Conversational Dynamics:
Discovering *When* Employee Language Matters

Working Paper

May 2021

ABSTRACT

Text analysis is increasingly used for marketing insight. But while work has shed light on *what* firms should say to customers, *when* to say those things in a given conversation is less clear. Customer service agents, for example, could adopt a certain speaking style early in a conversation, at the end, or throughout. *When* might specific language features be most beneficial? This paper introduces a method to address this question. To demonstrate its value, we apply it to two key dimensions of service language: warmth and competence. Prior research suggests an affective (i.e., warm) approach leads employees to seem less competent, so a cognitive (i.e., competent) style should be used. In contrast, analysis of hundreds of customer service calls across two firms suggests that service outcomes are better when *both* affective and cognitive language are used, but at separate, specific times. An experiment underscores the observed pattern and directly demonstrates causality. We also briefly show how the method can shed light on other noteworthy language features. Taken together, this work updates managerial guidance on, and deepens conceptual understanding of, the warmth-competence tradeoff and offers a means to support new explorations of conversational dynamics in marketing and beyond.

Keywords: Language, Dynamics, Customer Service, Functional Regression, Group-Lasso.

Language plays a central role in marketing. Advertising copy shapes attitudes, sales language impacts purchase, and service language drives satisfaction and retention (cf. Pogacar, Shrum, and Lowrey 2018). A great deal of work has considered how service employees should speak to customers (e.g., Parasuraman, Zeithaml, and Berry 1985; Blanding 1989), for example, and advances in text analysis have shed light on what specific words and speaking styles matter in a variety of domains (Berger et al. 2020).

But while it's clear that *what* companies and employees say matters, might *when* they say it also play an important role?

When calling customer service, for example, or speaking with a salesperson, the interaction usually involves multiple conversational turns. The customer says something, the employee responds, and the two go back and forth repeatedly. Research suggests that employees should ask questions (Drollinger, Comer, and Warrington 2006), use first person pronouns (Packard, Moore, and McFerran 2018), or speak in a rational, task-oriented way (e.g., Singh et al. 2018; van Dolen, Dabholar, and de Ruyter 2007), but they could do so at any point in an interaction. Should employees do these things throughout a conversation? Or might doing so at certain points be more beneficial? And might doing them at other points actually have negative effects?

This paper introduces a method to move beyond asking *whether* certain language features matter, to asking *when*. Conversations often involve dramatic moment-to-moment variations in content (Zhang, Wang, and Chen 2020), making make them remarkably difficult to analyze. To address these challenges, we combine functional data analysis (FDA; e.g., Foutz and Jank 2010) with machine learning. This allows us to recover time-based trajectories documenting the relationship between language and important marketing outcomes.

To demonstrate the approach, and its potential, we apply it to the two most important dimensions of person perception — warmth and competence (Fiske, Cuddy, and Glick 2007). These dimensions are usually seen in opposition: Trying to be warmer, or more affective, makes people seem less competent, and acting more rationally or task-oriented makes people seem less emotionally engaged (Godfrey, Jones, and Lord 1986; Holoien and Fiske 2013; Wang et al. 2017). Consequently, research suggests doing one *or* the other. In customer service, for example, research suggests agents should speak or behave competently rather than warmly (e.g., Kirmani et al. 2017; Li, Chan, and Kim 2019; Marinova, Singh, and Singh 2018; van Dolen et al. 2007). Accordingly, firms prioritize competence in this setting (Jasmand, Blazevic, and de Ruyter 2012).

In contrast, our approach suggests *both* warmth and competence may be valuable, but at different points in an interaction. Consistent with this notion, dynamic modeling of conversational language from two different firms, as well as an experiment, reveal that customer satisfaction and purchase are higher when employees speak more affectively at certain points in the conversation, and more cognitively in others. Further, our approach reveals that using the same styles at the wrong conversational moments can be detrimental.

The paper makes four main contributions. First, by moving beyond *whether* a communication or language style matters to showing *when* it matters, this method can help firms improve customer satisfaction and purchase behavior. Managers can train employees on the importance of rapport building at the start of frontline interactions before turning to competently addressing the customer's needs, as well as the value of returning to a warm style at the interaction's close. Our approach can be applied to enhance analysis of frontline sales and service transcripts, improving employee performance assessment and development. AI and chatbot developers can use the method to fine-tune machine agents' ability to handle complex multiple-turn interactions.[1]

Second, we provide deeper insight into the so-called warmth/competence trade-off reported by both marketing scholars and psychologists. While a great deal of research suggests that speakers are best off using only one of an affective *or* cognitive approach, our dynamic approach reveals that the warmth/competence "trade off" may not be so stark. Rather than pursuing just one dimension, communicators may benefit from prioritizing speaking affectively *and* cognitively at different, specific times within an interaction.

Third, our approach helps address modeling challenges in understanding communication dynamics in marketing (Grewal et al. 2021). Conversations include time-varying interactional and circumstantial features that affect both conversational content and outcomes (Zhang, Mullainathan, and Danescu-Niculescu-Mizil 2020). Our approach helps address interactional challenges by simultaneously accounting for various language and paralanguage (vocalization) dynamics for both conversation partners, as well as potential mimicry. To address the circumstantial challenge, we apply a machine learning method to account for features of conversation partners and their interaction that may drive conversational content and/or outcomes. The method also helps tackle the high dimensionality, irregularity, and sparsity inherent in conversational data. While it is nearly impossible to eliminate all potential sources of endogeneity in conversational dynamics, our approach accounts for the most plausible sources, enhancing inference making.

Fourth, scholars who examine language, paralanguage, and other time-varying interaction features (e.g., non-verbals such as posture or physical behaviors) can apply the method to expand

---

[1] R syntax for the approach is shared at [URL blinded for review] to facilitate such opportunities.

their empirical toolkit. The approach can help marketing and consumer researchers move beyond traditionally static, interaction-level analyses to better understand important situated and temporal factors in human interactions, enhancing the conceptual and substantive insights they offer. We hope the approach provides a useful framework for the growing use of text analysis for conceptual or substantive insights in marketing and beyond.

TALKING TO CUSTOMERS

Talking to customers is important. American companies spend over a trillion dollars a year on staffing, training, and supporting frontline sales and service. This is the single largest strategic investment for most firms, and nearly three times what they spend on marketing communications (Cespedes and Wallace 2017; Morgan 2017). Further, these costs are likely to rise, as channel complexity and technology make it harder than ever to deliver great service (McBain 2020).

Consistent with its importance, researchers have spent a great deal of time and effort trying to understand and improve frontline interactions. Thousands of articles have studied service quality (see Ladhari 2008; Parasuraman and Zeithaml 2002; Snyder et al. 2016 for reviews), examining how consumers evaluate salespeople (e.g., Zeithaml, Berry, and Parasuraman 1996), service initiatives shape customer attitudes (e.g., Bolton and Drew 1991), and service quality impacts financial outcomes (Rust and Chung 2006).

Recent work has explored the role of language in marketing and service outcomes. Ordenes and colleagues (2014), for example, find that language's topical content (e.g., firm vs. product) enhances sentiment analysis. Other work finds that replying in complete sentences (Castleberry et al. 1999), using more concrete language (Packard and Berger 2020), and applying more first person singular pronouns ("I" pronouns; Packard, Moore, and McFerran 2018) can all improve customer satisfaction.

But while these examples demonstrate language's importance, they all focus on *what* rather than *when*. Should service agents use these types of language throughout a conversation, for example, or might they be more beneficial at certain conversational moments than others? And might using them at the wrong moments actually backfire?

THE CURRENT RESEARCH: WHEN LANGUAGE MATTERS

To illustrate the value of understanding *when*, we examine the so-called "warmth/competence trade-off." Warmth and competence are two universal dimensions of social cognition, accounting for almost all person perception (Fiske et al. 2007). Warmth captures affective expression and attention to emotions while competence focuses on agency, rationality and cognitive efficiency (Abele and Wojciskzke 2007). Above all else, people evaluate one another on these two fundamental dimensions (Judd et al. 2005).

Importantly, however, a great deal of research suggests these two dimensions are inversely related. Being affectively-engaged reduces perceived competence, while acting rational and cognitively-oriented makes people seem less warm. This trade-off has led many to suggest that people should try to be warm or competent, but not both (Godfrey et al. 1986; Holoien and Fiske 2013; Wang et al. 2017).

Marketing research supports the cognitive side of the tradeoff, suggesting that companies prioritize a more competence-oriented approach (Kirmani et al. 2017). Work on customer queries, for example, finds that cognitive or competent language and behaviors are beneficial while warm language and non-verbals have null or negative effects on customer attitudes

(Marinova et al. 2018; Singh et al. 2018). Similarly, solutions oriented service advisors reportedly improve customer satisfaction more than socially oriented agents (van Dolen et al. 2007) and service employees who use emoticons are perceived as warmer, but less competent (Li, Chan, and Kim 2019), leaving customers feeling less satisfied.

But should service agents always prioritize a rational, cognitive manner of speaking? And how does this fit with other work encouraging employees to speak affectively to show the customer they care (e.g., de Ruyter and Wetzels 2000; Parasuraman et al. 1985; Spiro and Weitz 1990)? More generally, might there be a way to incorporate both aspects effectively?

Rather than speaking either affectively *or* cognitively, we suggest considering *when* within customer interactions each is beneficial. In customer service calls, for example, rather than diving straight into finding a solution, affective language may initially be beneficial. Indeed, social norms suggest some relationship-building before turning to the speakers' specific goals or task may be beneficial (Gabor 2011; Kaski, Niemi, and Pullins 2018; Placencia 2004). When employees and customers interact for the first time, affective language may be particularly helpful at building situated rapport (DeWitt and Brady 2003; Gremler and Gwinner 2000).

But while starting with affective language may be beneficial, it should only go so far. Eventually employees must address the customer's needs (Marinova et al. 2018; Singh et al. 2018). Consequently, competence should be important, and shifting to a more analytic, cognitive communication style may be valuable.

Finally, given the work on recency and end effects (Greene 1986), closing with affective language may help end things on a positive note. Wrapping up an interaction in a manner that seems considerate or empathetic is a key feature of successful conversations (Schegloff and Sacks 1973). Summarizing what happened in a positive and polite way may signal that a conversation is approaching its close without seeming overly cold or abrupt (Bardovi-Harlig et al. 1991).

To test these predictions, we present a novel modeling approach. Analyzing dynamic linguistic and paralinguistic features over conversational time allows us to examine *when* more affective or cognitive employee language may have positive or negative effects. We test this approach with an initial dataset, use a second dataset to check robustness, and conduct a simple experiment to further assess causality.

STUDY 1: MAIN FIELD DATA

We collected recordings of 200 customer service calls from a large US online retailer of apparel products. A professional transcription company converted the recordings to text, separating each conversational turn (e.g., turn 1 (agent): “How can I help you?”, turn 2 (customer): “I can’t find …”). Part of the conversation was inaudible for fifteen of the 200 recordings provided, leaving 12,410 turns from 185 conversations for analysis. The average conversation lasted 6.19 minutes (SD = 3.97) and included 66.75 turns (SD = 44.49).

Independent Measures: Agent Affective and Cognitive Language

Following prior work examining warmth and competence (Decter-Frain and Frimer 2016; Berry et al. 1997; Marinova et al. 2018; Singh et al. 2018), we measure affective and cognitive language through Linguistic Inquiry and Word Count (LIWC; Pennebaker et al. 2015).[2,3] As

---

[2] Prior literature uses a variety of terms interchangeably for affective and cognitive components of interpersonal language, behavior, and perception (e.g., warmth/competence, communion/agency, relating/resolving).
[3] The Marinova et al. (2018) and Singh et al. (2018) papers customize the LIWC dictionaries to a smaller set of words and provide new names for their linguistic features of warmth (“relating”) and competence (“resolving”). We use this custom adaptation of the LIWC dictionaries in a robustness check, which produces similar results.

noted previously, warmth is conveyed through emotional expression. Using affective words like *happy* (e.g., “I’m happy you like the pants”), *great* (“That’s great”), or *horrible* (“That’s horrible”) signals that an agent is attending to a customer’s emotional state or expressing their own. Following prior work, affective language is measured through LIWC’s affective processes module, which contains 1,388 words and word stems[4] related to emotional expression (e.g., happy, great, horrible).

Cognitive language involves rational expression suggesting instrumentality, intelligence, and agency. Using cognitive words like *diagnose* (e.g., “Let’s diagnose the cause”) or *think* (“I think that will do it”) signals that an agent is cognitively working to address the customer’s needs. Following prior work, cognitive language style is measured through LIWC’s cognitive processes module, which contains 780 words and word stems related to this construct (e.g., diagnose, think, and solve).

Dependent Measures: Customer Satisfaction and Purchase

We examine the relationship between agent conversational dynamics and two closely related customer outcomes. First, perceived helpfulness represents a crucial performance-based measure of customer satisfaction (Cronin and Taylor 1992; Parasuraman et al. 1991), so we collected the firm’s measure of this for each call (1 = not at all helpful, 4 = very helpful, measured at the end of the call). Second, we also collected a behavioral measure, the number of orders in the 30 days following the call.

[4] Word stems capture tense and part of speech variations of a single root. For example, the stem “bother*” captures bother, bothers, bothered, and bothering.

Control Variables

Many other factors may be associated with both our dynamic predictors and outcome measures, so to minimize endogeneity we control for a range of call, agent, customer, and interaction variables at both the call (static) and turn (dynamic) levels.

Call Content. The call's content could impact the agent's language and customer satisfaction, so we control for it in two ways. First, we include dummy variables for the four call reasons captured by the firm (*Order*, *Shipping*, *Return*, *Product*). Second, to provide a more fine-grained measure, we use the customer's language to uncover the hidden mixture of call topics via latent Dirichlet allocation (Blei, Ng, and Jordan 2003). Assessment by perplexity and interpretability supports 13 topical controls, each of which characterizes the proportion of the call's language corresponding to that topic (*Topic 1*, …, *Topic 13*).

Complexity. The complexity of the call could shape agent's language, and their ability to successfully solve the issue, so we control for it in two ways. First, we take the average of two judges who listened to each call and indicated perceived difficulty or severity of the call on a five-point scale ($r = .72$; *Severity*). Second, given that complex issues may require more discussion, we control for call length using the total number of words spoken (*Length*).

Resolution. Whether the agent was able to resolve the customer's issue during the call likely impacts how both the agent and customer speak, as well as customer satisfaction and purchase. To account for this, two judges read each call transcript and indicated whether the customer's main issue had been resolved. Judge disagreements were settled via discussion (*Resolved*).

Agent Observables. The employee's experience could shape both how they speak to customers and conversation outcomes, so we control for agent characteristics in two ways. First, to take organizational experience into account, we include how many days agents have been with the firm (*Agent Tenure*). Second, to account for direct experience with customers, we consider the total number of calls they have handled (*Agent Calls*), which is only moderately correlated with job tenure ($r = .38$, $p < .05$). The firm also provided agent gender, which we include as a dummy variable (*Agent Female*).

Customer Observables. The consumer's experience with a firm can affect customer satisfaction and behavior (e.g., loyalty effects; Neiderhoffer and Pennebaker 2002), so we control for customer characteristics in two ways. First, we use the number of days since the customer's first purchase with the firm (*Customer Tenure*). Second, we include their lifetime expenditure with the firm in dollars (*Customer LTV*). We also incorporate two demographics variables provided by the firm, including dummies for which of five geographic regions a customer resides in (*Customer Region*), and a dummy for customer gender (*Customer Female*).

The customer's attitude about other aspects of the firm could impact how they interact with the agent, and their satisfaction. To control for this possibility, we include measures of their attitude towards the website (*Attitude Web*) and shopping experience (*Attitude Shop*), which were captured by the firm after the customer satisfaction measure at the end of the call.

Having considered a range of static, interaction-level features, we then account for dynamic conversational features.

Dynamics of Other Major Agent Language Features. Beyond affective and cognitive language, other dynamic features of employee language may shape how customers perceive or speak to them. To control for this, we include turn-level measurement of LIWC's other main psychological process dictionaries (e.g., *Social processes, Perceptual processes, Drives, Temporal orientation, and Informality*; Pennebaker et al. 2015).

Dynamics of Agent Paralanguage. In addition to what was said, one could wonder whether how things were said (i.e., paralanguage) might drive the effects. The extent to which a speaker modulates pitch and intensity (volume) while talking has been linked to social perception and persuasion (Van Zant and Berger 2020). We control for these dynamic paralinguistic features using phonetics software at turn level (*Pitch* and *Intensity*; Boersma and van Heuven 2001).

Moment-to-Moment Synchronicity. To isolate the dynamic impact of agent's language, we also control for how it may be shaped by customer language. How someone speaks can impact their conversation partner, but also may reflect things that the conversation partner said previously (Zhang et al. 2020). Agents may use more affective language to respond to customers who are already speaking emotionally, and customers may adopt agents' language when discussing technical or detailed steps that need to be taken to competently (i.e. cognitively) solve an issue. To control for these possibilities, we use a moment-to-moment measure of agent-customer linguistic synchronicity (*Synchronicity)*. Specifically, following Zhang, Wang, and Chen (2020) we create a synchronicity measure using the R^2 of the moment-to-moment regression from customer language on agent language. Figure A1 in the Web Appendix summarizes the synchronicity observed.

Dynamics of Customer Affective and Cognitive Language. In addition to moment-to-moment synchronicity, an agent might mimic or otherwise repeat something the customer said much earlier in the conversation. To account for things not captured by moment-to-moment mimicry, we include the customer's own affective and cognitive language over the course of the conversation as dynamic controls.

Dynamics of Other Major Customer Language Features. Beyond affective and cognitive language, other aspects of customer language may shape how employees respond later in the conversation, so we control for these using turn level measurement of the same psychological process dictionaries applied to employee language (i.e., *Social, Perceptual, Drives, Time, and Informal*).

Overall, in addition to the two dynamic language predictors (agent affective and cognitive language) related to the two most important dimensions of person perception, our model incorporates 34 static and 18 dynamic language and paralanguage controls, each observed during the course of over 12,000 conversational turns observed. See Web Appendix Table A1 for summary statistics for all the independent, dependent and control variables.

While it is difficult to completely rule out endogeneity in conversational data, controlling for such an extensive variety of factors helps mitigate such concerns. Further, the temporal relationship between the predictors and outcome measures casts doubt on reverse causality.

Modeling Approach

To flexibly characterize the relationship between dynamic conversational features (e.g., affective and cognitive language) and static conversational outcomes (i.e., customer satisfaction or purchase behavior), we begin our modeling efforts with semiparametric tools from functional data analysis (FDA; Ramsay and Silverman 1997). Functional data has seen growing applications in marketing to help address dynamic modeling challenges. Sood and colleagues (2009) use functional regression to forecast new product penetration, for example, demonstrating FDA's superiority over the Bass model in predicting diffusion. Similarly, functional analysis has been used to predict pre-release demand of motion pictures (Foutz and Jank 2010), relate moment-to-moment consumer attitudes to overall judgements of TV shows (Hui et al. 2014), and explore how temporal variations in online chatter volume is linked to new product performance (Xiong and Bharadwaj 2014).

We extend FDA to conversations. We consider time-varying measurement of a conversation feature (e.g., affective or cognitive language) within the n-th conversation as a trajectory $X_n(t)$, $n = 1, \dots, N$, that is randomly drawn from an underlying stochastic function. The following functional regression relates the static outcome of the interaction y_n to the dynamic language measurement $X_n(t)$,

$$y_n = \alpha + \int_0^1 \beta(t)[X_n(t) - \mu(t)]dt + e_n \tag{1}$$

where α is the intercept, $\mu(t) = \mathbb{E}[X_n(t)]$ the mean function of $X_n(t)$, e_n the i.i.d. Gaussian error term, and $\beta(t)$ the sensitivity curve of interest that characterizes the dynamic impact of a linguistic feature at different moments during a conversation. To enable functional regression's requirement that the units of analysis have the same total duration, we standardize the varied conversation lengths to a common interval $[0,1]$ (Ramsey and Silverman 2005). Therefore, any conclusions should be viewed against the relative progress of a conversation rather than absolute time passed. To account for the potential impact on model estimates due to standardization, we include conversational length in seconds and word count as controls in the main model.

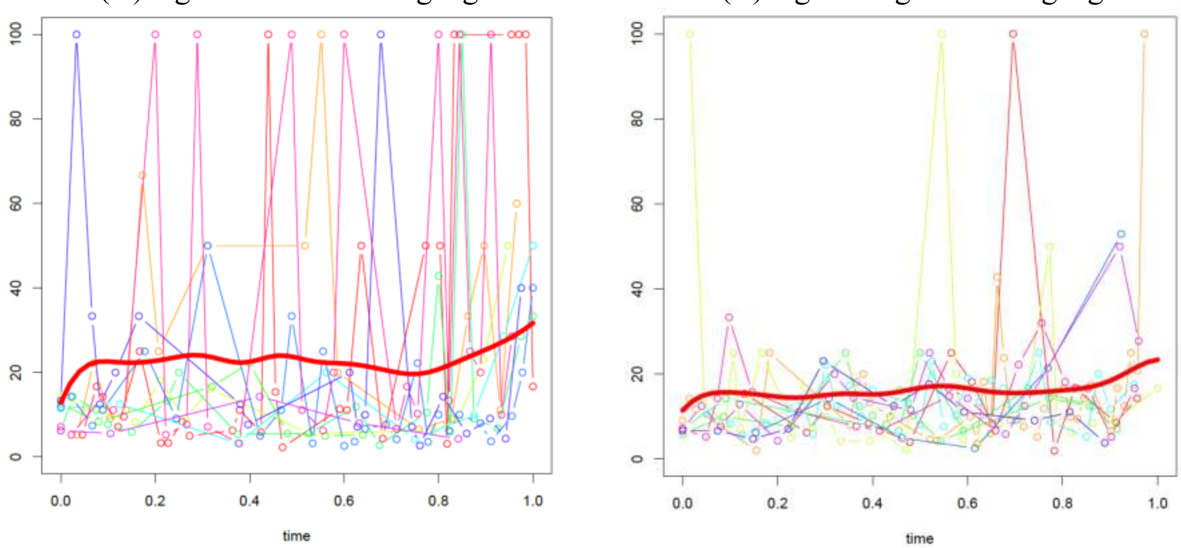
Sparseness and Irregularity in Conversational Dynamics. Before applying the functional regression model, however, major challenges specific to conversational data (i.e., irregularity, sparsity, and high dimensionality) need to be addressed. While virtual stock markets (Foutz and Jank 2010) and continuous user dials (Hui et al. 2014) provide evenly-spaced and dense measurements, conversational language occurs over a series of spontaneous conversational turns and tend to be *irregularly spaced* across time. Some turns ("Hi, my name is Chris, thanks for calling customer service."), for example, are longer than others ("My phone is broken.").

Further, given the use of fixed dictionaries to measure language features, a certain conversational feature may not appear every moment, resulting in *sparse* measurement of the feature.

Figure 1 demonstrates the irregularity in our focal language features. The bold lines in Figure 1 illustrate agents' average use of affective and cognitive language over the course of a conversation. The y-axis indicates turn-level LIWC measurement of a language feature (i.e., the percentage of words in a turn belonging to a LIWC category). The figure also depicts language from a random sample of 10 calls, which indicate the irregularity in language feature use.

Figure 1: Means and Samples of Linguistic Features over Conversational Time

(A) Agent Affective Language (B) Agent Cognitive Language



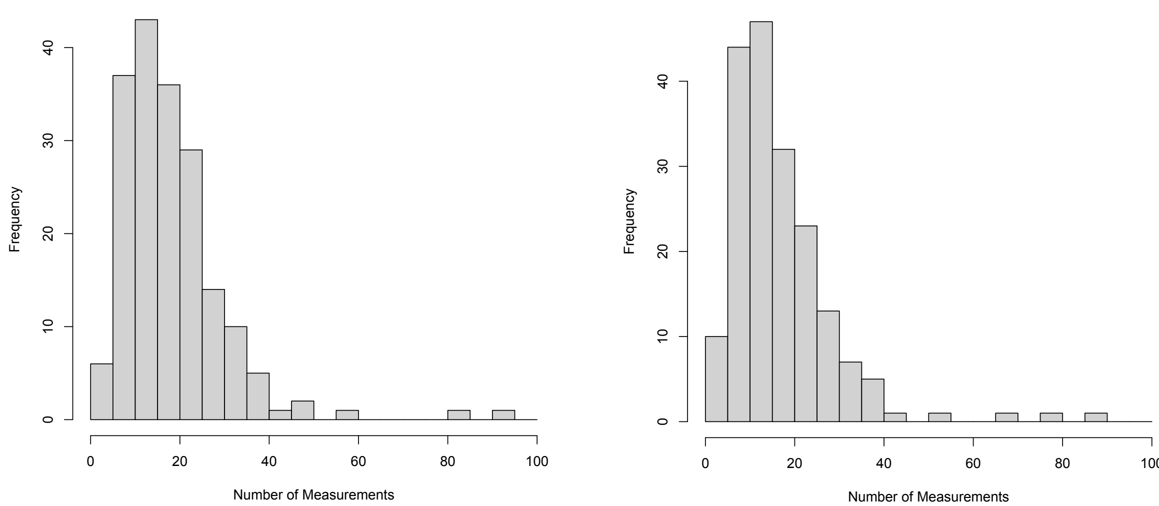
Notes: Y axis = proportion of words; Red line = mean over full data set.

Figure 2 illustrates the sparseness. Except for a handful of calls that contain close to 100 measures of these language features, most interactions have only 10 to 30 measurements. Consequently, functional regression for conversation must be able to handle the irregular and sparse measurement of conversational features.

In addition to irregularity and sparsity, human conversation is also complex, containing a wide variety of dynamic linguistic and paralinguistic features, as well as static observables. To control for their influence, we need to deal with a "wide" data situation in which the number of (both functional and scalar) variables may be comparable to or even greater than the number of observations (conversations). As noted, compared with the 185 call observations in the data, there are two focal language variables (agent affective and cognitive language), 18 dynamic language controls, as well as 34 static controls. The dynamic linguistic features alone translate to close to 100 regressors after the functional Karhunen-Loève expansion.

Figure 2: Sparseness in Linguistic Measurements of Conversation

(A) Agent Affective Language (B) Agent Cognitive Language



Moreover, as dependent variables may be recorded as nonlinear responses such as count data (i.e., purchase quantity post call), we need to employ an appropriate link function to generalize the functional linear regression specified in (1).

To address these challenges, we use recent developments in statistics and machine learning to extend the conventional functional regression model. In particular, we consider a dynamic language feature as a continuous trajectory $Z_n(t)$ over the course of conversation n. Across multiple conversations, we obtain a sample of measured trajectories that are assumed to be independently drawn from an underlying stochastic function, with unknown mean function $\mu(t) = \mathbb{E}[Z_n(t)]$ and variance function $\Sigma(t_1, t_2) = \mathrm{Cov}[Z_n(t_1), Z_n(t_2)]$. Due to measurement errors arising from using language dictionaries, the actual observation for the m-th measurement, $m = 1, \ldots, M_n$, of the n-th conversation is given by

$$X_n(t_m) = Z_n(t_m) + \varepsilon_n(t_m) \tag{2}$$

where t_m indicates the time of the sequential conversational turn at which the measurement was taken, and the measurement error ε_n is i.i.d. drawn from $N(0, \sigma^2)$. In call n, the M_n measurements are irregularly-spaced and sparse. We assume M_n is exogenous and control for its effect in our model.

For each functional variable, we apply scatterplot smoothing and surface smoothing, both via local linear regression, to estimate the mean and covariance functions respectively (Yao et al. 2005; Wang et al. 2016; Chen et al. 2017).[5] We use the entire sample simultaneously in the smoothing procedure to allow information shrinkage across observations to accommodate the data sparseness discussed above.

After smoothing, we apply Karhunen-Loève expansion to obtain eigen components of the conversations, $\{X_n(t)\}_{n=1}^N$, namely,

$$\Sigma(t_1, t_2) = \sum_{i=1}^{\infty} \lambda_i \phi_i(t_1) \phi_i(t_2) \tag{3}$$

---

[5] For both the smoothed mean and covariance functions, we apply the commonly-used Gaussian kernel and obtain the smoothing bandwidth via the generalized cross-validation bandwidth selection (Speckman 1988).

and so

$$X_n(t) = \mu(t) + \sum\nolimits_{i=1}^{\infty} \omega_{ni}\phi_i(t) + \varepsilon_n(t) \tag{4}$$

where $\phi_i(t)$ is the i-th eigen function, λ_i the associated eigen value, and ω_{ni} the i-th eigen score of the n-th conversation. If we expand the unknown $\beta(t)$ curve onto the same eigen bases,

$$\beta(t) = \sum\nolimits_{i=1}^{\infty} b_i\phi_i(t) \tag{5}$$

thanks to orthogonality, the functional regression in (1) can now be simplified to

$$y_n = \alpha + \sum_{i=1}^{\infty} b_i\omega_{ni} \approx \alpha + \sum_{i=1}^{I} b_i\omega_{ni} \tag{6}$$

In the above, the truncation I, or the actual number of eigen components to appear in the regression, is determined using AIC. We also tested other metrics such as BIC and leave-one-out cross-validation, and ended up with almost identical truncation points.

High Dimensionality in Conversational Dynamics. From the data we obtain a number of dynamic and static features that are possibly interdependent. Therefore, we write the following generalized functional regression to accommodate additional functional and scalar variables with nonlinear responses,

$$E[y_n|\{X_{ln}\}_{l=1}^{L}, \{W_{jn}\}_{j=1}^{J}] = g^{-1}(\alpha_a + \sum\nolimits_{l=1}^{L} \int_0^1 \beta_l(t)[X_{ln}(t) - \mu_l(t)]dt + \sum\nolimits_{j=1}^{J} \gamma_j W_{jn}) \tag{7}$$

where L and J denote the number of functional and scalar predictors respectively, W_{jn} is the j-th scalar control for the n-th call, γ_j represents the regression coefficients, and $g(\cdot)$ indicates the link function for nonlinear dependent variable. Besides using agent observables as controls, we further capture agent heterogeneity with a random intercept α_a for every agent.[6]

---

[6] We did not impose random coefficients on functional variables because of the limited number of call observations.

Applying the smoothing procedure and Karhunen-Loève expansion to the data, we obtain a simplified generalized regression as follows,

$$E[y_n|\{X_{ln}\}_{l=1}^{L},\{W_{jn}\}_{j=1}^{J}] = g^{-1}(\alpha_a + \sum_{l=1}^{L}\sum_{i=1}^{I_l} b_{li}\omega_{lni} + \sum_{j=1}^{J}\gamma_j W_{jn}) \tag{8}$$

where I_l for function variable $X_l(t)$ is determined by the truncation criterion discussed above.

As the total number of variables $(L + J)$ becomes comparable to the number of observations the model is likely to overfit, resulting in less meaningful results. To address this possibility, the regression needs to be regularized such that the dynamic and static controls can be automatically selected to yield efficient model inference.

However, conventional variable selection methods such as stepwise regression (e.g., Foutz and Jank 2010) are not appropriate in our context for two reasons. First, solutions from stepwise regression are path-dependent as the approach is a *greedy* algorithm that finds local optima in every step, but often fails to reach generally optimal variable selection. This limitation of stepwise regression is commonly described as the lack of oracle properties in variable selection (Zou 2006). Second, stepwise regression does not allow *group-wise* variable selection, whereas the selection of functional variables corresponds to selecting from the L groups of eigen scores in (8). That is, for a given functional variable $X_l(t)$, either $\{b_{li}\}_{i=1}^{I_l}$ are all suppressed to zero or are all selected to enter the regression. Similarly, categorical control variables associated with multiple dummies (e.g., call reasons, customer regions) also require group-wise selection.

To overcome these challenges, we utilize Group-Lasso regularization (Yuan and Lin 2006; Meier et al. 2008; Yang and Zou 2015) to avoid path-dependency and to retain the functional and categorical variable grouping after selection. The shrinkage and variable selection method, Lasso (Tibshirani 1996), has been widely applied in statistics and machine learning for high dimensional data analysis. Yuan and Lin (2006) proposed a generalization of Lasso for group-wise variable selection and regularization. To answer our central research questions around the dynamics of the well-established importance of affective and/or cognitive language in customer service and more broadly (e.g., Holoien and Fiske 2013; Kirmani et al. 2017; Marinova et al. 2018; Wang et al. 2017), we keep the two functional predictors unpenalized in the L1 regularization procedure (Chen et al. 2016). Assuming the controls in our model can be divided into D non-overlapping groups, where D is determined by the number of controls and the truncation of eigen components for each functional variable, Group-Lasso attempts to minimize

$$\frac{1}{2}\left\| g(E[y]) - \alpha_a - \vec{b}_A\vec{\omega}_A - \vec{b}_C\vec{\omega}_C - \sum_{d=1}^{D} \vec{b}_d\vec{\omega}_d \right\|_2^2 + \lambda \sum_{d=1}^{D} \sqrt{\dim(\vec{b}_d)}\left\|\vec{b}_d\right\|_2 \tag{9}$$

where subscripts "A" and "C" denote the affective and cognitive language components respectively. The Group-Lasso procedure suppresses a subset of groups of coefficients to zero to encourage a simpler and more efficient generalized linear model. Solving the above penalized least squares is computationally expensive, so we follow Yang and Zou (2015) and implement the groupwise-majorization-descent (GMD) algorithm to achieve fast computation of Group-Lasso for the simultaneous selection of functional and scalar variables. To determine the optimal value of penalty parameter λ, we first calculate the maximum penalty parameter λ_{max} such that none of the penalized groups are active in the model. Then we construct a multiplicatively decaying grid for possible λ values starting at λ_{max}, and use leave-one-out cross-validation to pick the best penalty parameter from the grid.

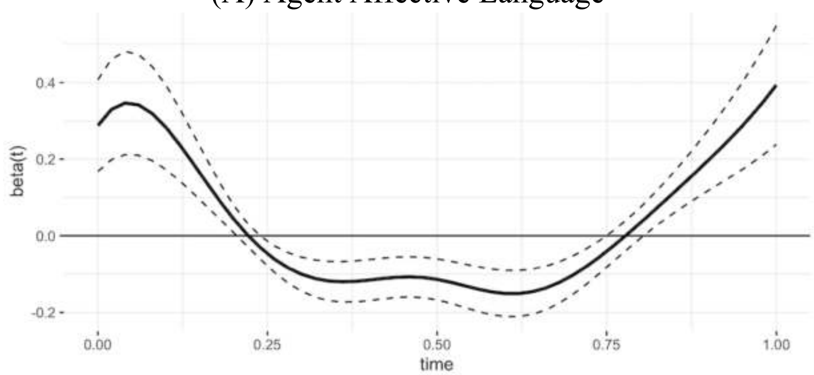
Key results are represented by the $\beta_l(t)$ curves estimated from the sparse functional regression in (8). Predictors have a positive (negative) relationship with the outcome of interest when a given $\beta_l(t)$ curve and its confidence interval lie above (below) zero. We examine the relationship between agent affective and cognitive language and both customer satisfaction and purchase.

Main Results

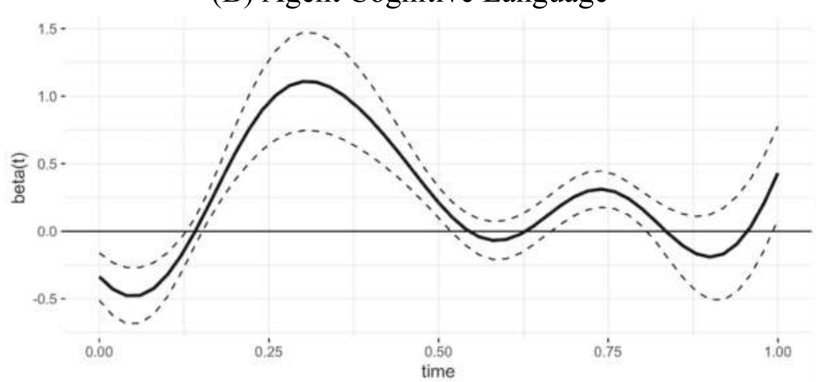
Customer Satisfaction. As predicted, customers are more satisfied (Figure 3A, pointwise 95% confidence interval above zero) when agents use more affective language at the conversation's beginning and end. In contrast, customers are more satisfied when agents avoid affective language during the middle of the call (pointwise 95% confidence interval below zero).

Figure 3: Beta Curves for Agent Affective (A) and Cognitive (B) Language
in Relation to Customer Satisfaction

(A) Agent Affective Language



(B) Agent Cognitive Language



Dotted lines: pointwise 95% confidence intervals

The results for cognitive language are quite different (Figure 3B). While customer satisfaction is higher when affective language is used at the call's beginning, speaking more rationally during this time appears to be costly. Instead, customers are more satisfied when agents use a cognitive language style in the middle of the conversation. Excluding control variables shows similar results (Web Appendix Figure A2). See Web Appendix Table A2 for parameter estimates of the model for customer satisfaction with a linear link function.

Taken together, these findings suggest that affective and cognitive language can both be beneficial, but at different times. Note also that customer service agents do not seem to follow the beta curves revealed (Figure 1A), casting doubt on the notion that these patterns are somehow already known and practiced.

Purchase Behavior. While the customer satisfaction results support our conceptualization, one might wonder whether they extend to subsequent purchase behavior.

To test this possibility, we apply a functional Poisson regression with a Log link function in (8) to estimate the relationship between agent affective and cognitive language and downstream purchase behavior (i.e., order count). The Poisson model has the same sets of functional and scalar variables as in the functional linear regression, and further includes a control for each customer's baseline buying behavior using the number of orders they placed up to 30 days prior to the conversation (*Orders 30 Pre*).

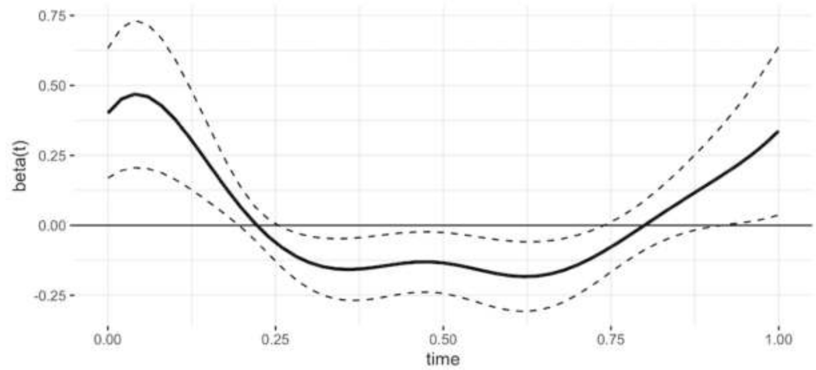
Even examining this more behavioral measure, however, results remain similar (Figure 4; parameter estimates are provided in Web Appendix Table A3).[7] Customers purchased more when agents use affective language at the beginning and end of the call, but cognitive language

[7] The beta curve for agent affective language is highly similar when we exclude the control variables, as is the beta curve for cognitive language, but with larger confidence bands (see Web Appendix Figure A3).

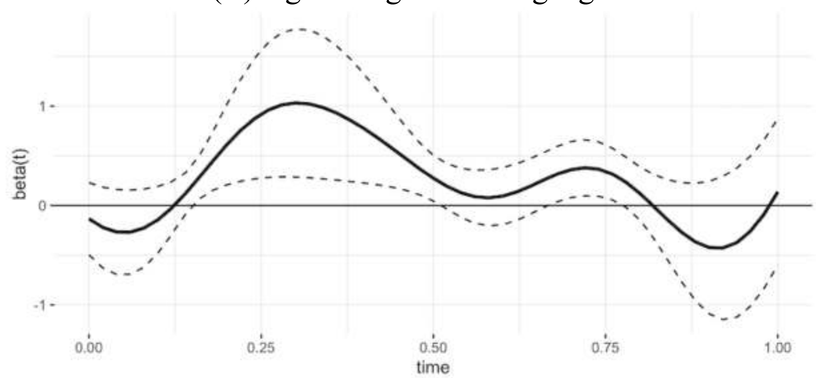
in the middle. This demonstrates the importance of language dynamics for firm performance, and casts further doubt on reverse causality (since purchase follows the language used in the call).[8]

Figure 4: Beta Curves for Agent Affective (A) and Cognitive (B) Language in Relation to Customer Order Count within 30 Days Post Interaction

(A) Agent Affective Language



(B) Agent Cognitive Language



[8] To account for the possibility of an interactive effect between agent's use of affective and cognitive language, we also considered models including a functional interaction term for affective and cognitive language. Three of the four resulting beta curves replicate the main results when we include this additional variable. The beta curve of agent affective language on purchases changed such that affective language remains important at the end of the conversation, but not at the start (see Web Appendix Figures A4 and A5).

Model Comparisons

Although our model can uniquely produce moment-to-moment insights over the course of a conversation, one may still wonder whether it performs better in both in-sample fit and out-of-sample prediction relative to several competing benchmarks. Therefore, we compare our main model, described here as Model 5 (M5), against the following alternative specifications:

M1: Simple "what" analysis. In this model we remove dynamics, aggregate data from turn level to conversation level, and estimate a static model of agent affective and cognitive language as predictors of customer satisfaction in a multivariate Lasso regression. The aggregate model includes all of the static controls, as well as the conversation-level averages of the dynamic agent and customer language and paralanguage features.[9] This benchmark thus compares our "when" approach to the "what" approach used in most prior research examining language in marketing.

M2: "What" analysis with conversational stages. This benchmark breaks the conversation into three conversational stages following Marinova, Singh, and Singh (2018). Judges dummy coded each turn as part of one of three conversational stages: *Sensing*, *Solving*, and *Settling*. The *Sensing* stage averaged 12% of the interaction, the *Solving* stage about 83%, and the *Settling* stage the last 5% of a given conversation. In each coded stage, we compute the turn-level average uses of affective and cognitive language by agent. We also include the same set of controls as in M1.

---

[9] Model estimates suggest that if we had only analyzed these language features at conversation-level, consistent with prior research, we would have concluded that agents should use only one of either affective or cognitive language, but not both. The call-level model estimates indicate that customer satisfaction has a positive relationship with agent affective language ($b = 0.05$, $p < 0.05$), and a negative but non-significant relationship with agent cognitive language ($b = -0.04$, $p > 0.1$). These findings are more consistent with the psychology literature's recommendation of prioritizing warmth (Godfrey, Jones, and Lord 1986; Holoien and Fiske 2013; Wang et al. 2017) than the competence-oriented speaking style recommended in recent customer service research (e.g., Kirmani et al. 2017; Singh et al. 2018).

M3: Functional model with wide data. Here we estimate the standalone model specified in (8) without Group-Lasso or the random intercept. In this case, the data that produce the results are particularly wide, i.e., the number of predictors is comparable to the number of observations.

M4: Homogeneous Functional model with Group-Lasso. In this model we integrate Group-Lasso into M3 but ignore the agent heterogeneous effect.

Table 1 reports the model comparison results based on root mean square error (RMSE), mean absolute deviation (MAD), and the correlation between the predicted value and the actual outcome. When conducting out-of-sample prediction, we hold out conversations from the entire data one by one using the leave-one-out cross-validation strategy (Hui et al. 2014).

Table 1: Measures of In-Sample Fit and Out-of-Sample Prediction
for Benchmark Models (M1-M4) versus Main Model (M5)

		M1	M2	M3	M4	M5
In-Sample Fit	RMSE	1.52	1.45	0.52	0.65	0.64
	MAD	1.35	1.22	0.37	0.49	0.47
	Correlation	0.30	0.21	0.96	0.91	0.92
Out-of-Sample Prediction	RMSE	1.65	1.63	2.28	0.98	0.97
	MAD	1.49	1.39	1.73	0.77	0.77
	Correlation	0.23	0.31	0.41	0.80	0.80

Traditional “what” analyses (M1 and M2) that ignore conversation dynamics yield clearly poorer in-sample and out-of-sample predictions than our functional framework (M4 and M5). The functional regression model that uses high dimensional data (M3) improves in-sample fit relative to its counterparts with Group-Lasso (M4 and M5), but its out-of-sample prediction deteriorates significantly due to overfitting. One can discern that the out-of-sample prediction of M3 is sometimes even worse than the static “what” analyses (M1 and M2), highlighting the importance of model regularization in functional regression on wide data. Further, the incorporation of the heterogeneous agent intercept offers little benefit (M5 vs. M4), likely because the number of observations (185) is too small for a panel dataset with 130 agents.

Taken together, the model comparison exercise suggests our approach offers superior predictive performance relative to previous models, in addition to its unique dynamic insights.[10]

Robustness

Alternative Measures of Affective and Cognitive Language Styles. The affective and cognitive language measures used in the current study have been extensively validated in prior work (cf. reviews by Kahn et al. 2007 and Tausczik and Pennebaker 2010), but one could wonder whether they might miss certain idiosyncratic features of customer service conversation. To address this possibility, we apply custom dictionaries from prior customer service research (Marinova et al. 2018; Singh et al. 2018). These works combined established dictionaries (LIWC) and human judging to develop custom lists of service-oriented "relating" (i.e., affective) words ($\mathrm{N} = 247$) and "resolving" (i.e., cognitive) words ($\mathrm{N} = 649$). We scored all agent and customer conversational turns using this approach, and estimated our main model with these alternative measures instead of LIWC to test robustness.

Results are similar. As before, customers are more satisfied when agents use the alternative affective language measure ("relating") during a conversation's start and end, but less satisfied when this language is used in the middle (see Web Appendix Figure A6). Similarly, for cognitive

[10] In addition to the benchmark model comparisons, one could still wonder whether prior work's suggestion to exclusively use an affective or cognitive style may be best, or how much "when" one uses these styles matters if one tries to use both. To probe these question, we performed a series of simulations comparing our model with a variety of alternatives. Simulation results further supported our dynamic model approach. See Web Appendix for detailed procedure and results.

language, customers are more satisfied when agents use "resolving" language in the middle of the call, but less satisfied when such language occurs at the beginning of the call. The modeling for purchase count dependent variable shows similar results (see Web Appendix Figure A7).

Note that these results differ from prior work. The research that developed these dictionaries (Marinova et al. 2018) found that only agent cognitive language was positively linked to their dependent measure (i.e., human judgement of customer emotion). They found that affective language impeded cognitive language's benefits when both were included in the model, supporting the warmth/competence trade-off and a recommendation to focus exclusively on competence-oriented cognitive language. These differences are likely driven by our dynamic modeling approach, but may also be due in part to distinctions in the specific customer service contexts (airline counter service vs. online retailing), or the different dependent measures (e.g., third-party judgment of displayed affect vs. customer satisfaction self-reports).

Valenced Subsets of Affective Language. While LIWC's affective process dictionary is often used to capture warmth, one could argue that "warm" affective language should contain only positive emotional words (e.g., happy and wonderful) and exclude negative ones (e.g., sad and disappointed). Agents often use negative affective language in a warm manner to convey empathy (e.g., "I'm disappointed we didn't deliver your order on time"), but to test the contribution of each valence we repeat the main analysis incorporating agents' positive and negative affective words as separate predictors.

Results are again similar. The beta curve for positive affective language is close to that of the full affective language dictionary, while negative affective language also appears to contribute positively, albeit only at the end (see Web Appendix Figure A8). A review of the negative affect words used in the conversational closings reveals that the presence of words like “sorry,” “problem,” and “wrong” are positively correlated with customer satisfaction (i.e., “sorry about that” or “Glad we could fix the problem”). Our functional approach appears to capture such subtle conversational language features well. Positive affective language has similar relationships with purchase, but the effects are reduced for negative affective language (see Web Appendix Figure A9).

STUDY 2: GENERALIZABILITY

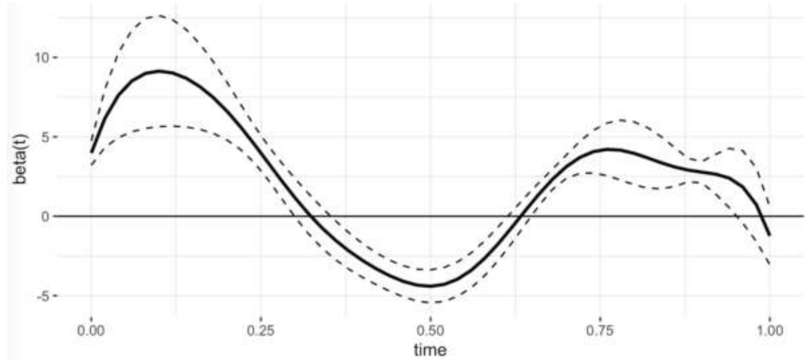
While these results are intriguing, one might wonder whether they are driven by the specific firm, industry, or customer satisfaction measure used. To test the generalizability of the results, we acquired an additional dataset of 204 customer calls (11,548 conversational turns) from a major U.S. airline, and applied the same analyses. Rather than agent helpfulness at the end of calls, the airline captured willingness to recommend as a Net Promoter Score (NPS; Reichheld 2003), a widely used approach to assessing customer satisfaction and purchase intent (Keiningham et al. 2007; van Doorn, Leeflang, and Tijs 2013).

We created a similar set of controls as in Study 1, including *Call Content* (six dummies provided by the firm, as well as the results of an LDA topic model capturing the latent mixture of call topics, *Complexity* (length in words), *Dynamics of Other Major Agent and Customer Language Features* (the same LIWC measures as in the main dataset), *Dynamics of Customer Affective and Cognitive Language,* and *Moment-to-Moment Synchronicity*. Unfortunately, data on agent and customer characteristics were not available.

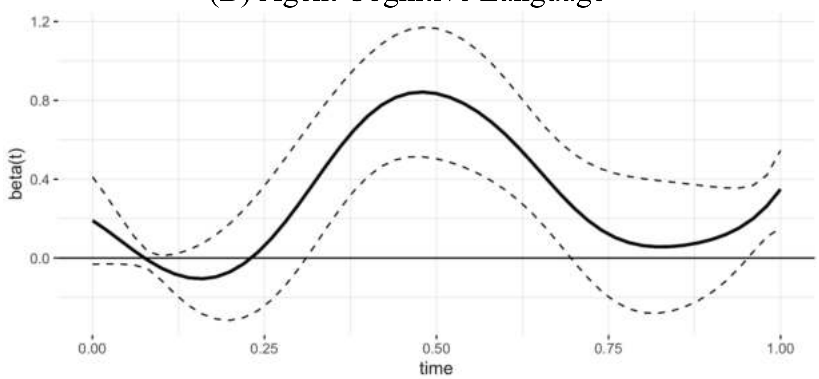
Even using a different company, in a completely different industry, results are similar to those found in the main analysis. Customers had higher willingness to recommend the firm when agents used more affective language at the start and end of the conversation, but more cognitive language in the middle (Figure 5).

Figure 5: Beta Curves for Agent Affective (A) and Cognitive (B) Language in Relation to Willingness to Recommend (NPS) for A Major U.S. Airline

(A) Agent Affective Language



(B) Agent Cognitive Language



STUDY 3: EXPERIMENTAL EVIDENCE

Finding consistent results across two completely different field datasets underscores their validity, and demonstrates how our modeling approach sheds light on conversation dynamics.

That said, one could still wonder whether the observed effects are truly causal. Including dozens of control variables casts doubt on alternative explanations, but to provide even more direct evidence, we conduct a simple experiment. We manipulate language to test whether, compared to the strategy recommended in prior research (i.e., emphasizing competence throughout), our findings from analyzing conversation dynamics (i.e., warm at the start and end, and competent in the middle) boosts customer satisfaction.

Method

Note that the challenges that make causal inference difficult in empirical data analysis also make it difficult to study conversations experimentally. Because conversations are interactive, they can quickly go in numerous different directions, making it hard to maintain experimental control. Consequently, to ensure careful control over the language used, we randomly assigned participants (N = 292, Amazon Mechanical Turk) to one of two versions of a simple scenario. They imagined calling an online retailer and read a six-turn conversation in which they asked the customer service agent to help them with free shipping (see Web Appendix for stimuli).

The only difference between conditions was the agent's language. For half of the participants (all cognitive condition), the agent used cognitive language throughout, as recommend by prior work. For the other half (dynamic condition), the agent's language followed the recommendations of the dynamic model (i.e., in the first and last 25% of the conversation, cognitive language was replaced by more affective language using the relevant LIWC dictionary). In the all cognitive condition, for example, the agent started by saying "What might I do for you today?", while in the dynamic condition they used the warmer "How can I help you today?"

Then, participants completed the key dependent variable (i.e., customer satisfaction, "How satisfied are you with the agent?"; 1 = not at all, 7 = very much). To test whether warmth and/or competence perceptions drove any observed effects, we also measured these perceptions ("How warm was the agent?" and "How competent was the agent?") on the same seven-point scale.

Results and Discussion

As predicted, revising agent language based on the dynamic model's findings improved customer satisfaction. Replacing cognitive language with affective language at the start and end of the conversation increased customer satisfaction ($M_{dynamic}$ = 6.06 vs. $M_{cognitive\ only}$ = 5.80; $F(1, 290) = 5.42$, $p = .021$).

Further, mediation analysis (PROCESS model 4; Hayes 2018) considering warmth and competence as parallel mediators confirmed that shifting agent language enhanced satisfaction because it increased perceived warmth (indirect effect = .045, 95% CI [.013, .087]). Using more affective words at the start and end made the agent seem more warm (b = .260, $p = .003$), which subsequently increased customer satisfaction (b = .173, $p < .001$).[11] Given competent language was included in both conditions when it was likely to matter, competence perceptions did not vary by condition and thus did not play a role (indirect effect = .091, 95% CI [-.006, .197]).

These results provide stronger causal evidence that the linguistic recommendations of our dynamic model can improve customer satisfaction.

GENERAL DISCUSSION

Language impacts a range of customer interactions. But while a great deal of research has examined customer service language and other marketing dialogues (e.g., social media conversations; Berger and Schwartz 2011; Ordenes et al. 2017), *when* different language features matter in conversation has received less attention.

[11] To replicate the online fashion retailer's satisfaction measure, we also asked "How helpful was the agent?". Results were the same. Warmth (indirect effect = .038, SE = .017, 95% CI [.010, .077]) but not competence (indirect effect = .085, SE = .047, 95% CI [-.007, .181]) perceptions drove the relationship between warmth/competence dynamics and helpfulness. Using more affective words at start and end increased perceptions of warmth (b = .139, $t = 1.82$, $p = .070$) which boosted perceived helpfulness (b = .146, $t = 3.74$, $p < .001$).

To address this gap, we developed an approach that examines how conversational dynamics relate to important customer service outcomes. As an initial demonstration of its potential, we applied it to the most important dimensions of person perception, warmth and competence. While existing work looking at interactions as a whole suggests that communicating in a manner that emphasizes competence is best, our more dynamic perspective suggests a different approach may be more impactful. Specifically, "bookending" the efficient, competent addressing of customer needs with warmer, more affective rapport building can increase satisfaction. Further, launching straight into the competence-oriented language endorsed by prior research may hurt satisfaction and purchase, as may using only a warmth-oriented approach. The result was supported across three studies, including field data analysis of two different firms, in two industries, using different dependent measures linked to customer attitudes and purchase behaviors, and in a simple lab experiment.

Our approach helps address three major challenges in modeling conversational dynamics: sparsity, irregularity, and high dimensionality. Linguistic measurement of human language is inevitably irregular and sparse, so we modeled the time-varying data as random trajectories realized from smooth underlying functions. Conversations also yield shallow (few observations) but wide data situations in which a large number of verbal and vocal features need to be accommodated to strengthen inference (Zhang et al. 2020). To achieve model regularization within the functional analysis framework, we incorporated Group-Lasso from the machine learning literature to automatically select functional and scalar variables to enter the functional regression and avoid overfitting the noise from the data. The flexibility inherent in this method allows us to retain the focal predictors of known interest (e.g., affective and cognitive language) while penalizing other variables to find the most statistically meaningful set of available controls.

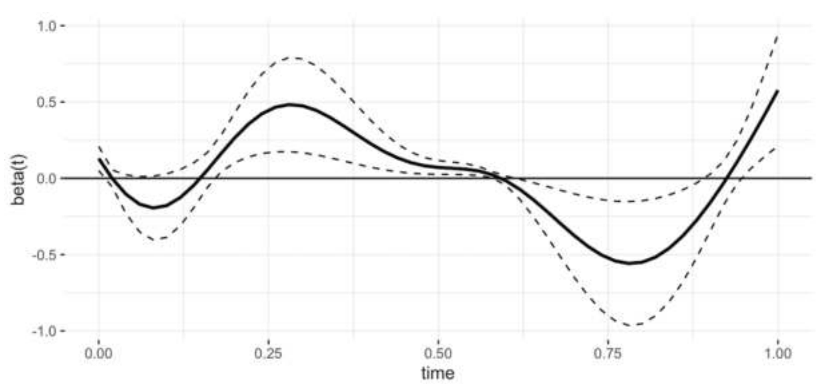
Applications to Other Linguistic Features

Our work focused on affective and cognitive language, but the same approach can be applied to any other conceptually or substantively important language feature.

Take asking questions. Prior research suggests asking questions is beneficial (Brooks and John 2018; Huang et al. 2017) because it signals interest in the customer's issue (Brody 1994; Drollinger and Comer 1997). Consumers also believe that asking questions is an important agent behavior, making it a common feature of scales used to evaluate agent performance (Drollinger et al. 2006; Ramsey and Sohi 1997).

But *when* should agents ask questions? Ignoring dynamics for a moment, our main dataset confirms that customers were indeed more satisfied when agents asked more questions (b = .13, p = .010). But is this true at any moment in a conversation?

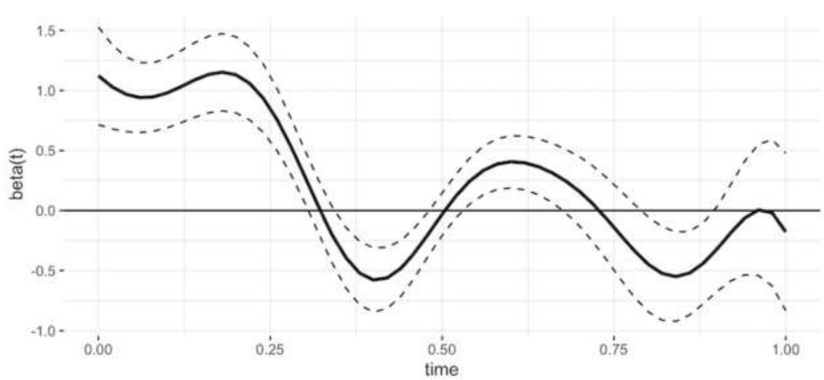
Figure 6: Beta Curve for Agent Question Asking in Relation to Customer Satisfaction



To illustrate how our method can test such ideas, we run the same model but with agent question-asking as the main dynamic predictor of customer satisfaction. Results indicate that customer satisfaction depends on *when* agents ask questions (Figure 6). Agent question asking has a positive relationship with satisfaction when used between 15% and 50% of the way through interaction, but a negative relationship afterwards. This suggests agents should only ask questions after the customer has a chance to describe their needs, potentially consistent with the notion of a "sensing" stage in customer service (Marinova et al. 2018). Results further suggest it may be beneficial to ask a question at the conversation's close. Asking "Is there anything else I can help with?" shortly before the conversation ends may well be an important practice.

To further explore the value of the method, we also looked at pronouns. Research suggests that using first person singular ("I") pronouns makes the agent seem more agentic and empathetic (Packard et al. 2018), but are such pronouns useful throughout a conversation? A traditional conversation level *what* analysis suggests that first person singular pronouns are positively related to customer satisfaction (b = .051, p = .040), but does not speak to when they might be more beneficial. To begin to address this question, we run the same model with agent first person singular pronouns as the main dynamic predictor of customer satisfaction.

Figure 7: Beta Curve for Agent First Person Singular Pronouns
in Relation to Customer Satisfaction



Results suggest that the benefits of first person singular pronouns occur mostly at the beginning of conversations (Figure 7). This is the same period when warm, affective language is beneficial. In contrast, first person singular pronouns may be costly for a brief period when cognitive language matters (i.e., the middle of the conversation). This pattern of results suggests that the empathetic (i.e., warm) dimension of "I" pronouns identified in Packard et al. (2018) may be more important than the agentic (i.e., competent) dimension.

Overall, these examples further underscore the potential value of language dynamics, demonstrating not only whether they matter, but *when*.

Implications, Limitations, and Future Research

Our findings have clear implications for both researchers and managers. For researchers, moving beyond identifying *what* firm agents should say, or types of language to use, our approach offers model-based suggestions on *when* to say it. This method expands the toolkit available to researchers who use text analysis to understand consumer behavior and marketing problems (Berger et al. 2020), offering more nuanced insight into the role of language features of interest. It could easily be applied to explore dynamics of paralanguage (Luangrath, Peck and Barger 2017) or non-verbal communications as well.

Managers might use the approach to understand not only what language to use, but when to use it more effectively. Moreover, as many organizations look to integrate digital presence for service agents into web experiences (Herhausen et al. 2020), or introduce automated chatbots and other forms of verbal artificial intelligence into their customer service experience, a better understanding of the optimal temporal application of language features may help make these conversational technologies more productive.

While we hope the method reported here is widely useful, the exact findings regarding the temporal importance affective and cognitive language may depend somewhat on the context. The results replicate across two different industries and firms, but as with many field investigations, different contexts may yield varying outcomes. The best time to use affective language in sales calls, for example, may be different from the ideal time in calls that involve solving customer issues. How important cognitive language is at different conversational times may depend on how severe the customer's issue seems. In the Web Appendix we present an example of such an analysis, considering moderation of the main model results by call severity.

In live store interactions, it is possible that service employees can build rapport using non-verbal information (e.g., facial expression and posture). In this case the importance of warm, affective language may be diminished to some extent. Applying the method outlined here to additional contexts may provide further insight.

Accounting for a large number of agent-, customer- and firm-level factors, as well as a range of dynamic language and paralanguage features helps account for most plausible sources of variation. As with any analysis of field data, however, our estimates remain subject to potential endogeneity. While the temporal sequence of our language predictors and outcomes do not support reverse causality, and an initial experiment supports causality, future research could pursue field experiments, where causality could be assessed with greater external validity. That said, our approach contributes a novel lens on the temporal importance of conversational features, offering a model to accommodate a rich array of interaction-level and moment-to-moment variations that have been largely overlooked.

Future work might build on these findings in a number of ways. The functional regression framework takes the dynamic language features as given, for example, without looking into the underlying mechanisms of how a particular feature emerges in conversation, or how different features may enhance or diminish each other. Future research could conduct a "cost assessment" of a language feature, supporting determination of an optimal level of that feature over conversational time. Future work could also use these methods to investigate people's tendency to adapt their own language to that of others (Danescu-Niculescu-Mizil et al. 2013). In addition, future work could study conversational dynamics across domains. When certain linguistic features are beneficial in doctor-patient or lawyer-client conversations, for example, may differ from what was observed in our marketing context.

This research takes an important step toward quantifying the dynamic role of language in conversation. While we focused on customer service language, the modeling approach should also be useful in studying word of mouth, negotiations, message recall, and various other topics. We hope this work provides a useful framework for those examining conversations in the marketing domain, and beyond.

REFERENCES

Abele, Andrea E. and Bogdan Wojciszke (2007), "Agency and Communion from the Perspective of Self Versus Other," *Journal of Personality and Social Psychology*, 93(5), 751-763.

Bardovi-Harlig, K., B. A. S. Hartford, R. Mahan-Taylor, M. J. Morgan, and D. W. Reynolds (1991), "Developing Pragmatic Awareness: Closing the Conversation," *ELT Journal*, 45(1), 4-15.

Berger, Jonah, Ashlee Humphreys, Stephan Ludwig, Wendy Moe, Oded Netzer and David Schweidel (2020), "Uniting The Tribes: Using Text for Marketing Insight," *Journal of Marketing*, 84(1), 1-25.

Berger, Jonah and Eric M. Schwartz (2011), "What Drives Immediate and Ongoing Word of Mouth?" *Journal of Marketing Research*, 48(5), 869-880.

Berry, Diane S., James W. Pennebaker, Jennifer S. Mueller, and Wendy S. Hiller (1997), "Linguistic Bases of Social Perception," *Personality and Social Psychology Bulletin*, 23(5), 526-537.

Blanding, Warren (1989), "Customer Service Communications: Telephone Systems and Procedures," *Practical Handbook of Customer Service Operations*, 377-407.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan (2003), "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, 3, 993-1022.

Boersma, Paul and Vincent van Heuven (2001), "Speak and UnSpeak with PRAAT," *Glot International*, 5(9), 341-347.

Bolton, Ruth N. and James H. Drew (1991), "A Longitudinal Analysis of The Impact of Service Changes on Customer Attitudes," *Journal of Marketing*, 55(1), 1-9.

Brody, Nathan (1994), "Cognitive Abilities," *Psychological Science*, 5(2), 63-68.

Brooks, Alison Wood, and Leslie John (2018), "The Surprising Power of Questions," *Harvard Business Review,* 96(3), 60-67.

Castleberry, Stephen B., C. David Shepherd, and Rick Ridnour (1999), "Effective Interpersonal Listening in The Personal Selling Environment: Conceptualization, Measurement, and Nomological Validity," *Journal of Marketing Theory and Practice*, 7(1), 30-38.

Cespedes, Frank V. and Christopher Wallace (2017). "Executives and Salespeople Are Misaligned — and The Effects Are Costly," *Harvard Business Review* (website).

Chen Kehui, Xiaoke Zhang, Alexander Petersen, and Hans-Georg Muller (2017), "Quantifying Infinite-Dimensional Data: Functional Data Analysis in Action," *Statistics in Biosciences*, 9, 582-604.

Chen Qingxia, Hui Nian, Yuwei Zhu, H. Keipp Talbot, Marie R. Griffin, and Frank E. Harrell Jr (2016), "Too Many Covariates and Too Few Cases? A Comparative Study," *Statistics in Medicine*, 35(25), 4546-4558.

Cronin, Joseph J. and Steven A. Taylor, "Measuring Service Quality: A Reexamination and Extension," *Journal of Marketing*, 56(3), 55-68.

Danescu-Niculescu-Mizil, Cristian, Robert West, Dan Jurafsky, Jure Leskovec, and Christopher Potts (2013), "No Country for Old Members: User Lifecycle and Linguistic Change in Online Communities," *Proceedings of the 22nd International Conference on the World Wide Web*, 307-318.

de Ruyter, Ko and Martin G. M. Wetzels (2000), "The Impact of Perceived Listening in Voice-to-Voice Service Encounters," *Journal of Service Research*, 2(3), 276-284.

Decter-Frain, Ari and Jeremy A. Frimer (2016), "Impressive Words: Linguistic Predictors of Public Approval of The U.S. Congress," *Frontiers of Psychology*, 23, 1-13.

DeWitt, Tom and Michael K. Brady (2003). "Rethinking Service Recovery Strategies: The Effect of Rapport on Consumer Responses to Service Failure," *Journal of Service Research*, 6(2), 193-207.

Drollinger, Tanya, and Lucette B. Comer (1997), "A Conceptual Framework of Active Empathetic Listening in The Sales Encounter," *National Conference on Sales Management Proceedings*, 142-145.

Drollinger, Tanya, Lucette B. Comer, and Patricia T. Warrington (2006), "Development and Validation of The Active Empathetic Listening Scale," *Psychology & Marketing*, 23(2), 161-180.

Fiske, Susan T., Amy J. C. Cuddy, and Peter Glick (2007), "Universal Dimensions of Social Cognition: Warmth and Competence," *Trends in Cognitive Sciences*, 11(2), 77-83.

Foutz, Natasha Z. and Wolfgang Jank (2010), "Prerelease Demand Forecasting for Motion Pictures Using Functional Shape Analysis of Virtual Stock Markets," *Marketing Science*, 29(3), 568-579.

Gabor, Don (2011), *How to Start A Conversation and Make Friends*, New York, NY: Simon and Schuster.

Godfrey, Debra K., Edward E. Jones, Charles G. Lord (1986), "Self-Promotion Is Not Ingratiating," *Journal of Personality and Social Psychology*, 50(1), 106-115.

Greene, Robert L. (1986), "Sources of Recency Effects in Free Recall," *Psychological Bulletin*, 99(2), 221-228.

Gremler, Dwayne D. and Kevin P. Gwinner (2000), "Customer-Employee Rapport in Service Relationships," *Journal of Service Research*, 3(1), 82-104.

Grewal, Dhruv, Dennis Herhausen, Stepan Ludwig, and Francisco Villarroel Ordenes (2021), "The future of digital communication research: Considering dynamics and multimodality," *Journal of Retailing*, in press, https://doi.org/10.1016/j.jretai.2021.01.007.

Hayes, Andrew F. (2018), "Partial, Conditional, and Moderated Moderated Mediation: Quantification, Inference, and Interpretation," *Communication Monographs*, 85(1), 4-40.

Herhausen, Dennis, Oliver Emrich, Dhruv Grewal, Petra Kipfelsberger, and Marcus Schoegel (2020), "Face Forward: How Employees' Digital Presence on Service Websites Affects Customer Perceptions of Website and Employee Service Quality," *Journal of Marketing Research*, 57(5), 917-936.

Holoien, Deborah S. and Susan T. Fiske (2013), "Downplaying Positive Impressions: Compensation between Warmth and Competence in Impression Management," *Journal of Experimental Social Psychology*, 49(1), 33-41.

Huang, Karen, Michael Yeomans, Alison W. Brooks, Julia Minson, and Francesca Gino (2017), "It Doesn't Hurt to Ask: Question-Asking Increases Liking," *Journal of Personality and Social Psychology*, 113(3), 430-452.

Hui, Sam K., Tom Meyvis, and Henry Assael (2014), "Analyzing Moment-to-Moment Data Using A Bayesian Functional Linear Model: Application to TV Show Pilot Testing," *Marketing Science*, 33(2), 222-240.li

Jasmand, Claudia, Vera Blazevic, and Ko de Ruyter (2012), "Generating Sales While Providing Service: A Study of Customer Service Representatives' Ambidextrous Behavior," *Journal of Marketing*, 76(1), 20-37.

Judd, Charles M., Laurie James-Hawkins, Vincent Yzerbyt, and Yoshihisa Kashima (2005), "Fundamental Dimensions of Social Judgment: Understanding The Relations between Judgments of Competence and Warmth," *Journal of Personality and Social Psychology*, 89(6), 899-913.

Kahn, Jeffrey H., Renee M. Tobin, Audra E. Massey, and Jennifer A. Anderson (2007), "Measuring Emotional Expression with the Linguistic Inquiry and Word Count," *American Journal of Psychology*, 120, 263-286.

Kaski, Timo, Jarkko Niemi, and Ellen Pullins (2018), "Rapport Building in Authentic B2B Sales Interactions," *Industrial Marketing Management*, 69, 235-252.

Keiningham, Timothy L., Bruce Cooil, Tor Wallin Andreassen, and Lerzan Aksoy (2007), "A Longitudinal Examination of Net Promoter and Firm Revenue Growth," *Journal of Marketing*, 71, 39-51.

Kirmani, Amna, Rebecca Hamilton, Debora Thompson, and Shannon Lantzy (2017), "Doing Well vs. Doing Good: The Differential Effect of Underdog Positioning on Moral and Competent Service Providers," *Journal of Marketing*, 81(1), 103-117.

Ladhari, Riadh (2008), "Alternative Measures of Service Quality: A Review," *Managing Service Quality: An International Journal*, 18(1), 65-86.

Li, Xueni, Kimmy W. Chan, and Sara Kim, "Service with Emoticons: How Customers Interpret Employee Use of Emoticons in Online Service Encounters," *Journal of Consumer Research*, 45(5), 973-987.

Luangrath, Andrea W., Joann Peck, and Victor A. Barger (2017), "Paralanguage and its Implications for Marketing Communications," *Journal of Consumer Psychology*, 27(1), 98-107.

Marinova, Detelina, Sunil Singh and Jagdip Singh (2018), "Frontline Problem-Solving Interactions: A Dynamic Analysis of Verbal and Nonverbal Cues," *Journal of Marketing Research*, 55(2), 178-192.

McBain, Jay (2020), "What I See Coming for The Channel in 2020," *Forrester Featured Blogs*.

Meier, Lukas, Sara Van De Geer, and Peter Buhlmann (2008), "The Group Lasso for Logistic Regression," *Journal of the Royal Statistical Society: Series B*, 70, 53-71.

Morgan, Blake (2017), "Customer Service is a $350 Billion Industry, and It's a Mess," *Forbes*.

Ordenes, Franscsco V., Stephan Ludwig, Ko de Ruyter, Dhruv Grewal, and Martin Wetzels (2017), "Unveiling What Is Written in The Stars: Analysis Explicit, Implicit, and Discourse Patterns of Sentiment in Social Media," *Journal of Consumer Research*, 43(6), 875-894.

Ordenes, Franscsco V., Babis Theodoulidis, Jamie Burton, Thorsten Gruber, and Mohamed Zaki (2014), "Analyzing Customer Experience Feedback Using Text Mining: A Linguistics-Based Approach," *Journal of Service Research*, 17(3), 278-295.

Packard, Grant and Jonah Berger (2020), "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 Pronouns Use in Customer-Firm Interactions," *Journal of Marketing Research*, 55(4), 541-555.

Parasuraman, Ananthanarayanan, Leonard L. Berry, and Valarie A. Zeithaml (1991), "Understanding Customer Expectations of Service," *Sloan Management Review*, 32(3), 39-48.

Parasuraman, Ananthanarayanan and Valarie A. Zeithaml (2002), "Understanding and Improving Service Quality: A Literature Review and Research Agenda," *Handbook of marketing*, 339-367.

Parasuraman, Ananthanarayanan, Valarie A. Zeithaml, and Leonard L. Berry (1985), "A Conceptual Model of Service Quality and Its Implications for Future Research," *Journal of Marketing*, 49(4), 41-50.

Pennebaker, James W., Ryan L. Boyd, Kayla Jordan, and Kate Blackburn (2015), *The development and psychometric properties of LIWC2015*, https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf

Placencia, Maria E. (2004), "Rapport-Building Activities in Corner Shop Interactions," *Journal of Sociolinguistics*, 8(2), 215-245.

Pogacar, Ruth, L. J. Shrum, and Tina M. Lowrey (2018), "The Effects of Linguistic Devices on Consumer Information Processing and Persuasion: A Language Complexity x Processing Mode Framework," *Journal of Consumer Psychology*, 28, 689-711.

Ramsay, James O. and Bernard W. Silverman (1997), *Functional Data Analysis*, New York: Springer.

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), 127-137.

Reichheld, Frederick F. (2003), "The One Number You Need to Grow," *Harvard Business Review*, 81, 46-54.

Rust, Roland T. and Tuck S. Chung (2006), "Marketing Models of Service and Relationships," *Marketing Science*, 25(6), 551-765.

Schegloff, Emanuel A. and Harvey Sacks (1973), "Opening Up Closings", *Semiotica*, 8(4), 289-327.

Singh, Sunil, Detelina Marinova, Jagdip Singh, and Kenneth Evans (2018), "Customer Query Handling in Sales Interactions," *Journal of the Academy of Marketing Science*, 46, 837- 856.

Snyder, Hannah, Lars Witell, Anders Gustafsson, Paul Fombelle, and Per Kristensson (2016), "Identifying Categories of Service Innovation: A Review and Synthesis of The Literature," *Journal of Business Research*, 69(7), 2401-2408.

Sood, Ashish, Gareth M. James, Gerard J. Tellis (2009), "Functional Regression: A New Model for Predicting Market Penetration of New Products," *Marketing Science*, 28(1), 36-51.

Speckman, Paul (1988), "Kernel Smoothing in Partial Linear Models," *Journal of the Royal Statistical Society, Series B*, 50(3), 413-436.

Spiro, Rosann L. and Barton A. Weitz (1990), "Adaptive Selling: Conceptualization, Measurement, and Nomological Validity," *Journal of Marketing Research*, 27(1), 61-69.

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.

Tibshirani, Robert (1996), "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society, Series B*, 58(1), 267-288.

Van Zant, Alex B. and Jonah Berger (2020), "How The Voice Persuades," *Journal of Personality and Social Psychology*, 118(4), 661-682.

Van Doorn, Jenny, Peter S. H. Leeflang, and Marleen Tijs (2013), "Satisfaction as A Predictor of Future Performance," *International Journal of Research in Marketing*, 30(3), 314-318.

Wang, Jane-Ling, Jeng-Min Chiou, and Hans-Georg Muller (2016), "Functional Data Analysis," *Annual Review of Statistics and Its Applications*, 3, 257-295.

Wang, Ze, Huifang Mao, Yexin Jessica Li, and Fan Liu (2017), "Smile Big or Not? Effects of Smile Intensity on Perceptions of Warmth and Competence," *Journal of Consumer Research*, 43(5), 787-805.

Xiong, Guiyang and Sundar Bharadwaj (2014), "Prerelease Buzz Evolution Patterns and New Product Performance," *Marketing Science*, 33(3), 401-421.

Yang, Yi and Hui Zou (2015), "A Fast Unified Algorithm for Solving Group-Lasso Penalized Learning Problems," *Statistics and Computing*, 25, 1129-1141.

Yao, Fang, Hans-Georg Muller, and Jane-Ling Wang (2005), "Functional Data Analysis for Sparse Longitudinal Data," *Journal of the American Statistical Association*, 100(470), 577-590.

Yuan, Ming and Yi Lin (2006), "Model Selection and Estimation in Regression with Grouped Variables," *Journal of the Royal Statistical Society, Series B*, 68(1), 49-67.

Zeithaml, Valarie A., Leonard L. Berry and Ananthanarayanan Parasuraman (1996), "The Behavioral Consequences of Service Quality," *Journal of Marketing*, 60(2), 31-46.

Zhang, Justine, Sendhil Mullainathan, and Cristian Danescu-Niculescu-Mizil (2020), "Quantifying the Causal Effects of Conversational Tendencies," *Proceedings of the ACM on Human-Computer Interaction*, 4(131), 2-24.

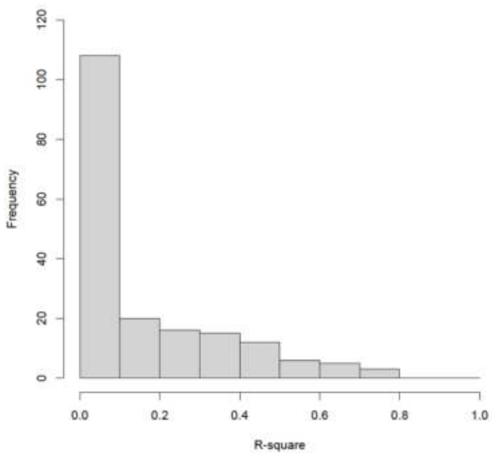
Zhang, Qiang, Wenbo Wang, and Yuxin Chen (2020), "In-Consumption Social Listening with Moment-to-Moment Unstructured Data: The Case of Movie Appreciation and Live Comments," *Marketing Science*, 39(2), 285-457.

Zou, Hui (2006), "The Adaptive Lasso and Its Oracle Properties," *Journal of the American Statistical Association*, 101(476), 1418-1429.

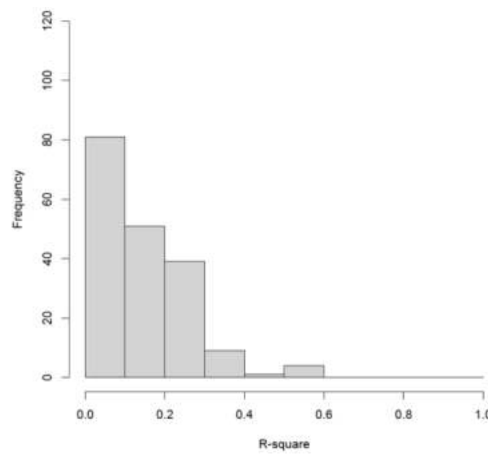
WEB APPENDIX

Figure A1: Variance in Agent Language Explained by Customer Language (R^2)

(A) Agent Affective Language



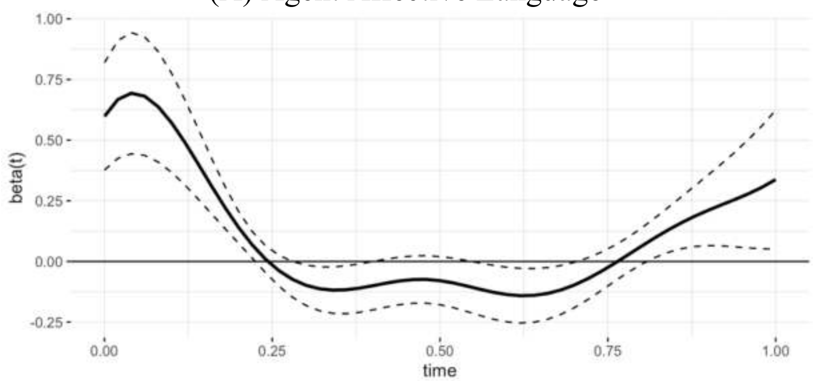
(B) Agent Cognitive Language



Note: The histograms summarize the linguistic synchronicity of agent's and customer's affective and cognitive language across the 185 conversations. Overall, some level of conversational synchronicity happens more frequently for cognitive language, but synchronicity occurs more deeply for affective language in the fewer conversations in which it is present.

Figure A2: Beta Curves for Agent Affective (A) and Cognitive (B) Language in Relation to Customer Satisfaction without Controls

(A) Agent Affective Language



(B) Agent Cognitive Language

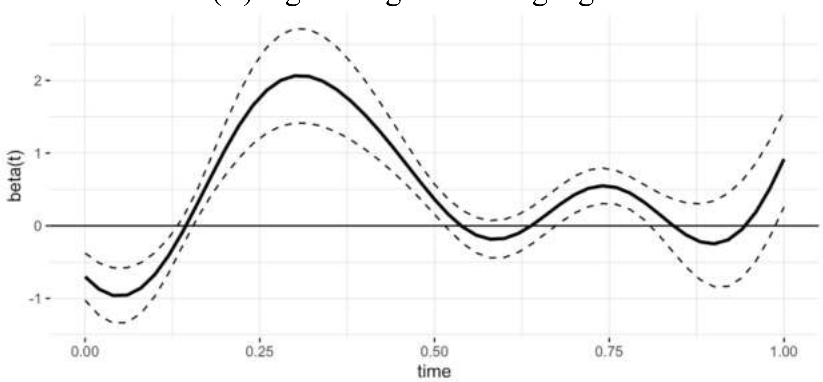
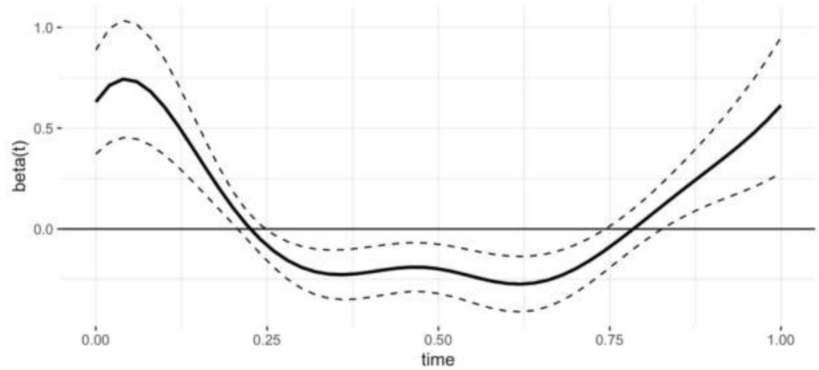


Figure A3: Beta Curves for Agent Affective (A) and Cognitive (B) Language in Relation to Order Count within 30 Days Post Interaction without Controls

(A) Agent Affective Language



(B) Agent Cognitive Language

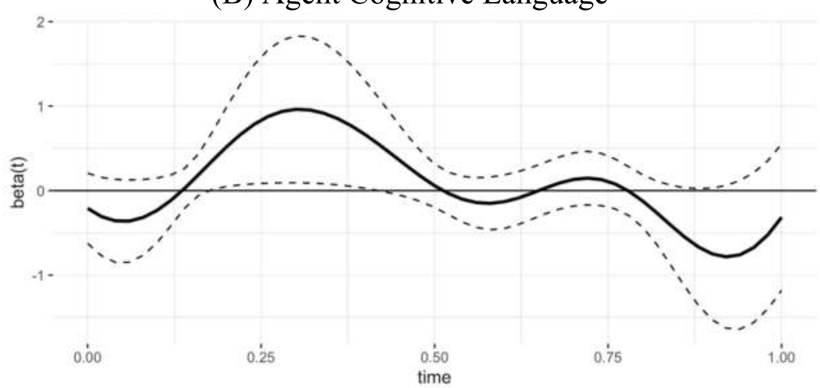
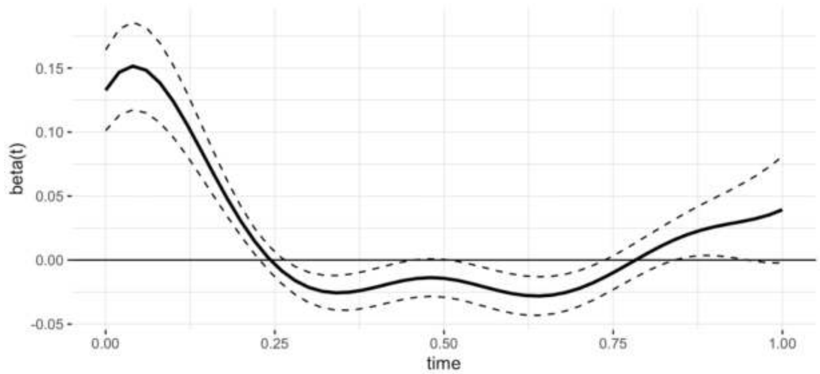


Figure A4: Beta Curves for Agent Affective (A) and Cognitive (B) Language in Relation to Customer Satisfaction, with an Agent Affective and Cognitive Language Interaction Term

(A) Agent Affective Language



(B) Agent Cognitive Language

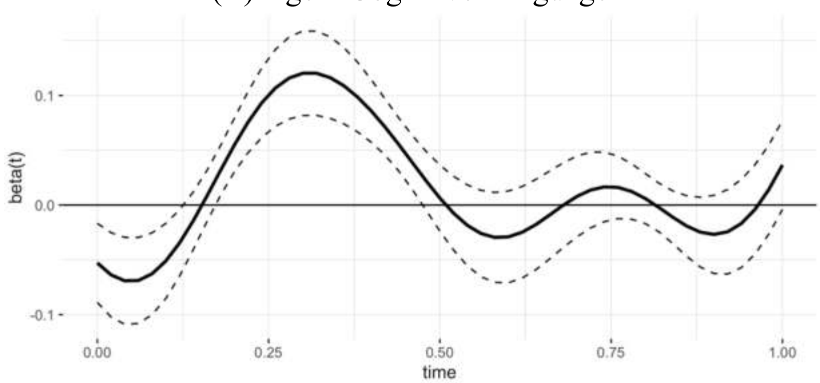
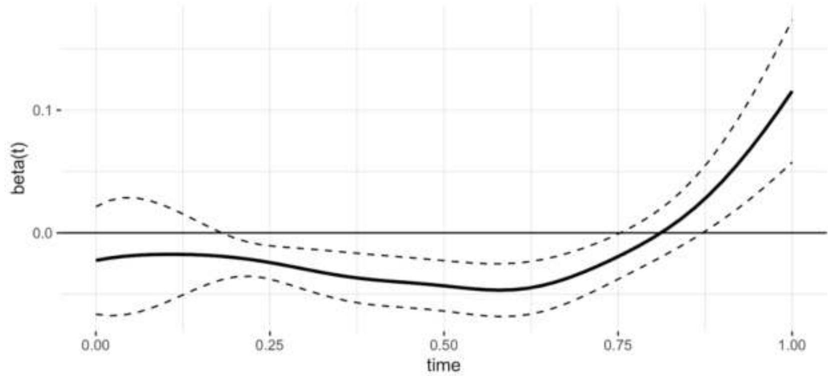


Figure A5: Beta Curves for Agent Affective (A) and Cognitive (B) Language in Relation to Order Count within 30 Days Post Interaction, with an Agent Affective and Cognitive Language Interaction Term

(A) Agent Affective Language



(B) Agent Cognitive Language

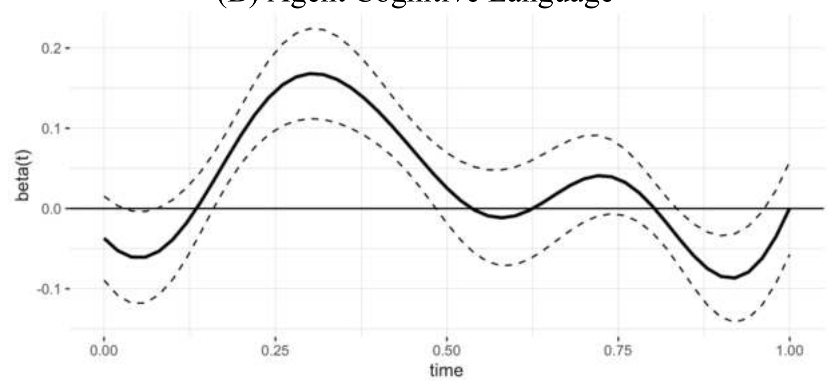
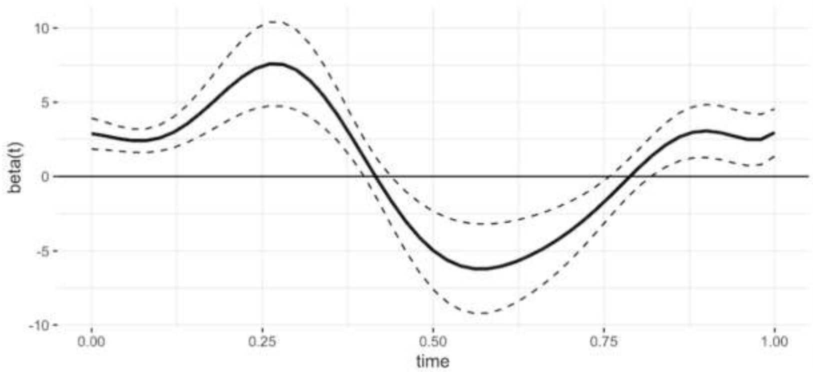


Figure A6: Beta Curves for Agent “Relating” (A) and “Resolving” (B) Language in Relation to Customer Satisfaction

(A) Agent “Relating” Language



(B) Agent “Resolving” Language

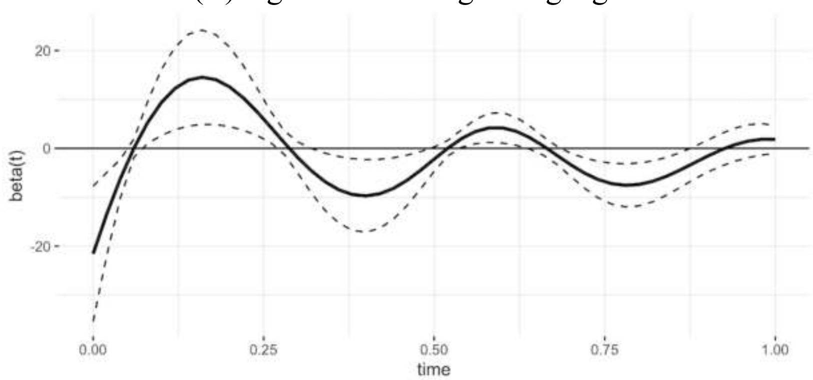
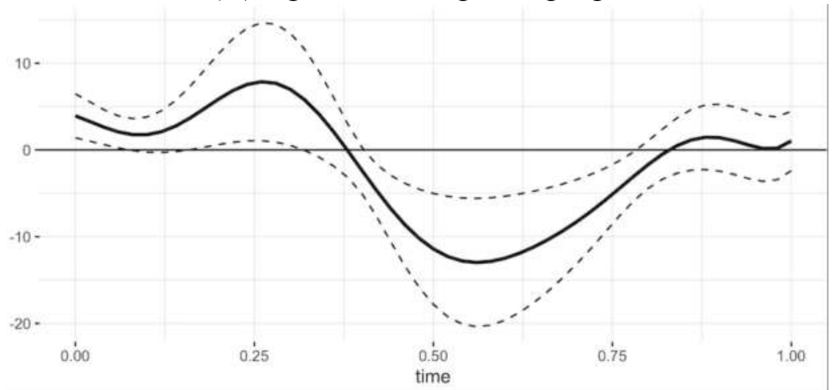


Figure A7: Beta Curves for Agent “Relating” (A) and “Resolving” (B) Language in Relation to Order Count within 30 Days Post Interaction

(A) Agent “Relating” Language



(B) Agent “Resolving” Language

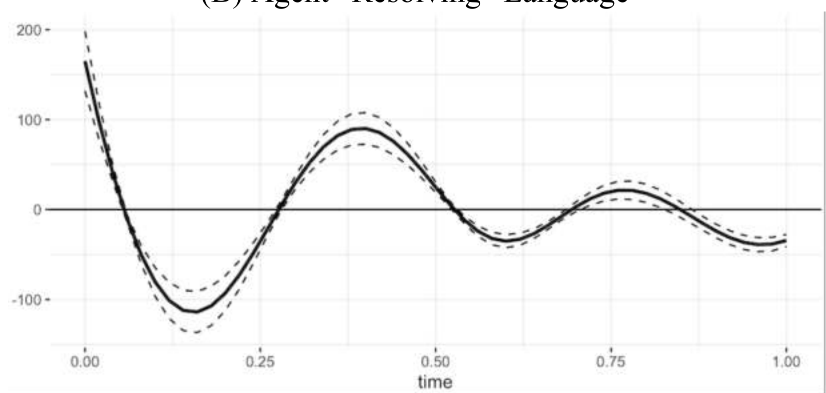
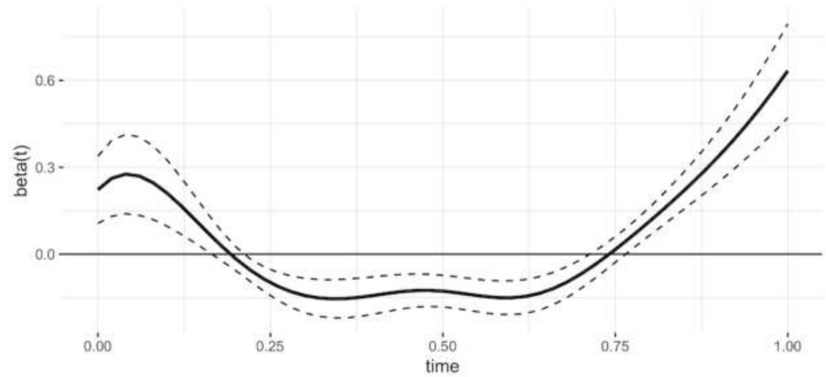


Figure A8: Beta Curves for Agent Positive Affective (A) and Negative Affective (B) Language in Relation to Customer Satisfaction

(A) Agent Positive Affective Language



(B) Agent Negative Affective Language

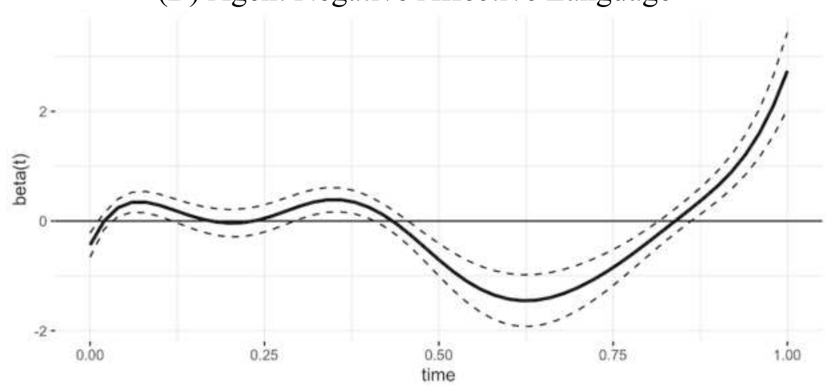
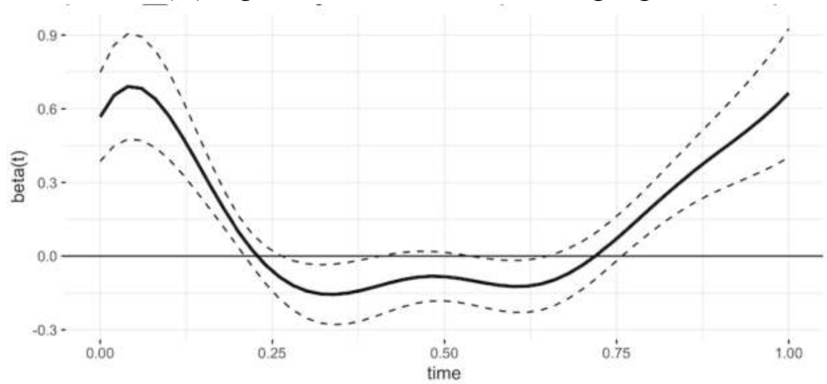


Figure A9: Beta Curves for Agent Positive Affective (A) and Negative Affective (B) Language in Relation to Order Count within 30 Days Post Interaction

(A) Agent Positive Affective Language



(B) Agent Negative Affective Language

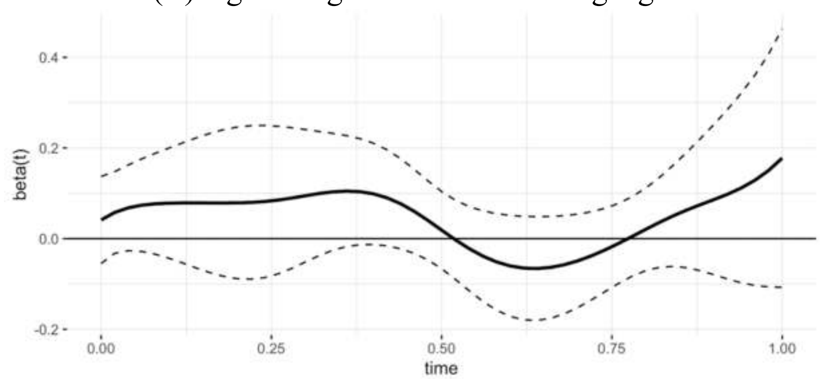


Table A1: Summary Statistics

	Mean	SD	Min	Median	Max
Independent Measures					
Agent Affective Language	22.74	27.42	0.00	11.11	100.00
Agent Cognitive Language	16.03	14.79	0.00	12.50	100.00
Dependent Measures					
Customer Satisfaction	3.34	1.61	1.00	3.00	4.00
Orders 30 Days Post	0.76	1.76	0.00	0.00	23.00
Controls					
Order	0.27	0.44	0.00	0.00	1.00
Shipping	0.27	0.45	0.00	0.00	1.00
Return	0.38	0.49	0.00	0.00	1.00
Product	0.05	0.22	0.00	0.00	1.00
Topic 1	0.09	0.06	0.02	0.07	0.41
Topic 2	0.07	0.06	0.01	0.04	0.35
Topic 3	0.08	0.05	0.02	0.07	0.45
Topic 4	0.08	0.07	0.02	0.07	0.60
Topic 5	0.07	0.06	0.01	0.05	0.45
Topic 6	0.09	0.10	0.02	0.06	0.61
Topic 7	0.07	0.06	0.01	0.05	0.44
Topic 8	0.08	0.05	0.01	0.07	0.28
Topic 9	0.07	0.05	0.01	0.06	0.30
Topic 10	0.07	0.04	0.01	0.06	0.38
Topic 11	0.08	0.05	0.02	0.06	0.28
Topic 12	0.08	0.07	0.02	0.06	0.58
Topic 13	0.09	0.05	0.01	0.07	0.29
Severity	2.61	0.94	1.00	2.50	5.00
Length	1082.03	853.54	112.00	854.00	4385.00
Resolved	0.80	0.40	0.00	1.00	1.00
Agent Tenure	412.38	650.85	0.00	216.00	3880.00
Agent Calls	4160.34	2456.80	37.00	4072.00	15010.00
Agent Female	0.61	0.49	0.00	1.00	1.00
Agent Social	12.35	16.85	0.00	8.57	100.00
Agent Perception	2.07	6.30	0.00	0.00	100.00
Agent Drive	6.48	10.73	0.00	0.00	100.00
Agent Time	17.10	15.06	0.00	17.39	100.00
Agent Informal	18.58	31.67	0.00	5.56	100.00
Agent Pitch	89.00	5.80	0.00	89.22	115.42
Agent Intensity	65.35	6.73	0.00	66.25	80.72
Customer Tenure	2177.19	1172.09	0.00	2123.00	4718.00
Customer LTV	6433.80	14600.02	68.00	2177.33	119762.85
Customer Region S	0.13	0.34	0.00	0.00	1.00
Customer Region E	0.36	0.48	0.00	0.00	1.00
Customer Region W	0.28	0.45	0.00	0.00	1.00
Customer Region MW	0.13	0.33	0.00	0.00	1.00
Customer Region OTHR	0.10	0.30	0.00	0.00	1.00
Customer Female	0.81	0.39	0.00	1.00	1.00

Att_Web	3.67	1.58	1.00	4.00	5.00
Att_Shop	3.47	1.71	1.00	4.00	5.00
Customer Affective Language	22.96	27.61	0.00	18.57	100.00
Customer Cognitive Language	21.51	19.79	0.00	16.67	100.00
Customer Social	7.88	16.00	0.00	0.00	100.00
Customer Perception	1.39	6.40	0.00	0.00	100.00
Customer Drive	4.85	13.28	0.00	0.00	100.00
Customer Time	14.79	17.16	0.00	12.50	100.00
Customer Informal	27.89	39.30	0.00	5.56	100.00
Customer Pitch	90.58	6.79	0.00	90.81	112.31
Customer Intensity	64.94	11.02	0.00	66.91	84.96
Orders 30 Days Pre	1.30	1.71	0.00	1.00	18.00

Table A2: Parameter Estimates for Customer Satisfaction after Group-Lasso

	Estimate	SE	p-stat
(Intercept)	1.18	0.39	0.003
affect_A_1	0.02	0.01	0.028
affect_A_2	-0.01	0.04	0.721
affect_A_3	0.15	0.07	0.031
affect_A_4	0.06	0.16	0.691
affect_A_5	1.42	3.28	0.666
affect_A_6	0.65	1.17	0.579
cognition_A_1	0.11	0.05	0.025
cognition_A_2	0.21	0.15	0.176
cognition_A_3	0.65	0.22	0.004
cognition_A_4	-1.29	0.99	0.197
cognition_A_5	7.19	7.45	0.336
cognition_A_6	2.64	3.70	0.476
cognition_C_1	0.01	0.03	0.753
cognition_C_2	0.03	0.05	0.488
cognition_C_3	-0.20	0.16	0.218
cognition_C_4	0.20	0.25	0.413
cognition_C_5	0.34	0.81	0.678
pitch_A_1	0.01	0.02	0.649
pitch_A_2	0.25	0.23	0.276
pitch_A_3	-0.73	0.45	0.104
pitch_A_4	-2.03	0.92	0.029
pitch_A_5	0.74	1.75	0.674
percept_C_1	0.00	0.03	0.940
percept_C_2	-0.08	0.10	0.435
percept_C_3	0.01	0.15	0.964
percept_C_4	2.83	2.57	0.274
percept_C_5	0.92	11.90	0.939
percept_C_6	-6.94	3.89	0.076
time_C_1	0.10	0.05	0.043

time_C_2	0.25	0.10	0.009
time_C_3	0.25	0.32	0.430
time_C_4	0.11	0.72	0.876
time_C_5	0.09	1.51	0.951
pitch_C_1	-0.04	0.02	0.057
pitch_C_2	-0.05	0.28	0.874
pitch_C_3	0.16	0.85	0.849
pitch_C_4	-1.29	1.61	0.423
pitch_C_5	-1.63	0.81	0.046
intensity_C_1	-0.04	0.02	0.008
intensity_C_2	0.07	0.05	0.205
intensity_C_3	0.21	0.14	0.124
intensity_C_4	-0.07	0.38	0.850
intensity_C_5	0.14	0.48	0.770
intensity_C_6	4.83	2.25	0.033
intensity_C_7	6.20	5.93	0.297
Topic1	2.97	1.42	0.039
Topic2	-4.53	1.13	0.000
Topic4	-1.04	1.26	0.414
Topic7	-4.11	1.13	0.000
Topic9	2.43	1.41	0.086
Topic11	1.43	1.38	0.301
Agent Tenure	0.00	0.00	0.233
Att_Web	0.19	0.06	0.001
Att_Shopping	0.46	0.05	0.000

Table A3: Parameter Estimates for Customer Purchase after Group-Lasso

	Estimate	SE	p-stat
(Intercept)	-0.97	0.13	0.000
affect_A_1	0.08	0.04	0.042
affect_A_2	0.01	0.05	0.845
affect_A_3	0.24	0.10	0.019
affect_A_4	-0.41	0.27	0.140
affect_A_5	-0.98	4.90	0.841
affect_A_6	-2.40	1.94	0.215
cognition_A_1	0.17	0.08	0.037
cognition_A_2	-0.18	0.26	0.490
cognition_A_3	0.59	0.36	0.105
cognition_A_4	-1.53	1.58	0.335
cognition_A_5	-2.49	1.26	0.049
cognition_A_6	-0.29	6.11	0.961
Orders 30 Pre	0.24	0.02	0.000

Table A4: Call-Level Linear Regression for Customer Satisfaction after Lasso

	Estimate	SE	p-stat
(Intercept)	0.50	0.47	0.29
Agent Affective Language	0.05	0.03	0.04
Agent Cognitive Language	-0.04	0.03	0.15
Topic 1	2.67	1.34	0.05
Topic 2	-3.99	1.17	0.00
Topic 7	-2.40	1.08	0.03
Cust. Region MW	-0.37	0.18	0.04
Att_Web	0.24	0.06	0.00
Att_Shop	0.46	0.05	0.00
Cust. Perception	0.12	0.05	0.01
Cust. Informal	0.03	0.01	0.05

Table A5: Call Level Poisson Regression for Customer Purchases after Lasso

	Estimate	SE	p-stat
(Intercept)	-0.07	0.47	0.89
Agent Affective Language	-0.08	0.04	0.05
Agent Cognitive Language	-0.02	0.04	0.54
Orders 30 Pre	0.21	0.01	0.00

Experimental Stimuli

Imagine you called customer service at Shopsite, an online retailer, and this was the conversation you had with a service agent:

Agent: Hi. [What might I do for you / How can I help you] today?

You: I can’t figure out how to get the free shipping.

Agent: I think I can find a solution. I know it can be a little complex to locate. I’ll explain where… scroll down a bit. See the dropdown menu at the bottom right?

You: Uh… ok. I got it.

Agent: I trust everything is [fixed / OK] then?

You: Yes, thank you. Bye now.

Simulations

Model comparisons presented in the main paper suggest our approach to capturing conversational dynamics enhances the predictive benefit of understanding when affective or cognitive language is beneficial. But one might still wonder how the model's dynamic recommendations (i.e., using more affective language at start and end, and cognitive language in the middle) should perform relative to the exclusively cognitive or affective approaches recommended in prior research. Similarly, one could ask how much *when* one uses each of these language styles matters if both affective and cognitive language are used in a single interaction.

To begin to answer these questions, we performed a series of simulations. Because our model identifies when affective and cognitive language should be used, but not the optimal level of these features at a given moment, the simulations utilize the average observed levels of affective and/or cognitive language at each conversational moment, and then turn that language feature "on" or "off" at different moments based on the simulation condition. We caution that these simulations compare alternative approaches to the dynamic language use suggested by our modeling estimates. Consequently, the simulated improvements in satisfaction and purchases should be considered optimistic ceilings rather than expected outcomes.

First, we compare the current approach to the marketing literature's recommendation to be competence-oriented throughout the interaction. The simulation suggests that employees who follow the timing of affective and cognitive language suggested in the current approach (Figures 3 and 4) would see a 2.50 point increase in customer satisfaction ($p < 0.01$) and 3.42 more purchases in the 30 days following the call ($p < 0.01$) over this simulated competence-only baseline. For a more conservative test, we also compare our approach to a competence-only approach that uses cognitive language only, but emphasized at the "right times" (per Figures 3 and 4). Results further support the notion that using both affective and cognitive language at the right times, rather than only cognitive language at the right times, should have beneficial effects, i.e., difference in customer satisfaction $= 2.06$ ($p < 0.01$) and in purchases $= 2.84$ ($p < 0.01$).

Results are similar when we compare the current approach to the psychology literature's suggestion to be affective (or warm) throughout the interaction, i.e., difference in customer satisfaction $= 2.42$ ($p < 0.01$) and in purchases $= 3.69$ ($p < 0.01$). A comparison to being affective only but at the "right times" shows similar results, i.e., difference in customer satisfaction $= 1.36$ ($p < 0.01$) and in purchases $= 1.87$ ($p < 0.01$).

Second, we consider a comparison which acknowledges that affective and cognitive language can fruitfully co-exist in a single interaction but ignores the possibility that *when* these speaking styles are used matters. To do so, we simulate a scenario in which the two speaking styles are turned on at the mean observed level at every point in conversational time. Speaking both affective and cognitively at the "right times" rather than at all times results in a simulated improvement of 1.49 points in customer satisfaction ($p < 0.05$) and an incremental 2.39 purchases in the 30 days after the call ($p < 0.05$).

Taken together, while the size of the results should be considered ceilings rather than expected values, they support the benefits of using *both* affective and cognitive language rather than only one, and of considering *when* to use each of these approaches over the course of a conversation.

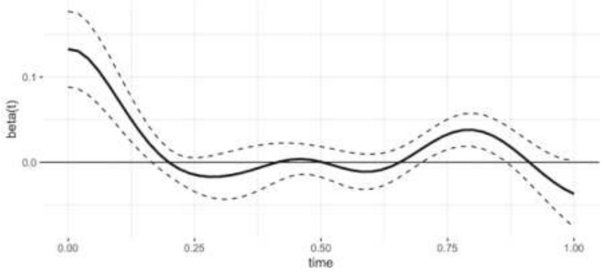
Moderation of Conversation Dynamics by Situated Factors

On top of the main results, one could further ask how situated features moderate the temporal importance of language features (Zhang et al. 2020). While our results accounted for over 50 such features, including both dynamic and static controls, the dynamics of language features could shift due to other situated factors. For instance, our model considers the severity of the customer's issue and conversational features that might be linked to this (i.e., pitch and intensity of customer voice), but agent affective or cognitive language may become particularly important when customers seek resolution of a more severe, difficult issue. To this end, we used the judged severity measure (two independent judges, 1 = Not at all severe, 7 = Very severe; $r = .57$) to perform a median split of our data and run our model separately on both data segments following the split. We find that for difficult issues, cognitive language is more important overall, while affective language becomes less important at the end of the call (Figure A10). Competently solving difficult issues may be more important than rapport building in this case. In contrast, more mundane service interactions may benefit most from a more personable, affective engagement approach, especially at the conversation's start.

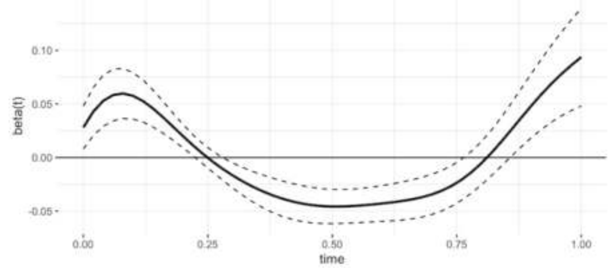
Figure A10: Beta Curves for Agent (A) Affective and (B) Cognitive Language in Relation to Customer Satisfaction Moderated by Call Severity

(A) Agent Affective Language

Low Severity

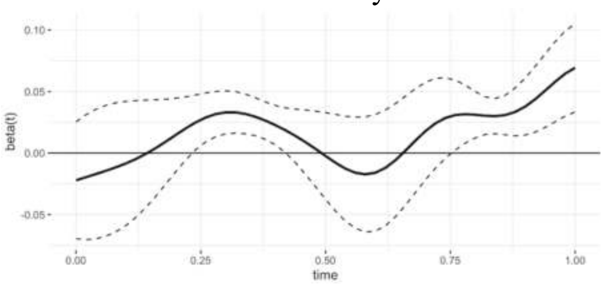


High Severity



(B) Agent Cognitive Language

Low Severity



High Severity