined with a monologue speech act (forming Expression), is less positive than when private speech is combined with dialogue speech (forming Interaction). These substitution effects expand B&B's assertions regarding the interaction effects between the two dimensions by showing which combinations erode the advantages of an individual speech act.

Contributions to Theory Advancement

This study provides several theoretical contributions. First, this study closes an important gap on the effectiveness of a proper response tone in the creation of positive emotions for complainants, contributing to the knowledge of complaint management on social media. Among the few studies that have examined corporate response strategies to service failures, primary attention has been given to response contents rather than tones. We contribute to the literature by demonstrating the importance of using a specific pair of speech acts to create positive emotions for the complainant. This result expands our understanding of proper response strategies to handling complaints by redirecting our attention from response contents to response tones.

Second, we contribute to expanding speech act theory in three areas: (1) the effects of speech acts on the listener's emotions, (2) the substitution effects between conventionality and dialogicality, and (3) complaint management. No prior studies, including B&B's, have proposed or tested any of these three propositions. B&B's focus was on a speaker's intentions and did not consider a listener's emotional reactions to the speech. Furthermore, although B&B introduced the interactions between the two dimensions of speech acts, they did not explain substitution effects between conventionality and dialogicality. In contrast, our finding shows that a particular set of combinations can erode the benefits of an individual speech act. We also extend speech act theory to the context of complaint management by building a bridge between automated complaint handling systems and the fast-growing AI field; this is because B&B's classification is useful for designing responses for automated agents involved in negotiations with a human listener (Chang & Woo, 1994; Woo & Chang, 1992). Spanning the boundaries between complaint management and AI will foster the advancement of knowledge in the emerging automated complainant support field.

Last, this study makes methodological contributions by demonstrating the benefits of deep learning for the interactive marketing field. Although many prior marketing studies have adopted machine learning algorithms, they have centered on topic identification and sentiment analysis. This study tackles a deeper and more challenging problem: the semantics of a tweet and the interactions between complainants and agents. In addition, our emotion RNN architecture overcomes the limitations of sentiment analyses that are reliant on a dictionary-based approach and which can therefore be easily outdated as syntax changes over time.

Implications for Practice

This study also provides important practical implications. First, we identify the specific response tones that lead to happy customers on social media. These findings enable firms to reduce the costs of providing financial redress and to prevent negative emotional contagion on social media. While acknowledging the importance of response contents (e.g., financial redress), our study highlights that service agents can effectively handle the complaint management process to some extent simply by using proper levels of dialogicality and conventionality.

Second, our findings help develop guidelines for training customer service agents. Using a private speech act on twitter is generally better than using a socially entrenched speech act. Nonetheless, it is safe to assume that many companies train customer service agents with pre-established scripts before they deploy the agents to interact with complainants (Nguyen, Groth, Walsh, & Hennig-Thurau, 2014; Walsh, Gouthier, Gremler, & Brach, 2012). Therefore, the challenge lies in providing customer service agents with various scenarios and possible responses which they can select so that the agents' replies ought to sound spontaneous; in this way, customers are not aware of the fact that agents' responses are based on a script.

Our study demonstrates the merit of utilizing a deep learning emotion classification model to gauge how successful a service agent is in addressing complaints. This model can be a replacement to the additional step of administering a post-service customer survey that suffers from selection and non-response biases. Firms can use the model to assess an agent's quality of service in a more systematic and transparent manner. Lastly, firms can utilize deep learning speech act classification models to assess the degree to which each agent follows the guidelines for using specific speech acts. Simultaneously, the results generated by these algorithms (e.g., “You used private speech acts 80% of the time”) will help agents reflect on their personal tendencies to use one speech act over the other.

Limitations and Suggestions for Future Research

Like all studies, this study has limitations to acknowledge. We operationalized our dependent variable by analyzing complainants' explicit expressions of emotions. Our operationalization was thus limited in its capacity to capture a complainant's use of sarcasm and politeness to disguise their feelings. By comparing changes in emotions from the same complainant across interactions in the same conversation, we managed to account for individual variance in expressing emotions—that is, even a polite person could show their changing emotions within their personal range. Despite the advantages of our chosen model in controlling for individual differences, the absence of complainants' individual characteristics in our hypothesis testing remains a limitation.

The second limitation is that we focused on emotions, rather than objective retweets and likes. This choice was made to accomplish our primary goal of identifying speech acts to handle a complainant's emotional state. However, we acknowledge that other objective measures may be appropriate for evaluating effects on the audience that observes agent-complainant interactions on social media.

This point leads to our third limitation and presents a promising area for future research. It is worthwhile to examine how service agents' speech acts used to handle complaints affect observers' attitudes toward the firm and their subsequent behaviors on social media (e.g., retweeting, liking). Unlike offline settings, how agents handle complaints is visible not only to individual complainants but also to the audience observing agent-complainant interactions (Johnen & Schnittka, 2019). Potential research questions include “Could the speech acts also garner positive responses from the audience and prevent negative WOM among the audience?” Indeed, observers of an agent-complainant interaction often also engage with the interaction, thereby forming a multi-party negotiation. It is not yet known whether B&B's classification centering on bi-directional negotiations is an appropriate theoretical lens for multi-party negotiations. Future researchers are recommended to consider extending the speech act theory beyond their focus on bi-directional negotiations and to identify if any additional dimensions outside conventionality and dialogicality are pertinent to multi-party negotiations.

Lastly, future researchers can consider whether communication accommodation occurs in twitter complaint management settings. Communication accommodation has been found on twitter in previous studies in that members with a shared social identity tend to mimic the other party's communication style (Danescu-Niculescu-Mizil, et al. 2011; Tamburrini et al. 2015). Our study did not center on individuals sharing a social identity and thus our results did not show mimicry (Appendix C); however, it is possible that companies with strong brand cultures that extend to customers could exhibit communication accommodation between an agent and the complainant.

Acknowledgment

This work was supported in part by the National Science Foundation grant, CNS 1551004.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.intmar.2021.02.001.

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA, US: Sage Publications, Inc.

Akaike, H. (1974). A new look at the statistical model identification. *Springer Series in Statistics, Springer, 215–222* *https://link.springer.com/chapter/10.1007%2F978-1-4612-1694-0_16*.

Austin, J. L. (1962), How to Do Things with Words, J. O. Urmson and M. Sbisá, eds., 2nd ed, Cambridge, MA: Havard University Press, https://www.hup.harvard.edu/catalog.php?isbn=9780674411524.

Austin, J. L. (1970). In J.O. Urmson & G.J. Warnock (Eds.) *Philosophical Papers,* Oxford, England: Oxford University Press.

Ballmer, T. T., & Brennenstuhl, W. (1981). Speech Act Classification: A Study in the Lexical Analysis of English Speech Activity Verbs. , Berlin, Heidelberg: Springer Berlin Heidelberghttps://doi.org/10.1007/978-3-642-67758-8.

Beebe, L. M., & Waring, H. Z. (2002). The pragmatics in the interlanguage pragmatics research agenda: the case of tone. *AAAL Annual Conference*, 6–9.

Beebe, L. H., & Waring, H. Z. (2004). The linguistic encoding of pragmatic tone: adverbials as words that work. In D. Boxer & A.D. Cohen (Eds.) *Studying Speaking to Inform Second Language Learning* (pp. 228–228) *Multilingual Matters*, .

Brandt, M. (2019). Infographic: 80% of Twitter's users are mobile. *Statista Infographics* https://www.statista.com/chart/1520/number-of-monthly-active-twitter-users/.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46*(3), 904–911.

Honey, C., & Herring, S. C. (2009). Beyond microblogging: Conversation and collaboration via Twitter. *2009 42nd Hawaii International Conference on System Sciences*, 1–10, https://doi.org/10.1109/HICSS.2009.89.

Chang, M. K., & Woo, C. C. (1994). A speech-act-based negotiation protocol: design, implementation, and test use. *ACM Transactions on Information Systems (TOIS), 12*(4), 360–382, https://doi.org/10.1145/185462.185477.

Chebat, J.-C., & Slusarczyk, W. (2005). How emotions mediate the effects of perceived justice on loyalty in service recovery situations: an empirical study. *Journal of Business Research, Special Section on the 2002 SMA Retail Symposium, 58*(5), 664–673, https://doi.org/10.1016/j.jbusres.2003.09.005.

Chung, J., Gulcehre, C., Cho, K. H., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *ArXiv:1412.3555 [Cs], December* http://arxiv.org/abs/1412.3555.

Danescu-Niculescu-Mizil, C., Gamon, M., & Dumais, S. (2011). Mark my words! Linguistic style accommodation in social media. *Proceedings of the 20th International Conference on World Wide Web*, 745–754, https://doi.org/10.1145/1963405.1963509.

Davidow, M. (2000). The bottom line impact of organizational responses to customer complaints. *Journal of Hospitality & Tourism Research, 24*(4), 473–490, https://doi.org/10.1177/109634800002400404.

Davidow, M. (2003). Organizational responses to customer complaints: what works and what doesn't. *Journal of Service Research, 5*(3), 225–250, https://doi.org/10.1177/1094670502238917.

DeWitt, T., Nguyen, D. T., & Marshall, R. (2008). Exploring customer loyalty following service recovery: the mediating effects of trust and emotions. *Journal of Service Research, 10*(3), 269–281, https://doi.org/10.1177/1094670507310767.

Einwiller, S. A., & Steilen, S. (2015). Handling complaints on social network sites–an analysis of complaints and complaint responses on Facebook and twitter pages of large US companies. *Public Relations Review, 41*(2), 195–204.

España-Boquera, S., Castro-Bleda, M. J., Gorbe-Moya, J., & Zamora-Martinez, F. (2011). Improving offline handwritten text recognition with hybrid HMM/ANN models. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 33*(4), 767–779, https://doi.org/10.1109/TPAMI.2010.141.

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

Goldstein, J., & Sabin, R. E. (2006). Using speech acts to categorize email and identify email genres. *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)*. IEEE.

Green, M. (2017). Speech acts. In E.N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy, Winter 2017*, Stanford, CA: Metaphysics Research Lab, Stanford University https://plato.stanford.edu/archives/win2017/entries/speech-acts/.

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

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

Hansen, N., Kupfer, A.-K., & Hennig-Thurau, T. (2018). Brand crises in the digital age: the short- and long-term effects of social media firestorms on consumers and brands. *International Journal of Research in Marketing, 35* (4), 557–574, https://doi.org/10.1016/j.ijresmar.2018.08.001.

He, S., Rui, H., & Whinston, A. B. (2017). Social media strategies in product-harm crises. *Information Systems Research, 29*(2), 362–380, https://doi.org/10.1287/isre.2017.0707.

Herhausen, D., Ludwig, S., Grewal, D., Wulf, J., & Schoegel, M. (2019). Detecting, preventing, and mitigating online firestorms in brand communities. *Journal of Marketing, 83*(3), 1–21, https://doi.org/10.1177/0022242918822300.

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

Jiang, H., & Huhns, M. N. (2005). An approach to broaden the semantic coverage of ACL speech acts. *Proceedings of International Conference on Conceptual Modeling*. Springer.

Johnen, M., & Schnittka, O. (2019). When pushing back is good: the effectiveness of brand responses to social media complaints. *Journal of the Academy of Marketing Science, 47*(5), 858–878, https://doi.org/10.1007/s11747-019-00661-x.

Keeling, K., McGoldrick, P., & Beatty, S. (2010). Avatars as salespeople: communication style, trust, and intentions. *Journal of Business Research, 63*(8), 793–800, https://doi.org/10.1016/j.jbusres.2008.12.015.

Kelleher, T., & Miller, B. M. (2006). Organizational blogs and the human voice: relational strategies and relational outcomes. *Journal of Computer-Mediated Communication, 11*(2), 395–414, https://doi.org/10.1111/j.1083-6101.2006.00019.x.

Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences, 111*(24), 8788–8790, https://doi.org/10.1073/pnas.1320040111.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, eds. *Advances in Neural Information Processing Systems, 60*(6), 84–90, https://doi.org/10.1145/3065386.

Lee, C., & Chau, D. (2018). Language as pride, love, and hate: archiving emotions through multilingual Instagram hashtags. *Discourse, Context & Media, 22*(April), 21–29, https://doi.org/10.1016/j.dcm.2017.06.002.

Lee, Y. E., & Benbasat, I. (2003). Interface design for mobile commerce. *Communications of the ACM, 46*(12), 48–52, https://doi.org/10.1145/953460.953487.

Lipton, Z. C., Berkowitz, J., & Elkan, C. (2015). A critical review of recurrent neural networks for sequence learning. *ArXiv:1506.00019 [Cs], October* http://arxiv.org/abs/1506.00019.

McCorkindale, T. (2010). Can you see the writing on my wall? A content analysis of the fortune 50's Facebook social networking sites. *Public Relations Journal, 4*(3), 1–13.

Miller, J. (1991). Short report: reaction time analysis with outlier exclusion: bias varies with sample size. *The Quarterly Journal of Experimental Psychology Section A, 43*(4), 907–912, https://doi.org/10.1080/14640749108400962.

Morris, S. (1988). How many lost customers have you won Back today? An aggressive approach to complaint handling in the hotel industry. *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior, 1*, 86–92.

Nguyen, H., Groth, M., Walsh, G., & Hennig-Thurau, T. (2014). The impact of service scripts on customer citizenship behavior and the moderating role of employee customer orientation. *Psychology & Marketing, 31*(12), 1096–1109.

Niu, R. H., & Fan, Y. (2016). To tweet or not to tweet? Exploring the effectiveness of service recovery strategies using social media. *International Journal of Operations & Production Management, 36*(9), 1014–1036, https://doi.org/10.1108/IJOPM-10-2013-0461.

Orthaber, S., & Márquez-Reiter, R. (2011). 'Talk to the hand'. Complaints to a public transport company. *Journal of Pragmatics, 43*(15), 3860–3876, https://doi.org/10.1016/j.pragma.2011.10.004.

Pennington, J., Socher, R., & Manning, C. (2014). Glove: global vectors for word representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, 1532–1543, https://doi.org/10.3115/v1/D14-1162.

Searle, J. R. (1999). *Mind, language and society: Philosophy in the real word*. London, England: Weidenfeld and Nicolson.

Searle, J. R., & Vanderveken, D. (1985). *Foundations of illocutionary logic*. Cambridge, England: Cambridge University Press.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research, 15*(1), 1929–1958.

Steiner, E. (2019). *A functional perspective on language, action, and interpretation: An initial approach with a view to computational modeling*. Walter de Gruyter GmbH & Co KG.

Su, S. Y. W., Huang, C., Hammer, J., Huang, Y., Li, H., & Wang, L., et al (2001). An internet-based negotiation server for e-commerce. *The VLDB Journal, 10*(1), 72–90.

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017). Inception-v4, inception-ResNet and the impact of residual connections on learning. *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, 4278–4284 http://arxiv.org/abs/1602.07261.

Tamburrini, N., Cinnirella, M., Jansen, A., AA., & Bryden, J., et al (2015). Twitter users change word usage according to conversation-partner social identity. *Social Networks, 40*, 84–89.

van Noort, G., Voorveld, H. A. M., & van Reijmersdal, E. A. (2012). Interactivity in brand web sites: Cognitive, affective, and behavioral responses explained by consumers' online flow experience. *Journal of Interactive Marketing, 26*(4), 223–234, https://doi.org/10.1016/j.intmar.2011.11.002.

Vieira, R., Moreira, Á. F., Wooldridge, M., & Bordini, R. H. (2007). On the formal semantics of speech-act based communication in an agent-oriented programming language. *Journal of Artificial Intelligence Research, 29*, 221–267.

Walsh, G., Gouthier, M., Gremler, D. D., & Brach, S. (2012). What the eye does not see, the mind cannot reject: Can call center location explain differences in customer evaluations? *International Business Review, 21*(5), 957–967, https://doi.org/10.1016/j.ibusrev.2011.11.002.

Woo, C. C., & Chang, M. K. (1992). An approach to facilitate the automation of semistructured and recurring negotiations in organizations. *Journal of Organizational Computing and Electronic Commerce, 2*(1), 47–76.

Yang, S.-U., Kang, M., & Johnson, P. (2010). Effects of narratives, openness to dialogic communication, and credibility on engagement in crisis communication through organizational blogs. *Communication Research, 37*(4), 473–497, https://doi.org/10.1177/0093650210362682.