What Leads to Longer Word of Mouth Discussion?

REIHANE BOGHRATI AND JONAH BERGER

ABSTRACT While a great deal of research has demonstrated word of mouth's impact, less is known about the broader discussions in which such sharing takes place. Some things are talked about for longer, and longer discussion tends to increase persuasion, but how might the language consumers use to talk about something shape discussion length? Natural language processing of hundreds of conversations begins to address this question. Results indicate that discussions last longer when conversation partners speak more concretely and asking narrow (but not broad) questions. These results hold even controlling for who is talking, what they are talking about, and a variety of other aspects of the discussion. Overall, the findings deepen understanding of word of mouth, shed light on linguistic drivers of conversation, and highlight how natural language processing can provide insight into consumer behavior.

Consumers have dozens of conversations each day. They talk about products they bought, services they use, and experiences they enjoyed. They text friends, call family, and chat face to face with their colleagues. Furthermore, such conversations have a huge impact on consumer behavior. Word of mouth shapes what consumers think, do, and buy (e.g., Chevalier and Mayzlin 2006; De Angelis et al. 2017; see Moore and Lafreniere [2020] for a review), and consistent with this impact, organizations invest significant resources in creating and managing word of mouth.

But while word of mouth is clearly valuable, there has been less attention to the broader conversations in which word of mouth is situated. Some discussions about products, services, and ideas last longer than others. Why?

While a burgeoning stream of research has begun to examine behavioral drivers of word of mouth (see Berger [2014] for a review), there has been less attention to actual back-and-forth discussions (see Villarroel Ordenes et al. 2019). Most work simply examines whether or not a product was mentioned (e.g., Berger and Schwartz 2011) or how the presence of mentions shapes sales (Godes and Mayzlin 2009).

Not surprisingly, though, discussion length matters. Longer discussions can provide more information, reasons, or details, all of which should increase word of mouth's effect. Indeed, a great deal of research demonstrates that longer discussions are more persuasive (Zhang et al. 2010) and helpful and boost purchase (Ghose et al. 2012; Kim et al. 2018) and retention (Irvine et al. 2023). Additional data we collected are also consistent with this notion (see study 1 and pilot studies 1 and 2 in the appendix).[1] So why are some products and services talked about for longer than others?

Natural language processing of hundreds of conversations using machine learning and other techniques begins to address this question. Concreteness is one of the most studied language features in marketing (e.g., Schellekens et al. 2010, 2013; De Angelis et al. 2017; Aerts and Verlegh 2018; Packard and Berger 2021), and questions are an integral part

Reihane Boghrati (corresponding author: rboghrat@asu.edu) is an assistant professor of Information Systems at the W. P. Carey School of Business, Arizona State University, Tempe, AZ 82581, USA. Jonah Berger (jberger@wharton.upenn.edu) is an associate professor of Marketing at the Wharton School, University of Pennsylvania, Philadelphia, PA 19104, USA. The authors thank Christophe Van den Bulte for his generous guidance on modeling approach, and Nicole Weinsten, Judy Liu, Allen Zhu, George Ferridge, Ryan Bush, Isabella Cruz, Ruchi Sharma, Kelsey Warren, Theodore Damiani, Joseph Arias, Hyo Kim, Seungmin Suh, Logan Shalit, Jonathan Goldschlag, Matthew Karkafi, William Chan, Noah Katz, and Emily Xi for their helpful research assistance. The authors also thank Wharton Behavioral Lab for their help with data collection.

1. We don't mean to suggest that longer discussions always have more impact. One piece of useful information, for example, can have greater impact than five pieces of useless information. But, on average, longer discussions create the opportunity for more information, reasons, or detail, all of which should increase word of mouth's impact.

Issue Editors: Ann Kronrod, Sarah Moore, and Peeter Verlegh

Published online March 10, 2026.

Journal of the Association for Consumer Research, volume 11, number 2, April 2026. Published by The University of Chicago Press for the Association for Consumer Research. https://doi.org/10.1086/740059

of almost every conversation. Consequently, we examined both these aspects. Results indicate that speaking more concretely and asking certain types of questions (i.e., narrow rather than broad), seem to encourage longer discussion.

This work makes three main contributions. First, we deepen understanding around linguistic drivers of word of mouth. As noted, while a burgeoning stream of research has begun to examine *whether* people talk about something, less work has studied the consequential outcome of *discussion persistence*, or how long things are discussed. We examine how language shapes discussion length and suggest directions for future research.

Second, we highlight how natural language processing and machine learning can shed light on consumer behavior. By integrating traditional approaches (e.g., dictionaries), with more complex ones (e.g., deep learning) we showcase how they can work together to deepen insight. We also improve the identification of question breadth by building an automated classifier available for other researchers to encourage work in the area.

Third, from a substantive perspective, these findings can help marketplace actors generate word of mouth. Everyone from brand managers and public health officials to campaign managers and activists want to harness word of mouth's power. Our results highlight simple linguistic shifts that may encourage longer discussion.

DRIVERS OF WORD OF MOUTH

There has recently been a great deal of interest in drivers of word of mouth. Spurred, in part, by the emergence of social media and word of mouth's impact on sales (Chevalier and Mayzlin 2006; Godes and Mayzlin 2009), researchers have started to examine why some products or brands are more likely to be mentioned.

Some work has focused on attributes of the product, service, or brand being discussed. Things get talked about more if they are original (Moldovan et al. 2011), for example, more publicly visible (Berger and Schwartz 2011), or evoke more emotion (Berger and Milkman 2012). Other work has examined how features of the audience shape word of mouth. Consumers tend to share different things with smaller versus larger groups (Barasch and Berger 2014), for example, and interpersonal closeness impacts word of mouth valence (Dubois et al. 2016). But while this literature has provided important insights into why people mention and share some things rather than others, there has been less attention to actual back-and-forth conversations.

CONSUMER CONVERSATION

Conversations are social interactions in which multiple people act in coordinated ways to reach shared understanding (Clark and Schaefer 1989). Consumers converse to seek and share information, manage impressions, and strengthen social bonds, among other reasons (e.g., Dunbar 1996; Baumeister et al. 2004).

Empirical work has started to examine the consequences of conversations (Yeomans et al. 2023; Di Stasi et al. 2024; Brooks 2025). Work on customer service interactions, for example, finds that using certain pronouns (Packard et al. 2018) or language patterns (Villarroel Ordenes et al. 2019) can increase customer satisfaction.

But while this work has provided important insight into the effects of conversation, the underlying causes are less clear. Why are some things talked about longer than others? We suggest linguistic concreteness and asking questions both play an important role.

LINGUISTIC CONCRETENESS AND DISCUSSION LENGTH

Concreteness is one of the most studied language features in marketing (Schellekens et al. 2010, 2013; De Angelis et al. 2017; Packard and Berger 2021; Berger et al. 2023) and refers to the degree to which something is real or tangible and perceivable to the mind or eye (i.e., vivid or imaginable; Semin and Fiedler 1988; Brysbaert et al. 2014). Concrete language arises from immediate sensory experience (e.g., things one can touch, feel, or see, like shoes or a table), while abstract language refers to intangible qualities, concepts, or ideas (e.g., love or truth; Langacker 1987; Hansen and Wänke 2010). Furthermore, marketing research has examined when consumers use more abstract language (e.g., when describing experiences that are congruent with their attitudes; Schellekens et al. 2010), and how concreteness impacts referral persuasiveness (De Angelis et al. 2017), and customer satisfaction (i.e., customers are more satisfied when employees talk concretely; Packard and Berger 2021).

Particular topics (e.g., buying shoes) are more concrete than others (e.g., love), but even when talking about the same topic, consumers can use more or less concrete language. Someone can say that they had "dinner" or "chicken teriyaki," that they "got" or "grabbed" it, and that it was "really" or "mouthwateringly" good. In all cases, the latter uses more concrete language to talk about the same thing. But how might linguistic concreteness shape discussion length?

We suggest that concrete language should encourage discussion. Consistent with the fact that they refer to real or perceptible entities, concrete things are easier to visualize or imagine and require less cognitive resources to process (Connell and Lynott 2012). It's easier to picture chicken teriyaki, for example, than dinner. This, in turn, can impact interest and comprehension. Consistent with this notion, work in education and learning finds that concrete texts are more comprehensible and interesting (Sadoski et al. 2000). Furthermore, concrete language can also hold attention (Berger et al. 2023) and make people feel like someone else is listening to them (Packard and Berger 2021), which may also encourage further discussion. Beyond concreteness, though, we also examine questions.

QUESTIONS AND DISCUSSION LENGTH

Questions play an integral role in almost every conversation. They facilitate information exchange, build rapport, and shape value capture (Huang et al. 2017; Brooks and John 2018; Di Stasi et al. 2022). But how might they shape the length of discussion?

One could argue that any question should encourage discussion. A key function of questions is to solicit information (Chafe 1970; Kearsley 1976; Dillon 1982) and when someone asks a question, others often feel compelled to reply, which might encourage continued discussion. Asking questions also makes people seem responsive (Huang et al. 2017), which could encourage discussion. While people certainly want to talk about what they want to talk about, given conversations social nature, people also try to appraise how interested their partner is (Yeomans and Brooks 2020). If someone senses their partner is disinterested in a particular topic, they may switch to something else. Consequently, signals that a conversation partner is engaged in a particular topic (like questions) should encourage persistence. Questions can suggest someone understood, cares about, and is interested in what was said (Chen et al. 2010; Huang et al. 2017), which could encourage continued discussion.

Importantly, however, not all questions are the same. Prior work notes that questions vary in their breadth or openness (Miles 2013). Broad, or more open questions, broadly seek information without constraining the appropriate response. For example, the question "why did you go to Chili's last night?" could generate a range of possible answers. Consumers could say that they like the brand, that they didn't have food at home, or any number of other replies.

Narrow, or more closed questions, however, tend to suggest a narrower range of responses. The question "who did you eat dinner with last night?" for example, begs a more directed answer. The respondent could mention different people, but there is only so much variation in how they can respond or where they can go with the question.

We suggest that question breadth should shape discussion. One could argue that while broad questions should encourage discussion, narrow questions may not. After all, while broad questions encourage in depth responses that could open up many avenues (Kearsley 1976), narrow questions may be so specific that people run out of things to say. Consequently, narrow questions could lead people to stop talking about a particular topic.

In contrast, we suggest that while narrow questions should encourage discussion, broad questions may not. Narrow questions should encourage further progress down a particular line or direction. By asking a question, they not only provide fuel to further discussion but encourage discussion of the topic at hand. Furthermore, to the degree that they follow up an existing line of thought, they suggest that the questioner is interested in that line of thought and thus would be happy to continue discussing it (Huang et al. 2017).

Broad questions, in contrast, may be less likely to have the same effect. While they encourage a response, by calling for more complex or exploratory answers, they may encourage people to think about (or bring up) other conversation topics or directions, beyond the one being discussed. Responses to broad questions should have similar effects. Consequently, while people may not switch topics right away, broad questions may eventually encourage talking about something else.

THE CURRENT RESEARCH

In summary, we suggest that concreteness and question type should impact the length of discussion. Two studies test these possibilities using hundreds of real word of mouth conversations. Given that most word of mouth happens in offline, synchronous conversations (Keller and Fay 2012), we focus on these types of interactions.

The simplest way to analyze conversational features would be to examine how their presence relates to discussion length. Whether discussions that involve more narrow questions, for example, tend to last longer.

Examining at the discussion level, however, ignores when those particular language features occur. Two discussions could include the same number of narrow questions, for example, but one could include them earlier. Furthermore, if certain linguistic features truly encourage discussion, then not only should discussions that include them last longer, but these discussions should last longer *after* those discussion-encouraging features occur. Consequently, we measure linguistic features at the turn (rather than discussion) level and use hazard modeling to examine whether people are more likely to keep talking after key linguistic features occur.

STUDY 1

Study 1 examines 583 discussions about products, services, and brands between 192 participants (96 pairs). We test whether longer discussions about a topic are more persuasive, and, controlling for a range of other aspects, how the language used relates to discussion length.

Method

Participants ($N = 192$, 76% females, $M_{age} = 22.36$) completed a conversation study.[2] Initial recruitment focused on laboratory conversations, but given COVID disruptions, some participants had conversations over Zoom. As we discuss in the appendix, however, this factor did not affect the results. To ensure real word of mouth conversations, rather than just getting to know each other chitchat, participants signed up to do the study with someone they already knew relatively well. Each participant started by listing five brands (e.g., Adidas), products (e.g., ice cream), services, or consumption experiences that they thought their conversation partner didn't know much about.

To control for tie strength, we asked participants to rate how well they know their conversation partner (1 = *did not know at all*, 7 = *knew very well*). To control for how interesting participants found each topic, we also measured topic interest. For each topic listed by them or their conversation partner, participants rated how interested they would be in talking about it (1 = *not at all interested*, 7 = *extremely interested*).[3]

Then, participants were given 10 minutes to converse, during which they could discuss as many topics as they wanted. They were randomly assigned to one of the things they listed, and told they could talk about it for as long or short as they wanted before moving on to the next thing (see the appendix). Microphones recorded the discussion, and Amazon Web Services was used to convert the recordings into text. To ensure transcriptions were correct, research assistants listened to the recordings and fixed any errors.

Research assistants broke each conversation down into participants' preselected topics (e.g., Adidas). They read each conversation, marked turns where topics changed (e.g., where participants stopped talking about Adidas and started talking about coffee), and matched the preselected topics to those discussed in the transcripts. Any discussions where participants talked about the conversation itself (e.g., experimenter instructions) were removed, resulting in a data set of 583 discussion topics ($M_{\text{no. of turns}} = 9.33$, $SD_{\text{no. of turns}} = 7.12$) and almost 5,500 conversational turns.

A great deal of prior work (e.g., Zhang et al. 2010; Ghose et al. 2012; Kim et al. 2018), as well as pilot studies 1 and 2, demonstrate that longer discussion increases persuasion. To provide an additional test, after the conversation ended, participants provided how persuasive they found each topic's discussion (1 = *not at all*, 7 = *extremely*). To examine what persuades others, we focus on the ratings of the conversation partner who did not list the topic initially. We also measured how much they enjoyed discussing each topic (1 = *didn't enjoy at all*, 7 = *enjoyed a great deal*). Finally, we measured demographic information (i.e., age and gender).

Extracting Key Linguistic Features

Natural language processing was used to extract key linguistic features from each conversational turn.

Linguistic Concreteness. Following prior work (Packard and Berger 2021), we measure linguistic concreteness using a boot-strapped extension of the MRC Psycholinguistic Database (Paetzold and Specia 2016). This includes scores for over 85,000 English language words.[4] Averaging concreteness scores across the words in each turn provided a score for that turn.

Question Breadth. Building on prior work in computational linguistics (Stolcke et al. 2000), we determined which sentences contained a question by training a deep learning model (see Lee et al. [2018] and Liu et al. [2019] for recent applications in marketing).[5] This approach was also used to

2. Data are available upon request to researchers who agree to comply with the privacy restrictions set forth by our Institutional Review Board.

3. We collected topic interest before, rather than after the conversation, to avoid participants using discussion persistence as a cue for interest.

4. MRC Psycholinguistic Database was selected because of the larger coverage (83%) compared to alternative measures (74%).

5. While one could try to use question marks to identify questions, an initial read though suggested this would not be sufficiently accurate. A number of questions appeared without question marks at the end (e.g., "what if we went at like 1:00." or "what else is interesting."). Also, dialogue

extract other dialogue acts (e.g., statements; see Control Variables section for details).

For classification, we adopt an approach which uses both what is being said and who says it. Incorporating speaker embedding into utterance embedding creates a speaker-aware contextualized embeddings, which helps predict a dialogue act for each sentence. To represent utterance embedding, Robustly Optimized Bidirectional Encoder Representation for Transformers (Liu et al. 2019) was used, and for speaker embedding, we use a shallow neural network with a speaker embedding layer. Next, speaker embeddings are added to utterance embeddings to obtain speaker-aware utterance embedding and finally a fully connected layer is used for dialogue act classification (see the appendix for details). For training data, we use a standard dialogue act data set of manually coded phone conversations (Jurafsky et al. 1997). After fine-tuning the model with this data set, we apply it on our data to predict a dialogue act label per sentence. The model achieves 83% accuracy (He et al. 2021) similar to human accuracy of 84% (Stolcke et al. 2000).

While dialogue act prediction helps identify questions, it does not classify question breadth. After predicting questions using the automated classifier, we identified broad and narrow questions manually. Three research assistants were given a description of narrow and broad questions and coded each question in the data set as one or the other. Examples of narrow questions included things like "what is the scent of this candle?" and "Are you interested in climbing mountains yourself?;" examples of broad questions included things like "what do you think?" and "I don't get that at the land border, why they do that?" While we used manual ratings to ensure accuracy, to make it easier for future research to study question breadth, we built an automated classifier (see the appendix).

Control Variables

One could wonder whether any observed effects are driven by who is talking, what they are talking about, or other linguistic features. Consequently, we control for these aspects in a number of ways.

Who Is Talking. Rather than being driven by language, one could wonder the results are just picking up an effect of particular people. Maybe certain people tend to use particular linguistic features, for example, and also have longer conversations. Alternatively, maybe the pairing between conversation partners is driving the effect.

We address these possibilities a few ways. First, we control for aspects of each individual (e.g., age). Second, we control for features of the conversation partners, such as gender match (i.e., both men, both women, or mixed gender), how well they know each other (i.e., tie strength), and to control for unobserved aspects of the pairing, conversation random effects.

What Is Being Talked About. Beyond who is conversing, one could wonder whether the results are driven by what is being discussed. Certain discussion topics (e.g., tech products) may tend to be discussed for longer than others (e.g., toilet paper), and could potentially involve different linguistic features. Alternatively, maybe conversation partners use different language for topics they find personally more or less interesting.

We address these possibilities a few ways. First, to control for the topic of discussion, two research assistants categorized discussion topics into higher-level categories such as cosmetics or technology.[6]

Second, we control for how interesting participants said they found each topic. The main model uses average interest between the partners, but other ways of controlling for interest (i.e., difference in interest between the partners, maximum interest, minimum interest, or only considering topics above a threshold) show similar effects.

Third, one might wonder whether discussion order might impact discussion length. Things that are talked about early on might be talked about for a while, but as conversation partners get tired, later things may get cut short. The topics were randomly ordered, but to further control for this, we also control for topic order.

How People Talk About It. One could also wonder whether the results are driven by some other aspect of content. Maybe asking questions, for example, is correlated with another linguistic feature (e.g., words per turn) and that correlated feature, rather than questions themselves, is driving the result. Consequently, following prior work (Berger and Packard 2018) we control for word count, six-letter words (as a control for writing complexity), and baskets of words empirically linked to social or psychological constructs

act classification allows removal of questions that signal nonunderstanding (e.g., "excuse me?") or that just open discussion (e.g., "how are you?") which do not seek broad or narrow information.

6. Because dyads specified the topics they talked about, it was more accurate to do it manually instead of using an approach like topic modeling.

(e.g., affective processes and cognitive processes) from linguistic inquiry and word count (LIWC) at the turn level.

One might also wonder whether the results are driven by linguistic mimicry (Moore and McFerran 2017). To rule out this possibility, for each discussion, we calculated linguistic style matching (LSM) between conversation partners (Ireland and Pennebaker 2010) and included it as a control.

We also control for the other dialogue acts (e.g., statements or acknowledgments) in each conversational turn. Building upon prior work (Jurafsky et al. 1997), the method used to identify questions (i.e., He et al. 2021) was used to identify other dialogue acts. The data have 42 different labels, so chance accuracy is slightly over 2%, but we achieve 83%, similar to human accuracy (84%; Stolcke et al. 2000). Since dialogue acts are defined per sentence, we predict dialogue act tags for each sentence, in each turn. Consequently, each turn can have multiple tags (e.g., an agreement for its first sentence and a question for the second sentence).

Finally, we also control for discussion emotionality. To rule out this possibility, for each discussion, we calculated language emotionality using LIWC and included it as a control.

Modeling Approach

Given our interest in predicting how long it takes for an event to occur (i.e., the end of discussion) based on time-varying explanatory variables (e.g., linguistic features for a given turn), we used hazard modeling. Hazard models relate the time that passes before an event occurs to variables that may be associated with that quantity of time. A central concept of these types of models is *hazard rate*, which is the probability of an event happening at current time, given it has not happened yet. In our case, hazard is the probability of discussion ending on a given turn.

Key features of the data informed what specific approach to use. First, some of key predictors are time-varying. Linguistic concreteness, for example, varies across the course of the conversation. Second, discussion ends at a particular turn. Given these facets, a time-varying discrete-time hazard model seemed most appropriate. These models are used when explanatory variables (such as language features) are varying by time and time is discrete (Allison 1982).

The model was estimated using logistic regression (Allison 1982). Concreteness scores and questions, which vary at the turn level are used to predict the end of discussion. Control variables entered the model at the turn level (e.g., other dialogue acts and emotionality), or discussion level (e.g., linguistic style matching and topic interest), depending on when they occurred. To control for the fact that pairs of participants talk about different topics and may tend to talk for longer or shorter amounts, conversation ID was included as a random effect. Time may impact discussion persistence, so we control for turn number. To avoid assumptions about the functional form between time and discussion persistence, time (i.e., turn number) enters the model nonparametrically. This allows each period to get a separate coefficient and is generally a more recommended approach (Allison 2014). Given the volume of data, however, it was not possible to give every single turn a separate coefficient, so turns were grouped into five-turn buckets. That is, if a discussion has 100 turns, there are 20 coefficients, each belongs to a five-turn bucket (results are robust to other approaches; see table A2 models 4–5).

Formally, we denote the hazard rate by h_{it}, the probability that discussion i ends at turn t, given the discussion has not yet ended. Since h_{it} is a probability, it can only have numbers between 0 and 1, while the explanatory variables have a larger range of numbers. Therefore, by taking the logit transformation of h_{it} the model is defined as:

$$logit(h_{it}) = \alpha(0) + \sum_{j=1}^{n} \alpha(j) f_{it}(j) + \delta_t + \gamma_i$$

$$\gamma_i = N(0, \sigma^2)$$

Where n is number of linguistic features in the model and f_{it} denotes a linguistic feature value for discussion i in turn t. In the main models, turn enters the model nonparametrically, but results are also robust to them entering parametrically. Thus δ_t is a turn-specific fixed effect allowing the baseline hazard to vary freely across turns. Further, participants pairs are controlled for by random effect γ_i for conversation IDs. The model with controls is estimated similarly.

Given the interest in understanding what drives longer discussion, to ease interpretability, we flip the signs of the coefficients so that positive coefficients indicate a variable is linked to longer discussion.

Results

Persuasion. First, we examined the relationship between discussion length and persuasion. As predicted, and consistent with prior work (Zhang et al. 2010; Ghose et al. 2012; Kim et al. 2018) longer discussions about a product, service, or brand were more persuasive ($\beta = 0.045$, $SE = 0.011$, $p < .001$). Longer discussions were also more enjoyable ($\beta = 0.068$, $SE = 0.009$, $p < .001$).

Female to Female. Second, we examined concreteness and questions. As predicted, discussion lasted longer when people used more concrete language ($\beta = 0.153$, SE = 0.060, $p = .011$, OR = 1.165, table 1, model 2). Discussion also lasted longer when people asked questions ($\beta = 0.157$, SE = 0.058, $p = .007$, OR = 1.17, table 1, model 1),[7] but the effect depended on the type of question asked. Discussion lasted longer when consumers asked narrow questions ($\beta = 0.155$, SE = 0.058, $p = .007$, OR = 1.168), but broad questions did not have the same effect ($\beta = 0.033$, SE = 0.055, $p = .547$, OR = 1.034, table 1, model 2).

Robustness Checks. To test generalizability, and explore alternative explanations, we conducted several robustness checks, including (1) testing reverse causality, (2) using fixed effects for conversation ID, (3) seeing whether results differed between Zoom and face to face conversations, and examining whether the results could be driven by (4) the time limit imposed, (5) short discussions, (6) the modeling approach used, (7) or how time is modeled. In all cases, results remain the same (see appendix).

Discussion

Results of study 1 provide initial support for our theorizing. Discussions lasted longer when consumers spoke concretely and asked certain types of questions (i.e., narrow questions). The results also demonstrate discussion length matters. Consistent with prior work (Zhang et al. 2010; Ghose et al. 2012; Kim et al. 2018), longer discussions were more persuasive.

We also examined whether the key variables might interact. Whether concreteness might be more important for more interesting topics, for example, or when language is more emotional. There were no significant interactions, but see Appendix for more detail.

STUDY 2

While the results of study 1 are consistent with our theorizing, one could wonder whether they are somehow driven by the fact that participants had to list conversation topics ahead of time. Consequently, to ensure that didn't drive the results, study 2 uses an even more naturalistic setting. Participants could share word of mouth about whatever they wanted, and we examined whether discussion lasted longer when consumers used more concrete language and asked narrow questions.

Method

Participants ($N = 222$, 72% female, $M_{\text{age}} = 21.2$) had word of mouth conversations in the lab. As in study 1, to ensure real conversations, rather than just getting to know you chitchat between strangers, participants signed up for the study with someone they already knew.

Participants were given 10 minutes to talk about anything they wanted. They talked about brands like Starbucks and Trader Joe's, products like shoes and wine, services like TikTok and insurance, and experiences like weight lifting and boating. A handful of participants were part of multiple sessions, and 125 conversations were collected.

Microphones recorded the discussion, and a professional transcription service manually converted the recordings to text. As in study 1, each conversational turn was treated as a separate record. Research assistants then broke each conversation down into chunks based on the topic discussed (e.g., Starbucks or shoes).

As in study 1, discussion length was defined as the number of turns in each discussion. Any discussions where participants talked about the conversation itself (e.g., experimenter instructions) were removed, resulting in a data set of 745 discussion topics ($M_{\text{no. of turns}} = 22.14$; $SD_{\text{no. of turns}} = 21.68$) and more than 16,000 conversational turns.[8]

As in study 1, we then used natural language processing to extract key features and the same hazard model analyzed the drivers of discussions.

Results

What Drives Discussion Length?. Even using a more naturalistic setting, where participants did not pre-define their topics and could share word of mouth about whatever they wanted for as long as they wanted, results were similar to study 1 (table 2 and table A3). As predicted, discussion lasted longer when concrete language ($\beta = 0.105$, SE = 0.047, $p = .025$, OR = 1.111), and narrow questions were used ($\beta = 0.414$, SE = 0.073, $p < .001$, OR = 1.513). Broad questions

7. Note that this is not simply a result of questions requiring a response and thus the topic lasting longer. The average distance between a questions and topic death was 8 turns.

8. Participants did not list their topics prior to the conversations, so we were not able to collect their interest per topic (i.e., discussion) in the same session. That said, we followed up with participants after the study and asked them to rate how interested they would be in talking about different topics. Many participants did not complete the follow up survey, but results remain similar even with this smaller sample (see theappendix for details).

Table 1. Study 1 Results for Language and Discussion Length

Model Variables	Main (1)	Question Breadth (2)
Key independent variables:		
Linguistic concreteness	.154*	.153*
	(.06)	(.06)
Questions	.157**	
	(.058)	
Narrow questions		.155**
		(.058)
Broad questions		.033
		(.055)
Controls:		
Who is talking:		
Average age	.111	.111
	(.083)	(.083)
Female to female	.08	.076
	(.198)	(.198)
Male to male	−.01	−.011
	(.354)	(.354)
Tie strength	.015	.015
	(.087)	(.087)
Conversation ID	Yes	Yes
What is being discussed:		
Average interest	.133*	.133*
	(.061)	(.061)
Topic order	−.094	−.094
	(.053)	(.053)
Topics	Yes	Yes
Other language controls:		
Word count	.004	.005
	(.095)	(.095)
Six letters	.082	.082
	(.054)	(.054)
Linguistic style matching	.658***	.659***
	(.053)	(.053)
Positive emotion	1.029	1.04
	(1.19)	(1.19)
Negative emotion	.283	.286
	(.299)	(.299)
Other LIWC features	Yes	Yes
Other dialogue acts	Yes	Yes
Nonparametric 4turn	Yes	Yes
Performance measure:		
AIC	3313.136	3314.792
BIC	3814.866	3823.125
Log. loss	.276	.276

Note.—AIC = Akaike information criterion; BIC = Bayesian information criterion; LIWC = linguistic inquiry and word count.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

Table 2. Study 2 Results for Language and Discussion Length

Model Variables	Main (1)	Question Breadth (2)
Key independent variables:		
Linguistic concreteness	.107*	.105*
	(.047)	(.047)
Questions	.345***	
	(.065)	
Narrow questions		.414***
		(.073)
Broad questions		.054
		(.05)
Controls:		
Who is talking		
Average age	−.011	−.012
	(.047)	(.047)
Conversation ID	yes	yes
What is being discussed:		
Topic order	−.1*	−.099*
	(.041)	(.042)
Other language controls:		
Word count	.094	.091
	(.078)	(.078)
Six letters	.077	.073
	(.044)	(.044)
Linguistic style matching	.814***	.816***
	(.044)	(.046)
Positive emotion	−1.558	−1.611
	(2.088)	(2.101)
Negative emotion	−.749	−.772
	(.94)	(.946)
Other LIWC features	yes	yes
Other dialogue acts	Yes	Yes
Nonparametric Turn	Yes	Yes
Performance measure:		
AIC	5463.303	5456.124
BIC	5888.067	5888.612
Log. loss	.158	.158

Note.—AIC = Akaike information criterion; BIC = Bayesian information criterion; LIWC = linguistic inquiry and word count.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

did not have the same effect ($\beta = 0.054$, SE = 0.050, $p = .282$, OR = 1.055, table 2, model 2).

Robustness Checks. To test the generalizability of the effects, and explore alternative explanations, we conducted the same robustness checks as study 1 (see the appendix).

Discussion

Study 2 underscores the findings of study 1. Even in a more naturalistic setting, where participants could talk about whatever they wanted and did not have to prespecify topics, results were similar. The discussion lasted longer when consumers used more concrete language or asked narrow questions. Broad questions did not have the same beneficial effect.

GENERAL DISCUSSION

Consumers talk about dozens of products and services each day, and such word of mouth has an important impact on consumer behavior. But while it is clear that word of mouth is frequent, and important, less is known about the broader context in which such interpersonal transmission is situated. Why are some products, services, or brands discussed for longer than others?

Natural language processing of hundreds of everyday word of mouth conversations begins to shed light on this question. Specifically, they suggest that conversations last longer when people speak more concretely and ask narrow (but not broad) questions.

Contributions and Implications

These findings make several contributions. First, while what is being talked about, or who is talking is certainly important, our results suggest that how things are talked about, or the specific linguistic features used, also plays an important role. Specifically, beyond how interesting the discussion topic is, or how well conversation partners know each other, using more concrete language or asking narrow questions seems to encourage longer discussion.

Second, from a substantive perspective, these findings should be useful in shaping discussion. Similar to prior work (Zhang et al. 2010; Ghose et al. 2012; Kim et al. 2018), our results indicate that longer discussions are more persuasive. Consequently, understanding how to encourage longer discussion is key. Our results suggest specific approaches that may be helpful. Specifically, speaking concretely and asking the right types of questions, are linked to longer discussions. Rather than speaking abstractly, for example, using more concrete language to make it easier for others to understand what one is talking about. Similarly, *who*, *where*, or *which* questions are almost always narrow in scope and thus may encourage longer discussion.

Third, we highlight how computational linguistic techniques can shed light on consumer behavior more broadly (Berger et al. 2020). While consumer researchers are well familiar with dictionary-based approaches (e.g., LIWC or evaluative lexicon), more complex features (e.g., dialogue acts) and techniques (e.g., topic modeling and word embeddings) are only starting to receive more attention. Hopefully, this work will encourage more researchers to begin to adopt these powerful and revelatory tools.

Building a Question Breadth Classifier

While we had RAs manually code broad and narrow questions, to enable future researchers examine question breadth without having to start from scratch, we built a classifier. To do so, we build upon current NLP models that use transfer learning (see the appendix).

Results indicate that the classifier significantly outperforms the traditional methods on wh- questions (i.e., those that include a wh- phrase or the word how). The traditional method only relies on the first word, but to give it more flexibility, we consider the presence of wh- phrase in the first three words of a question (instead of only the first). Even so, the classifier is significantly more accurate (89% vs. 77%, $t(545) = 3.00$, $p < .05$). The classifier also performs well on yes/no questions (84%) though the traditional method also performs decently well there (85%, $t(1169) = -1.50$, $p = .20$) given most yes/no questions are truly narrow. Overall, the question breadth classifier improves accuracy and should be particularly useful when manual coding is not practical (e.g., large amount of data).

Directions for Future Research

Future work might examine how conversation type might shape language's effects. We focused on word of mouth conversations, but calling customer service or speaking to a boss are different situations with different norms that might shape conversational dynamics. Subsequent research might examine how things like conversational goals (e.g., making a good impression or solving a problem) and relationship dynamics between conversation partners (e.g., familiarity, social distance, or relative status) might impact language's effects.

Second, work might examine other question types. "Check questions" (e.g., questions that start with "that sounds like . . ."), for example, indicate someone understood what was said without suggesting advice or opinion (Althoff et al. 2016). Which type of questions are important should depend on the domain. While check questions may be valuable in counseling conversations, where it's important to signal listening, they are likely less useful in customer service calls because customers are explicitly seeking advice.

Third, while narrow questions had a larger effect than broad questions, we do not mean to suggest that broad questions are bad. Rather, it depends on the outcome of interest. If the goal is to encourage topic switching, broad questions may sometimes be helpful. Future work might examine other downstream consequence of different types of questions, such as attitudes toward one's conversation partner, or interest in talking again further.[9]

Fourth, work could examine the underlying mechanisms behind concreteness and questions. As discussed, using concrete language may make things easier to understand, increase interest, or make people feel like others are listening. Similarly, narrow questions not only indicate interest, but encourage attention to the topic at hand.

Finally, we don't mean to suggest the identified features are the only aspects that shape conversation length. Other linguistic features may also play a role. Consequently, to guide future work, we conducted some preliminary analyses exploring factors like pronouns, linguistic style similarity, linguistic contribution, and other dialogue acts (see the appendix). We hope this will encourage future research in this area.

REFERENCES

Aerts, Goele, and Peeter Verlegh (2018), "Say Your Name: How Language Concreteness in Online Consumer Reviews Mediates the Effect of Reviewer Identification on Review Helpfulness," in *International Conference on Research in Advertising* (*ICORIA*), ed. J. Aldás and N. Vila, University of Valencia.

Allison, Paul D. (1982), "Discrete-Time Methods for the Analysis of Event Histories," *Sociological Methodology*, 13, 61–98.

——— (2014), *Event History and Survival Analysis: Regression for Longitudinal Event Data*, SAGE, 46.

Althoff, Tim, Kevin Clark, and Jure Leskovec (2016), "Large-Scale Analysis of Counseling Conversations: An Application of Natural Language Processing to Mental Health," *Transactions of the Association for Computational Linguistics*, 4, 463–76.

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

Baumeister, Roy F., Liqing Zhang, and Kathleen D. Vohs (2004), "Gossip as Cultural Learning," *Review of General Psychology*, 8 (2), 111–21.

Berger, Jonah (2014), "Word of Mouth and Interpersonal Communication: A Review and Directions for Future Research," *Journal of Consumer Psychology*, 24 (4), 586–607.

9. Indeed, while our results might seem to contradict Di Stasi et al. (2022), who find that asking more of certain types of questions is associated with personal gain in negotiations, key differences likely explain the differing results. First, they bucket question types differently. While they contrast yes-no questions with ones that ask what, when, where, who, why, and how (described as "open-ended" questions), we follow Miles's (2013) definition of broad and narrow questions based on whether they broadly seek information or encourage a narrower response. Both why and who questions encourage more than a yes or no answer, for example, but while why questions broadly seek information, who questions often encourage a narrower response. Second, the outcome is different. While asking more than yes-no questions might encourage gains in negotiations because the longer responses generate an informational edge, we argue (and find) that narrower questions may encourage longer discussion of a particular topic because it keeps attention focused on that topic.

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

Berger, Jonah, and Katherine L. Milkman (2012), "What Makes Online Content Viral?" *Journal of Marketing Research*, 49 (2), 192–205.

Berger, Jonah, Wendy W. Moe, and David A. Schweidel (2023), "What Holds Attention? Linguistic Drivers of Engagement," *Journal of Marketing*, 87 (5), 793–809.

Berger, Jonah, and Grant Packard (2018), "Are Atypical Things More Popular?" *Psychological Science*, 29 (7), 1178–84.

Berger, Jonah, and Eric M. Schwartz (2011), "What Drives Immediate and Ongoing Word of Mouth?" *Journal of Marketing Research*, 48 (5), 869–80.

Brooks, Alison Wood (2025), *Talk: The Science of Conversation and the Art of Being Ourselves*, Crown.

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

Brysbaert, Marc, Amy Beth Warriner, and Victor Kuperman (2014), "Concreteness Ratings for 40 Thousand Generally Known English Word Lemmas," *Behavior Research Methods*, 46 (3), 904–11.

Chafe, Wallace L. (1970), *Meaning and the Structure of Language*, University of Chicago Press.

Chen, Frances S, Julia A. Minson, and Zakary L. Tormala (2010), "Tell Me More: The Effects of Expressed Interest on Receptiveness During Dialog," *Journal of Experimental Social Psychology*, 46 (5), 850–53.

Chevalier, Judith A., and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal Of Marketing Research*, 43 (3), 345–54.

Clark, Herbert H., and Edward F. Schaefer (1989), "Contributing to Discourse," *Cognitive Science*, 13 (2), 259–94.

Connell, Louise, and Dermot Lynott (2012), "Strength of Perceptual Experience Predicts Word Processing Performance Better than Concreteness or Imageability," *Cognition*, 125 (3), 452–65.

De Angelis, Matteo, Vito Tassiello, Cesare Amatulli, and Michele Costabile (2017), "How Language Abstractness Affects Service Referral Persuasiveness," *Journal of Business Research*, 72, 119–26.

Dillon, James T. (1982), "Problem Finding and Solving," *Journal of Creative Behavior*, 87 (5), 793–809.

Di Stasi, Matteo, Alison Wood Brooks, and Jordi Quoidbach (2022), "Asking Open-Ended Questions Increases Personal Gains in Negotiations," https://doi.org/10.31234/osf.io/9rh4j.

Di Stasi, Matteo, Emma Templeton, and Jordi Quoidbach (2024), "Zooming out on Bargaining Tables: Exploring Which Conversation Dynamics Predict Negotiation Outcomes," *Journal of Applied Psychology*, 109 (7), 1077–93.

Dubois, David, Andrea Bonezzi, and Matteo De Angelis (2016), "Sharing with Friends Versus Strangers: How Interpersonal Closeness Influences Word-of-Mouth Valence," *Journal of Marketing Research*, 53 (5), 712–27.

Dunbar, Robin I. M. (1996), "Groups, Gossip, and the Evolution of Language," in *New Aspects of Human Ethology*, Springer, 77–89.

Ghose, Anindya, Panagiotis G. Ipeirotis, and Beibei Li (2012), "Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowdsourced Content," *Marketing Science*, 31 (3), 493–520.

Godes, David, and Dina Mayzlin (2009), "Firm-Created Word-of-Mouth Communication: Evidence from a Field Test," *Marketing Science*, 28 (4), 721–39.

Hansen, Jochim, and Michaela Wänke (2010), "Truth from Language and Truth from Fit: The Impact of Linguistic Concreteness and Level of Construal on Subjective Truth," *Personality and Social Psychology Bulletin*, 36 (11), 1576–88.

He, Zihao, Leili Tavabi, Kristina Lerman, and Mohammad Soleymani (2021), "Speaker Turn Modeling for Dialogue Act Classification," in *Findings of the Association for Computational Linguistics: EMNLP 2021*, Association for Computational Linguistics, 2150–57.

Huang, Karen, Michael Yeomans, Alison Wood 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), 430–52.

Ireland, Molly E., and James W. Pennebaker (2010), "Language Style Matching in Writing: Synchrony in Essays, Correspondence, and Poetry," *Journal of Personality and Social Psychology*, 99 (3), 549–71.

Irvine, Robert, Douglas Boubert, Vyas Raina, Adian Liusie, Vineet Mudupalli, Aliaksei Korshuk, Zongyi Liu, et al. (2023), "Rewarding Chatbots for Real-World Engagement with Millions of Users," https://doi.org/10.48550/arXiv.2303.06135.

Jurafsky, D., E. Shriberg, and D. Biasca (1997), "Switchboard Dialog Act Corpus," Technical Report, International Computer Science Institute.

Kearsley, Greg P. (1976), "Questions and Question Asking in Verbal Discourse: A Cross-Disciplinary Review," *Journal of Psycholinguistic Research*, 5 (4), 355–75.

Keller, Ed, and Brad Fay (2012), "Word-of-Mouth Advocacy: A New Key to Advertising Effectiveness," *Journal of Advertising Research*, 52 (4), 459–64.

Kim, Su Jung, Ewa Maslowska, and Edward C. Malthouse (2018), "Understanding the Effects of Different Review Features on Purchase Probability," *International Journal of Advertising*, 37 (1), 29–53.

Langacker, Ronald W. (1987), "Nouns and Verbs," *Language*, 63 (1), 53–94.

Lee, Dokyun D. K., Emaad Manzoor, and Zhaoqi Cheng (2018), "Focused Concept Miner (FCM): Interpretable Deep Learning for Text Exploration," Working Paper No. 3304756, Carnegie Mellon University.

Liu, Xiao, Dokyun Lee, and Kannan Srinivasan (2019), "Large-Scale Cross-Category Analysis of Consumer Review Content on Sales Conversion Leveraging Deep Learning," *Journal of Marketing Research*, 56 (6), 918–43.

Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov (2019), "Roberta: A Robustly Optimized Bert Pretraining Approach," https://doi.org/10.48550/arXiv.1907.11692.

Miles, Edward W. (2013), "Developing Strategies for Asking Questions in Negotiation," *Negotiation Journal*, 29 (4), 383–412.

Moldovan, Sarit, Jacob Goldenberg, and Amitava Chattopadhyay (2011), "The Different Roles of Product Originality and Usefulness in Generating Word-of-Mouth," *International Journal of Research in Marketing*, 28 (2), 109–19.

Moore, Sarah G., and Katherine C. Lafreniere (2020), "How Online Word-of-Mouth Impacts Receivers," *Consumer Psychology Review*, 3 (1), 34–59.

Moore, Sarah G., and Brent McFerran (2017), "She Said, She Said: Differential Interpersonal Similarities Predict Unique Linguistic Mimicry in Online Word of Mouth," *Journal of the Association for Consumer Research*, 2 (2), 229–45.

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

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

Paetzold, Gustavo, and Lucia Specia (2016), "Inferring Psycholinguistic Properties of Words," in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, ed. K. Knight, A. Nenkova, and O. Rambow, Association for Computational Linguistics, 435–40.

Sadoski, Mark, Ernest T. Goetz, and Maximo Rodriguez (2000), "Engaging Texts: Effects of Concreteness on Comprehensibility, Interest, and Recall in Four Text Types," *Journal of Educational Psychology*, 92 (1), 85–95.

Schellekens, Gaby A. C., Peeter W. J. Verlegh, and Ale Smidts (2010), "Language Abstraction in Word of Mouth," *Journal of Consumer Research*, 32 (3), 207–23.

——— (2013), "Linguistic Biases and Persuasion in Communication About Objects," *Journal of Language and Social Psychology*, 32 (3), 291–310.

Semin, Gün R., and Klaus Fiedler (1988), "The Cognitive Functions of Linguistic Categories in Describing Persons: Social Cognition and Language," *Journal of Personality and Social Psychology*, 54 (4), 558–68.

Stolcke, Andreas, Klaus Ries, Noah Coccaro, Elizabeth Shriberg, Rebecca Bates, Daniel Jurafsky, Paul Taylor, Rachel Martin, Carol Van Ess-Dykema, and Marie Meteer (2000), "Dialogue Act Modeling for Automatic Tagging and Recognition of Conversational Speech," *Computational Linguistics*, 26 (3), 339–74.

Villarroel Ordenes, Francisco, Dhruv Grewal, Lauren Grewal, and Panagiotis Sarantopoulus (2019), "Language Dynamics in Employee-Customer Text-Based Interactions," in *Winter AMA Academic Conference: Understanding Complexity Transforming the Marketplace*, ed. M. Giesler, S. K. Lam, and X. Luo, American Marketing Association, 189.

Yeomans, Michael, K. F. Boland, Hanne K. Collins, Nicole Abi-Esber, and Alison Wood Brooks (2023), "A Practical Guide to Conversation Research: How to Study What People Say to Each Other," *Advances in Methods and Practices in Psychological Science*, 6 (4), https://doi.org/10.1177/25152459231183919.

Yeomans, Michael, and Alison Wood Brooks (2020), "Topic Preference Detection: A Novel Approach to Understand Perspective Taking in Conversation," Harvard Business School Working Paper No. 20-077, February.

Zhang, Jason Q., Georgiana Craciun, and Dongwoo Shin (2010), "When Does Electronic Word-of-Mouth Matter? A Study of Consumer Product Reviews," *Journal of Business Research*, 63 (12), 1336–41.