Insights From Textual Data and Machine Learning Algorithms For Consumer Behavior

Ada Aka, University of Pennsylvania
Sudeep Bhatia, University of Pennsylvania
Michael Yeomans, Imperial College London
Alison Wood Brooks, Harvard University
Ishita Chakraborty, Yale School of Management
Minkyung Kim, University of North Carolina-Chapel Hill
K. Sudhir, Yale School of Management
Vincent Chen, University of California, Berkeley
Zhihao Zhang, University of California, Berkeley
Ming Hsu, University of California, Berkeley

People generate vast amounts of text data as they search for information on the Internet, share articles, and write reviews. The existence of such data, along with advances in machine learning, have created new opportunities for scientists.Our session explores the use of these data/methods for topics central to consumer behavior.

Insights From Textual Data and Machine Learning Algorithms For Consumer Behavior

Chair: Ada Aka, The Wharton School, University of Pennsylvania, USA
Discussant: Ming Hsu, University of California, Berkeley, USA

Paper #1: Machine Learning Models For Predicting, Understanding and Influencing Health Perception

Ada Aka, University of Pennsylvania, USA
Sudeep Bhatia, University of Pennsylvania, USA

Paper #2: Topic Preference Detection: A Novel Approach to Understand Perspective Taking in Conversation

Michael Yeomans, Imperial College London, UK
Alison Wood Brooks, Harvard University, USA

Paper #3: Attribute Sentiment Scoring with Online Text Reviews: Accounting For Language Structure and Missing Attributes

Ishita Chakraborty, Yale School of Management, USA
Minkyung Kim, University of North Carolina-Chapel Hill, USA
K. Sudhir, Yale School of Management, USA

Paper #4: Mapping 25 Years of Consumer Knowledge from Text Corpora

Vincent Chen, University of California, Berkeley, USA
Zhihao Zhang, University of California, Berkeley, USA
Ming Hsu, University of California, Berkeley, USA

SESSION OVERVIEW

People generate and are exposed to vast amounts of text data as they search for information on the Internet, read and share news articles, generate social media posts, chat with friends and family, and write online reviews. The existence of such large amounts of data, along with recent advances in machine learning and natural language processing, have created new opportunities for social and behavioral scientists. Our session explores the use of these data and novel methods (that are different than dictionary-based approaches) for four topics central to the study of consumer behavior.

The first paper of the session examines consumer perceptions for over 700 medical conditions. The authors obtain quantitative representations of online text explanations of these medical conditions through state-of-the-art language models, and use these representations to accurately predict consumer health perception. The authors also use these models to study the psychological correlates of health perception, and understand how language influences health perception.

The second paper examines an important question that consumers constantly face when engaging in everyday conversation: whether to stay on the same topic or switch to a new one. Across multiple studies, the authors demonstrate that while people want to accommodate their partner's topic preferences, they consistently underperform compared to machine learning algorithms that use natural language processing methods for extracting text-based features of the conversations. Thus, this paper presents a novel, text-based approach for topic selection in conversation.

The third paper of the session integrates deep learning-based text analytics methods and structural econometric modeling to construct a real-time, scalable market intelligence tool from freely available online text reviews from websites such as Yelp.com. To accomplish this goal, they overcome two important challenges: computing accurate numerical sentiment scores from free-flowing online reviews and addressing the difficulties with missing attributes in text data. This novel approach is highly accurate and have the potential to answer additional novel marketing-related questions.

The fourth paper seeks to develop a text analytic approach to perform large-scale inferences of the evolution of consumer knowledge over time. Compared to traditional methods that are costly and limited in scope, the authors demonstrate that this text-based vector semantic algorithm is able to accurately capture and predict many aspects of the evolution of consumer knowledge in the last 25 years, therefore providing a flexible tool for uncovering theoretical and managerial insights into consumer knowledge.

To conclude, the four papers presented in this session will show how different types of text data, combined with various machine learning techniques, can be used to obtain new insights about consumer behavior. The prevalence of text data in everyday life, and the growing power and popularity of machine learning methods, has made easy for researchers to study nuanced behavioral phenomena in naturalistic settings, opening up new avenues for research across diverse areas of Marketing. For this reason, we believe our session will appeal to a broad audience of researchers, as well as practitioners, interested in understanding and influencing consumer behavior.

Machine Learning Models for Predicting, Understanding and Influencing Health Perception

EXTENDED ABSTRACT

Health perception has significant implications for healthcare funding. Unsurprisingly, such funding decisions depend on how consumers, voters, and donors perceive the severity of health states, and changes to media coverage and popular perceptions of a disease can have considerable effects on how much funding is allocated to health programs aimed at combating the disease (Casamayou, 2001).

How can we predict, understand, and influence people's health perceptions for common disease states? One possibility is to use quantitative measures such as "disability adjusted life years" (DALYs) that assess the severity of different medical conditions associated with the disease state (Calvert & Freemantle, 2003). Yet considerable research in psychology and marketing has found that such objective measures are not good predictors of people's health state perceptions (Slovic & Peters, 2006). That is, people are not actually good at evaluating the severity of different health states. Rather, their judgments rely on emotion, memory, linguistic, social, and other psychological cues, which occasionally lead to perceptions that deviate from objective measures such as the mortality rates or DALYs (Chapman, 2019).

More recently, the Internet has become an important information resource, with millions of people using health websites to inform health perceptions and guide health decisions. Our goal, in this paper, is to use information communicated on these websites to model health perceptions for hundreds of common disease states. We use textual information presented on the National Health Service (NHS) website, which is one of the main online sources of health information in the United Kingdom (Powell, Inglis, Ronnie, & Large, 2011). We additionally rely on recent advances in machine learning, known as word and sentence embeddings, which can quantify textual content for use in quantitative analysis (Bhatia et al., 2019; Günther et al., 2019; Jones et al., 2015). By using embedding methods to quantify the informational content of health descriptions on the NHS we can build a machine learning model that is capable of predicting health perception given a textual description of a health state.

We collected health perceptions for a large set of medical conditions, diseases, and other health states with an online experiment in Prolific. We first scraped online text explanations of 777 different health states such as ‘acne’ and ‘brain aneurysm’ on the NHS website. Each of these health states is associated with multiple sentences of summary information (M = 3.88, SD = 2.30) as well as many pages of additional details. In the experiment, participants (N = 782 UK residents) were asked to read NHS summaries for ten randomly selected health states, to imagine that they were diagnosed with each of the health states, and to report their evaluations of the health states by using a standard EQ-5D’s visual analogue scale.

We used two state-of-the-art language models called DistilBERT and Word2Vec (Sanh et al., 2019; Mikolov et al., 2013) to represent our NHS text explanations as high-dimensional vectors. In Study 1, these vector representations were mapped onto aggregate human judgments using a Ridge regression. Our machine learning approach was able to accurately predict how participants perceived different health states (out-of-sample correlation of r = 0.70 between predicted and observed health ratings, p < .0001). This correlation is close to the split-half reliability correlation of 0.75 which is the theoretical upper-bound in making such predictions. In Figure 1, we also demonstrate the power of our model by comparing its performance to other competing models and measures, including those that rely only on objective statistics like mortality rates, frequency in language, or simpler features extracted from the text data (e.g. text length, concreteness, and sentiment).

A natural next question in our investigation is to interpret the information contained in health state text explanations and discussions that gives rise to these successful predictions. Since the embedding vectors are based on word co-occurrence statistics in natural language, they quantify the extent to which words and concepts are associated with each other in language, and more generally, in the minds of lay people. Thus, in Study 2, we used our embeddings approach to explore which concepts and constructs are most associated with high (and low) health state judgments. Using Linguistic Inquiry and Word Count (LIWC; Tausczik & Pennebaker 2010) and other participant generated keywords as inputs into our model, we found that health states which contained text related to the constructs of “death”, “risk”, and “money” were more likely to be perceived as bad health states. In contrast, those with text related to other constructs such as “present-focused” (e.g., looks, work), “negation words” (e.g., shouldn’t, don’t) were more likely to be rated as good health states.

Finally, in Study 3, we used our embedding models to predict how different descriptions of the same disease state can be associated with different health perceptions. Here, our goal was to test the utility of our approach for studying health communication and its effects. In an online study run on Prolific Academic, we asked a separate group of participants (N = 80 UK residents) to complete our initial health states ratings survey after reading a health state summary from one of six health states obtained from health-related websites like AmericanBreastCancerFoundation.org. We also obtained a prediction from our model and compared our model predictions with actual participants’ average ratings. We found that we were able to correctly predict changes in participants’ ratings as a function of disease description. This illustrates one way in which our approach can be used to inform policy-insights and behavioral interventions for better health communication.

In summary, we presented a novel machine learning approach to estimate how people use online summaries of medical conditions, and other health states, to form health perceptions. Our approach uses recent advances in natural language processing, and is able to predict lay health perceptions with very high accuracy. This technique is not only unique in predicting health judgment, but also has significant time and cost efficiencies as it can be easily applied to out-of-sample diseases without participant data. Policymakers, marketers, and researchers can use our method to quantify people’s perceptions about different health states as well as to better understand the psychological cues that people use to make health state judgments. We look forward to future applications of textual data and machine learning to the study of lay health perception.

Topic Preference Detection: A Novel Approach to Understand Perspective Taking in Conversation

EXTENDED ABSTRACT

Conversation is one of the most common and important tasks humans do together (Dunbar, Marriott & Duncan, 1997; Pickering & Garrod, 2004). People reveal their preference to converse constantly, and while conversations can serve instrumental or strategic goals (Crawford & Sobel, 1982; Berger, 2014), they can be intrinsically and mutually enjoyable, as well (Mehl, Vazire, Holleran & Clark, 2010; Epley & Schroeder, 2014; Kumar & Gilovich, 2015; Sun, Harris & Vazire, 2019). But even when people share a seemingly simple, cooperative goal such as enjoyment, the decisions necessary for generating conversation can be quite complex. While people may want to talk, they must still mutually decide what to talk about.

Topics are a fundamental structure of conversation, allowing speakers to jointly maintain common ground (Hardin & Higgins, 1996; Passoneau & Litman, 1997; Stalnaker, 2002; Schegloff, 2007). But consider a judgment that everyone confronts during every turn of every conversation: Should we stay on topic, or switch topics?

This decision can be difficult. Although people may have a general sense of their own preferences for different topics, those preferences may vary depending on whom they are talking to. Furthermore, others’ preferences can be ambiguous, even after the topic has begun. Formally, we define this task as “topic preference detection”: Can people tell whether someone else wants to stay on a topic based on what they have said about it?

Here, we study how well people learn their partners’ topic preferences, and whether that affects their own topic choices. We develop this as a naturalistic perspective-taking task, that tests how well people can learn about one another in conversation (Eyal, Steffel, & Epley, 2018). Across three studies we find that (i) people want to accommodate their partner’s topic preferences, but (ii) they routinely fail to detect what topics other people prefer. All data, code and preregistrations can be found (anonymized) on OSF at https://goo.gl/gmxyR3.

To define a common set of topics as stimuli, we conducted a pilot study (n=199), using fifty topic-starting questions drawn from previous research (Aron et al., 1997, Huang et al., 2017). Participants wrote a response to each topic question and then rated their preference (-10 to +10) for staying on topic. From that we chose a list of 12 topics (e.g. “what is the strangest thing about where you grew up?”), using topics that had average (but high-variance) ratings. Throughout, we estimate preference detection accuracy as the non-parametric correlation between predicted and actual preferences over the twelve topics.

Study 1A had an asynchronous design, similar to the pilot. We asked mTurk participants (n=392) to read the twelve topic questions one at a time, writing a response, and reporting their preference for staying on-topic. In Study 1B, other mTurkers (n=654) rated their own preferences for each topic, and then predicted the writer's topic preferences for a set of 24 responses (6 topics each for 4 writers). These human judges were less able to detect writer's preferences (τ = .142, CI_{95} = [.127, .158]), than a simple NLP algorithm (τ = .174, CI_{95} = [.151, .196]) that parsed the same text responses that humans had seen, using ngrams + politeness + word2vec features in a cross-validated LASSO regression (Freidman, Hastie & Tibshirani, 2010; Mikolov et al, 2017; Jurafsky & Martin, 2017; Yeomans, Kantor & Tingley, 2018). The judges also rated their preference for each topic with each of the four writers whose responses they saw. Their responses revealed a strong preference for mutually enjoyable topics – they preferred topics they thought their partner liked (τ = .531, CI_{95} = [.511, .550]). We empirically disentangle this topic accommodation from egocentric projection, which was also common among judges in every study we ran.

In Study 2, we again used the asynchronous topic preference detection paradigm from Study 1, recruited pairs of participants who know each other well (e.g. friends, family) to predict one another's preferences (n=172, preregistered). Again, the machine learning algorithms (τ = .155, CI_{95} = [.121, .189]) outperformed humans (τ = .188, CI_{95} = [.156, .219]). Participants were also overconfident in their prediction accuracy. And when they had more information about their partner - by having known their partner for longer, or by reading their response (vs. only the question) - their confidence increased, but not their accuracy.

In Study 3, we recruited laboratory participants (n=196) to take part in a synchronous paradigm. Participants reported their preference for each topic before meeting their partner, then had 10-minute dyadic conversations. Afterwards, they reported their own topic preferences, and predicted their partner's. Pre-conversation preferences strongly predicted speaking time on each topic (β = .117, SE = .023). And like Study 1, post-conversation preferences revealed a desire for mutually enjoyable topics (β = .557, SE = .031). However, speakers were less accurate predicting topic preferences (τ = .203, CI_{95} = [.166, .240]) than a simple word count of each person's on-topic speech (τ = .247, CI_{95} = [.206, .288]). A separate set of lab participants (N=330) watched videos of the conversations, but were not as accurate as the speakers themselves (τ = .150, CI_{95} = [.128, .171]). Conversely, we applied new a preference detection NLP algorithm to the transcripts, using the same features as before, as well new dialogic cues (e.g. laughter, pauses, follow-up questions), and it again outperformed humans (τ = .338, CI_{95} = [.305, .369]).

This research provides a novel framework for topic selection in co-operative conversation. Our finding show that topic preferences drive conversational behavior, and conversationalists want to find mutually enjoyable topics with one another. However, they are constrained by their ability to listen and reason about other people's conversational behavior. This suggests important limits to perspective taking even in open-ended natural language.

Attribute Sentiment Scoring with Online Text Reviews: Accounting for Language Structure and Missing Attributes

EXTENDED ABSTRACT

Many firms conduct routine tracking surveys on product/service performance on selected attributes chosen by managers that they believe drive overall customer satisfaction (Mittal et al. 1999, Mittal et al. 2001). The summary scores from these surveys are used as dashboard metrics of overall satisfaction and used as performance metrics at firms. However, surveys are costly, suffer from response biases and get outdated quickly (Culotta and Cutler 2016, Bi et al. 2019). Therefore, crowd-sourced online review platforms have emerged as an alternative and less expensive source of scalable, real-time feedback for businesses to listen in on their markets for both performance tracking as well as competitive benchmarking (e.g., Xu 2019, Li et al. 2019).

In this paper, the authors combine deep learning-based text analytics methods with structural econometric modeling to develop a real-time, scalable market intelligence tool from freely available online reviews. While we use restaurant reviews from Yelp.com as an empirical illustration of our method, this tool can be applicable in a range of industries like hotels, education etc. where firms care about monitoring a fixed set of managerially important attributes over time and benchmarking against competitive performance. Even when not used as a replacement for tracking surveys of performance, such quantitative summary metrics are valuable for managers because consumers use review platforms when making choices (e.g., Zhu and Zhang 2010, Luca and Vats 2013). In a study conducted on Amazon Mechanical Turk, we find evidence that providing attribute-level sentiment scores instead of an overall rating and text improves consumer decision-making by reducing the cognitive burden in making a choice. Moreover, with employee compensation and performance being directly linked to online review performance in many firms, the need to develop reliable quantitative metrics that capture attributes and related sentiments from online UGC both for tracking gaps in customer satisfaction as well as managing one's e-reputation has gained critical importance.

Deriving attribute-sentiment scores from text reviews requires addressing two novel and challenging problems: The first challenge lies in coming up with accurate numerical sentiment scores from free-flowing online reviews. For this, they develop a deep learning convolutional-LSTM hybrid model to account for language structure, in contrast to bag-of-words methods that rely on word frequency alone. Bag-of-words based approaches are limited in their ability to adequately score attribute sentiments especially for certain classes of hard sentences. Examples of hard sentences include various types of negations like contrastive (``but", "yet"), long sentences and instances of sarcasm that account for almost 50% of sentences in online reviews. Consider the following examples where sentiment degree is modified, as in (i) "horrible," "not horrible," "not that horrible" and (ii) "delight, "just missed being a delight". When words are just counted as in bag-of-words, making the connections between the key sentiment words "horrible" and "delight" with their degree modifiers will be difficult, without considering how they are grouped adjacently to form phrases—i.e., spatial structure. Likewise, it is difficult to capture the true sentiment in a long or contrastive sentence without accounting for the order in which the words/phrases occur (sequential structure). In our model, the convolutional layer accounts for the spatial structure (adjacent word groups or phrases) and LSTM accounts for the sequential structure of language (sentiment distributed and modified across non-adjacent phrases). Our deep learning model brings about significant accuracy improvements not only in 5-level granular sentiment classification but also in polarity detection (positive and negative) for both "easy" and "hard" sentences.

The second challenge is addressing the problem of missing attributes in text in constructing attribute sentiment scores—as reviewers write only about a subset of attributes and remain silent on others. It is important to understand what causes people to remain silent on certain attributes because assuming ``missing" as "unimportant" can bias attribute-sentiment scores. Further, behavioral science research has long recognized the importance of the right imputation for missing values because people do not ignore missing attributes in evaluations and often make complex and imperfect inferences from them (Gurney and Loewenstein 2019). For addressing attribute silence, the paper develops and estimates a structural model of reviewer rating behavior that takes into account the data generating process to develop a model-based imputation procedure. This econometric model of rating behavior also helps to identify the different incentives of various groups of reviewers to engage in online WOM. We find three segments of reviewers—the smallest but most active reviewers ("Status Seeking Regulars") who write mainly for being informative to others and maintaining platform status; the largest segment ("Altruistic Mass") who review without reward expectations, and a mid-size segment of "Emotive Irregulars" who review infrequently but write about attributes they are extremely satisfied or dissatisfied with. Our insights around attribute silence in reviews shows that informativeness and need to praise/vent drive more of the writing than the importance of the attribute. Not only does this contribute to the literature on why people engage in online word of mouth (Berger 2014), it also has implications for using reviews as a source of data for needs/benefits identification. In particular, contrary to conventional wisdom, the frequency of mentions of a benefit or a topic may not necessarily be a proxy of its importance. Overall, the paper illustrates the value of combining "engineering" thinking underlying machine learning approaches with "social science" thinking from econometrics to answer novel marketing questions.

Mapping 25 Years of Consumer Knowledge from Text Corpora

EXTENDED ABSTRACT

Consumer knowledge, the set of consumers' acquired understanding of brands, products, and other offerings, is known to be an important driver of consumer responses to brands and product offerings (Alba & Hutchinson, 2000; Hadar et al., 2013; Berger et al., 2020). Notably, consumer knowledge is constantly evolving and can change substantially over time (Smith and Lux, 1993; Polanyi, 1957). Just fifteen years ago, Facebook and MySpace were both rising stars in the social media category. Now, one is nearly synonymous with social media, while the other has fallen out of the mind of many. The ability to understand and predict such changes are thus valuable from both scientific and managerial perspectives.

To our knowledge, however, no method exists that can provide a quantitative, data-driven description of the evolution of consumer knowledge. Although longitudinal databases (especially commercial ones) exist, it is difficult to capture the full trajectory of every brand due to difficulties in keeping up with the constant entry and exit of firms and brands in the marketplace. Short of having a time machine, researchers cannot survey consumers from the past.

This study proposes a novel text analytic technique, Principal Semantic Component Analysis (PSCA), to track the changing meanings of words from time-indexed text corpora and characterize the evolution of consumer knowledge. Specifically, PSCA represents a combination of Principal Component Analysis (PCA) and the Dynamic Word Embedding (DWE) model (Yao et al., 2018), a recently developed diachronic natural language processing (NLP) approach that captures semantic changes.

Compared to static NLP methods such as traditional word embedding models, where the meaning of a word is invariant across time, DWE assigns a distinct meaning to a word per time period. For example, Hamilton et al. (2016) used an early version of DWE to trace the change in meaning of the word "gay" through the past hundred years from emotion to sexuality. Building on this line of literature, we hypothesize that a DWE approach can capture the rapid and sometimes subtle changes in consumer knowledge. Specifically, in the studies below, we use a DWE model based on New York Times articles from 1996 to 2019, containing 1.3 billion words.

Study 1 presents PSCA findings on changing meanings of brands. Although the method is applicable to any arbitrary brand, only a select set of brands are shown given space constraints. Fig. 2A shows the DWE trajectory of the word "blackberry," which tracks the significant change of the word over the past 25 years, from being associated with the fruit to the smartphone brand and back again. A quantitative account of the above statement therefore should contain two parts: the *vector* of the largest semantic change, and the *timing* of the movements. In the case of "blackberry," a successful model should tell us (a) where "blackberry" was moving to and from, e.g., "smartphone", and (b) when "blackberry" was at the peak or nadir in terms of strength of association, e.g., peaking around 2008 to 2011, which corresponds to the peak years of the brand Blackberry according to its stock prices and revenues.

Specifically, PSCA captures the vector of the largest semantic changes by computing the first principal component of DWE, referred to as Principal Semantic Component (PSC) hereafter. To identify the timing of movements, we then computed the time series of semantic similarities between the brand and the PSC. In Fig. 2B, the PSC of "blackberry" has the meaning of "android, app, smartphone." That is, PSCA identifies that "blackberry" moved most along the direction of "smartphone" from 1996 to 2019. The time series also gives information about the timing of Blackberry's rise and fall along the "smartphone" dimension. In particular, PSCA identifies 2010 as the time when Blackberry reached its peak as a smartphone brand, before dropping sharply after.

To explore the generalizability of PSCA beyond brands, we applied PSCA to detect changes in more intangible concepts, such as fashion trends and fads. For example, PSCA identifies that the word "Atkins" had a sharp but transient rise in association with diets in 2004, likely reflecting the rise and fall of the Atkins diet, and it uncovers how the word "selfie" came to popularity since 2010 and is still going strong (Fig. 2C).

These and other results that match with common impressions provide initial support for the utility of PSCA. In Study 2, we show that PSCA findings correspond with, and can even forecast, commonly used marketing metrics. Specifically, we compare PSCA results with several external metrics related to consumer knowledge. As a benchmark for existing text-based metrics for popularity, we also include word counts of the brand names. Due to space constraints, here we only present comparisons with Interbrand's Best Global Brand ranking, a ranking based on brand equity. Other metrics, including number of users and Brand Asset Valuator scores, lead to similar results.

There are several consistent trends which show that PSCA captures at least part of the consumer knowledge (Fig. 2DE). First, for brands which have peaks in the external metrics, e.g. Blackberry, peaks of PSCA coincide with those of external metrics. Moreover, we found that PSCA acts as a leading indicator of other metrics in the rising stage. This suggests that the semantic information leveraged by PSCA allows the algorithm to detect the rising trends earlier than considering word counts alone. Another notable feature is that external metrics often delay in providing data about brands. For example, Interbrand did not include Facebook until 2011, when the social media platform had already plateaued out according to PSCA. This highlights the important role of textual data in providing information about brands in their early stages, when longitudinal data is scarce.

Collectively, the results demonstrate that (a) textual data indeed contain rich information about consumer knowledge which is otherwise difficult to obtain, and (b) PSCA offers an accessible solution to extracting such information with similar temporal resolution as survey-based methods, but at much lower costs.

REFERENCES

Alba, J. W., & Hutchinson, J. W. (2000). Knowledge calibration: What consumers know and what they think they know. Journal of consumer research, 27(2), 123-156.

Aron, A., Melinat, E., Aron, E. N., Vallone, R. D., & Bator, R. J. (1997). The experimental generation of interpersonal closeness: A procedure and some preliminary findings. *Personality and Social Psychology Bulletin, 23(4),* 363-377.

Berger J (2014) Word of mouth and interpersonal communication: A review and directions for future research. Journal of consumer psychology 24(4): 586–607

Berger, J. (2014). Word of mouth and interpersonal communication: A review and directions for future research. *Journal of Consumer Psychology, 24(4),* 586-607.

Berger, J., Humphreys, A., Ludwig, S., Moe, W. W., Netzer, O., & Schweidel, D. A. (2020). Uniting the tribes: Using text for marketing insight. Journal of Marketing, 84(1), 1-25.

Bi JW, Liu Y, Fan ZP, Zhang J (2019) Wisdom of crowds: Conducting importance-performance analysis (ipa) through online reviews. Tourism Management 70:460–478

Calvert, M. J., & Freemantle, N. (2003). Use of health-related quality of life in prescribing research. Part 1: why evaluate health-related quality of life?. *Journal of clinical pharmacy and therapeutics*, *28*(6), 513-521.

Casamayou, M. H. (2001). *The politics of breast cancer*. Georgetown University Press.

Chapman, G. B. (2019). A decision-science approach to health-behavior change. *Current Directions in Psychological Science*, *28*(5), 469-474.

Corbett, J. B., & Mori, M. (1999). Medicine, media, and celebrities: News coverage of breast cancer, 1960–1995. *Journalism & mass communication quarterly*, *76*(2), 229-249.

Crawford, V. P., & Sobel, J. (1982). Strategic information transmission. *Econometrica: Journal of the Econometric Society,* 1431-1451.

Culotta A, Cutler J (2016) Mining brand perceptions from twitter social networks. Marketing science35 (3):343–362.

Dunbar, R. I., Marriott, A., & Duncan, N. D. (1997). Human conversational behavior. *Human nature, 8(3),* 231-246.

Epley, N., & Schroeder, J. (2014). Mistakenly seeking solitude. *Journal of Experimental Psychology: General, 143(5),* 1980.

Eyal, T., Steffel, M., & Epley, N. (2018). Perspective mistaking: Accurately understanding the mind of another requires getting perspective, not taking perspective. *Journal of personality and social psychology, 114(4),* 547.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, 33(1),* 1-22.

Günther, M., Thiele, M., & Lehner, W. (2019). Retro: Relation retrofitting for in-database machine learning on textual data. *arXiv preprint arXiv:1911.12674*.

Gurney N, Loewenstein G (2019) Filling in the blanks: What restaurant patrons assume about missing sanitation inspection grades. Journal of Public Policy & Marketing 0743915619875419.

Hadar, L., Sood, S., & Fox, C. R. (2013). Subjective knowledge in consumer financial decisions. Journal of Marketing Research, 50(3), 303-316.

Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016). Diachronic word embeddings reveal statistical laws of semantic change. arXiv preprint arXiv:1605.09096.

Hardin, C. D., & Higgins, E. T. (1996). Shared reality: How social verification makes the subjective objective. In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition, Vol. 3*. Guilford Press.

Huang, K., Yeomans, M., Brooks, A.W., Minson, J., & Gino, F. (2017). It doesn't hurt to ask: Question-asking increases liking. *Journal of Personality and Social Psychology, 113(3),* 430-452.

Jones, M. N., Willits, J., Dennis, S., & Jones, M. (2015). Models of semantic memory. *Oxford handbook of mathematical and computational psychology*, 232-254.

Jurafsky, D., & Martin, J. H. (2017). *Speech and language processing (Vol. 4).* London: Pearson.

Kumar, A., & Gilovich, T. (2015). Some "thing" to talk about? Differential story utility from experiential and material purchases. *Personality and Social Psychology Bulletin, 41(10),* 1320-1331.

Li Y, Lu SF, Lu LX (2019) Do yelp reviews influence consumer choice in the presence of government ratings? evidence from us nursing homes. Evidence from US Nursing Homes (October1, 2019)

Luca M, Vats S (2013) Digitizing doctor demand: The impact of online reviews on doctor choice. Cambridge, MA: Harvard Business School

Mehl, M. R., Vazire, S., Holleran, S. E., & Clark, C. S. (2010). Eavesdropping on happiness: Well-being is related to having less small talk and more substantive conversations. *Psychological science*, *21*(4), 539-541.

Mikolov, T., Grave, E., Bojanowski, P., Puhrsch, C., & Joulin, A. (2017). Advances in Pre-Training Distributed Word Representations. *arXiv preprint arXiv:1712.09405.*

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *arXiv preprint arXiv:1310.4546*.

Mittal V, Katrichis JM, Kumar P (2001) Attribute performance and customer satisfaction over time: evidence from two field studies. Journal of Services Marketing.

Mittal V, Kumar P, Tsiros M (1999) Attribute-level performance, satisfaction, and behavioral intentions over time: a consumption-system approach. Journal of Marketing 63(2):88–101.

Passonneau, R. J., & Litman, D. J. (1997). Discourse segmentation by human and automated means. *Computational Linguistics, 23(1),* 103-139.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences, 27(2),* 169-190.

Polanyi, K. (1957). The great transformation: The political and economic origins of our time. Boston: Beacon Press.

Powell, J., Inglis, N., Ronnie, J., & Large, S. (2011). The characteristics and motivations of online health information seekers: cross-sectional survey and qualitative interview study. *Journal of medical Internet research*, *13*(1), e20.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1978). A simplest systematics for the organization of turn taking for conversation. In *Studies in the organization of conversational interaction* (pp. 7-55). Academic Press.

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

Schegloff, E. A. (2007). *Sequence organization in interaction: Volume 1: A primer in conversation analysis (Vol. 1).* Cambridge University Press.

Slovic, P., & Peters, E. (2006). Risk perception and affect. *Current directions in psychological science*, *15*(6), 322-325.

Smith, R. A., & Lux, D. S. (1993). Historical method in consumer research: Developing causal explanations of change. Journal of Consumer Research, 19(4), 595-610.

Stalnaker, R. (2002). Common ground. *Linguistics and philosophy*, *25*(5/6), 701-721.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology*, *29*(1), 24-54.

Xu X (2019) Examining the relevance of online customer textual reviews on hotels' product and service attributes. Journal of Hospitality & Tourism Research43(1):141–163

Yao, Z., Sun, Y., Ding, W., Rao, N., & Xiong, H. (2018). Dynamic word embeddings for evolving semantic discovery. In Proceedings of the eleventh acm international conference on web search and data mining (pp. 673-681).

Yeomans, M., Kantor, A., & Tingley, D. (2018). The politeness Package: Detecting Politeness in Natural Language. *R Journal*, *10*(2).

Zhu F, Zhang X (2010) Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. Journal of marketing74(2):133–148