Journal of the Academy of Marketing Science
https://doi.org/10.1007/s11747-018-0581-x

REVIEW PAPER

Unstructured data in marketing

Bitty Balducci[1] · Detelina Marinova[1]

Received: 7 November 2017 / Accepted: 4 April 2018

Abstract
The rise of unstructured data (UD), propelled by novel technologies, is reshaping markets and the management of marketing activities. Yet these increased data remain mostly untapped by many firms, suggesting the potential for further research developments. The integrative framework proposed in this study addresses the nature of UD and pursues theoretical richness and computational advancements by integrating insights from other disciplines. This article makes three main contributions to the literature by (1) offering a unifying definition and conceptualization of UD in marketing; (2) bridging disjoint literature with an organizing framework that synthesizes various subsets of UD relevant for marketing management through an integrative review; and (3) identifying substantive, computational, and theoretical gaps in extant literature and ways to leverage interdisciplinary knowledge to advance marketing research by applying UD analyses to underdeveloped areas.

Keywords Unstructured data · Machine learning · Deep learning · Artificial intelligence · Nonverbal · Image · Video · Voice · Text · Linguistics · Acoustic · Big data · Text mining

The contemporary world is characterized by rapid advances in technology that are pervasive in everyday life (Huang and Rust 2017). This swift development has also spurred an unprecedented influx of unstructured data (UD). UD is commonly understood as "information that either does not have a pre-defined data model or is not organized in a pre-defined manner" (Wikipedia 2017). An estimated 80% of data held by firms today are unstructured (Rizkallah 2017), and they are growing 15 times faster than structured data (SD) (Nair and Narayanan 2012). This global expansion has not gone unnoticed; 87% of marketers cite data as their most underutilized resource but note that deriving value from various sources of UD remains a key challenge (Howatson 2016). The obstacles faced in extracting knowledge from UD mean that firms often sit idly on expansive troves of it, earning UD the designation "dark analytics" (Briggs and Hodgetts 2017). Yet, unlocking the insights embedded in this burgeoning resource has the potential to be particularly valuable in marketing, sales and service settings where UD volumes are an estimated five times greater than SD (Davies 2015). A wealth of unique information can be derived from analyses of UD for managerially relevant domains of interest such as competitive advantages (Coughlin 2017), social networks (Lohr 2012), and data privacy (Rizkallah 2017).

The rapid emergence and growth of technologies capable of analyzing vast amounts of UD through machine learning and other artificial intelligence methods (Marr 2017) has also made UD increasingly prominent in marketing literature. However, current applications of UD in marketing are fragmented, reflecting the scattered domains that house requisite theories (communication, linguistics, or psychology), novel methods (computer science), and untapped data owned by organizations. A critical review of this new and thriving field, to create an organizing framework, thus is essential to identify opportunities for future research.

This article makes three main contributions to the literature. First, it contributes to the growing body of research analyzing UD by offering a unifying definition and conceptualization of UD in marketing. This effort can assist scholars pursuing research with UD by shedding light on its characteristics, which can be used to leverage unique insights compared with traditional, SD analyses. Recent research gives examples of data

Mark Houston served as Area Editor for this article.

✉ Detelina Marinova
marinovad@missouri.edu

Bitty Balducci
bitty.balducci@mail.missouri.edu

[1] Robert J. Trulaske College of Business, University of Missouri, 424 Cornell Hall, Columbia, MO 65211, USA

that are unstructured (e.g., Hewett et al. 2016; Wedel and Kannan 2016), but to the best of our knowledge no formal definition or conceptualization of UD appears in the marketing literature.

Second, this article unifies disjoint literature within an organizing framework that synthesizes numerous subsets of UD relevant for marketing management through a review of marketing and other relevant literature. This contribution reflects the conceptual goals of explicating and relating (MacInnis 2011), in that we detail the unique characteristics of UD, "taking stock or reducing what is known" about the use and unique contribution of UD to a manageable set of key takeaways (MacInnis 2011, p. 144) and we draw connections between scattered research that uses UD across substantive domains. This integrative framework also provides insights on the dynamic nature of UD and reveals the theoretical richness and computational advancements that can be gained from other disciplines.

Third, this article identifies substantive (Table 2), theoretical (Table 3), and computational (Table 4) gaps in prior literature that warrant further research and demonstrates ways to leverage interdisciplinary knowledge to advance marketing research with UD in underdeveloped areas. For our critical review and research framework development, we consider articles that use UD for quantitative analysis relevant for marketing management. The types of UD we review include text, video, voice, images, nonverbal (e.g., facial and gestural cues), and select automated methods (Fig. 1). Although other types of UD are present in the marketing literature (e.g., physiological), they fall outside the scope of this paper. Nor do we include manipulations of UD such as alterations of images (Bashir and Rule 2014) or voice pitch (Lowe and Haws 2017) that are ultimately measured with surveys or other numerical methods. Similarly, studies that have an UD source, such as text, but analyze the data through structured approaches such as star ratings for online reviews (e.g., Ho-Dac et al. 2013; Kostyra et al. 2016), conduct analysis through qualitative methods (e.g., Kozinets et al. 2010) or simply count occurrences (Yadav et al. 2007) are not included. Finally, we focused on articles in *Journal of Marketing*, *Journal of Marketing Research*, *Journal of the Academy of Marketing Science*, *International Journal of Research in Marketing*, and *Marketing Science* published in the last 15 years, but we also include articles from other journals and prior decades to gather information on the types of UD with scant research.[1] For a thorough review of text analysis in consumer behavior research, see Humphreys and Wang (2018).

[1] For example, no papers published between 2002 and 2017 in the listed journals provided quantitative analyses of voice data. Therefore, we surveyed additional journals (e.g., *Journal of Personal Selling and Sales Management*) and included issues published before 2002 to capture the state of prior literature on voice data.

Unstructured data: definition and characteristics

We conceptualize the structure of any single data unit on a continuum from highly unstructured to highly structured (Fig. 2). A data unit's location on this continuum, determined by the perspective of the researcher, reflects the ease with which structure can be added to the data to make it suitable for quantitative analysis that aims to yield generalizable insights (Fig. 3). Some data units, such as video data, contain many simultaneous data points (e.g., nonverbal cues, acoustic vocal cues, words spoken) that flow concurrently requiring the researcher to assign values to them, manually or automatically, prior to quantitative analysis. However, other data units require relatively less (e.g., survey data) or almost no (e.g., net sales data in USD) effort on the part of the researcher to ready them for analysis. For this study, we consider the extent to which a *single* data unit is (un)structured. Text data from online reviews would constitute a single data unit; multiple data points pertaining to the same event derived from various units would not. For example, retailers may collect many data points about a customer during a purchase transaction, but such information comes from multiple data units (e.g., time, cost, location of purchase), so each unit must be separately considered and prepared for quantitative analysis.

We define UD as a single data unit in which the information offers a relatively concurrent representation of its multifaceted nature without predefined organization or numeric values. Table 1 clarifies what constitutes UD according to three characteristics that differentiate highly UD from highly SD, as well as specifying how UD might be employed to develop theory and obtain novel conceptual and managerial insights beyond what can be gleaned from SD.

Nonnumeric

The first characteristic of highly UD is that they are nonnumeric. They lack predefined numeric assignments for the constructs of interest and researchers must conduct manual or automatic coding prior to analysis. For example, to determine the level of customer expressed affect through nonverbal cues in a service exchange, a researcher must first consider which nonverbal cues embody different levels of positive, negative, and neutral affect before determining the degree of expressed affect in each cue and counting the number of occurrences in the interaction. Conversely, highly SD frequently do have a predefined numeric representation of the construct of interest and in instances when numeric values are not innate, the researcher can easily assign them (e.g., categorical data).

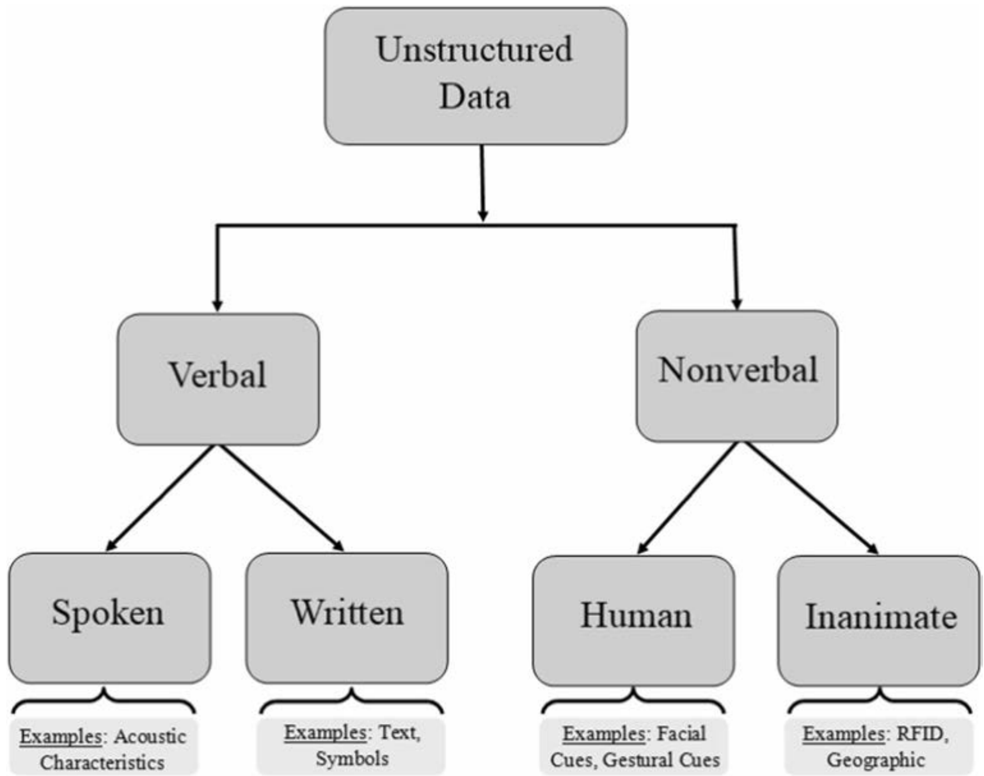

Fig. 1 Overview of unstructured data

Multifaceted

The second characteristic of highly UD is that they are multifaceted. A *single unit* of highly UD possesses multiple facets, each offering unique information enabling the researcher to select and analyze facet(s) based on the research goals. For example, voice data contains many facets (e.g., pitch, speech rate, intensity) that all provide unique information since each of these facets conveys different information about the speaker (e.g., affective state, persuasiveness). Despite the abundance of unique facets in voice data, researchers might draw on psychology and communication literature to identify distinct psychological responses that each prosody measure conveys and only include those that make theoretical and logical sense for the research question at hand. Conversely, highly SD is unifaceted meaning that a single data unit on this end of the continuum only offers one unique piece of information. For example, although a net promotor score can be viewed from different perspectives (e.g., service quality, customer satisfaction), this data unit only possess one unique facet—the numeric score. Whether this data unit is viewed as an indicator of service quality or customer satisfaction is the result of the

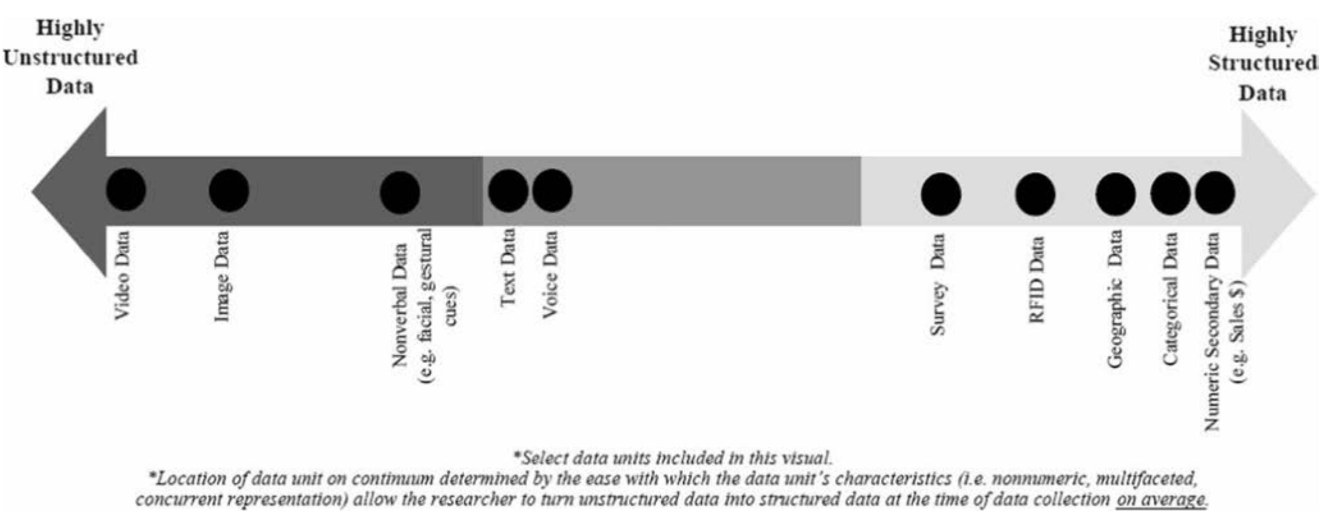

Fig. 2 Unstructured-structured data continuum

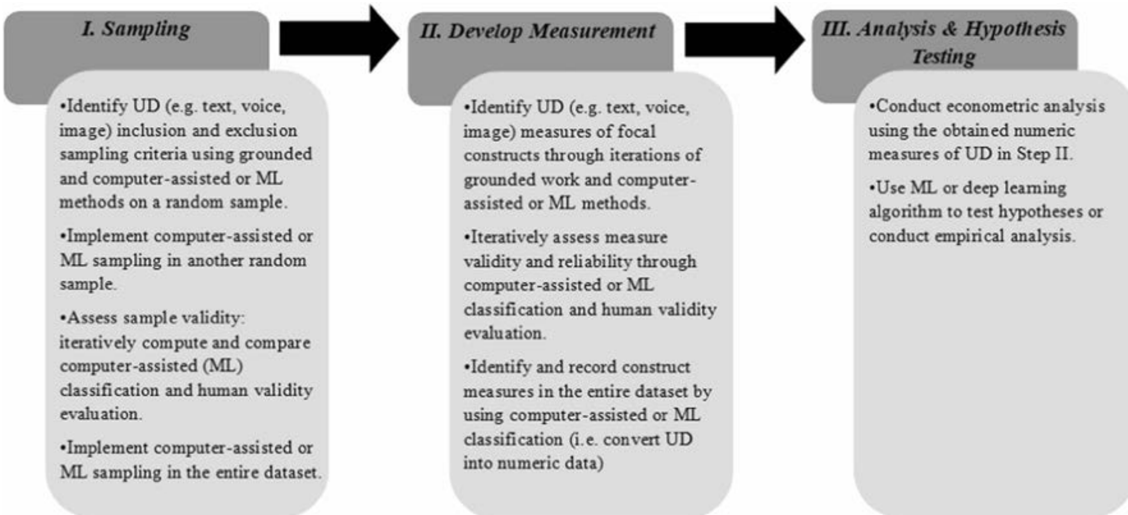

Fig. 3 Implementing unstructured data analysis in marketing research

researcher's interpretation of the data rather than some fundamental difference in a selected facet of that data unit.

Concurrent representation

The final characteristic of highly UD is that they maintain concurrent representation. The simultaneous presence of a single data unit's multiple facets that each provide unique information allows an UD unit to represent different phenomena at the same time. Thus, the scholar can examine diverse research questions with a single highly UD unit through examination of the concurrent flow of these unique facets. Consider text data documenting an email exchange between a salesperson and a customer. A single unit of this text data contains many unique facets (e.g., syntax, semantics) that occur simultaneously, and each of these facets provides distinct information that the scholar can use to assess different phenomena (e.g., persuasion, affect). This example reveals how the facets of a single unit of highly UD are unique, yet they are woven together and morph in parallel. Due to SD's unifaceted nature, these data exhibit non-concurrent representation. Even if highly SD might be fluid (e.g., number of Google searches per hour for a sampling period), a single unit of highly SD can only report on one facet at a time. Therefore, highly SD must move from one data unit to the next and cannot develop concurrent representation. For ease of exposition, we refer to data units located in the left and middle panel of Fig. 2 as UD, whereas data units located in the right panel represent SD.

Leveraging unstructured data for unique theoretical insights

Marketing management is understood as a process of creating, delivering, and communicating value to customers (Kotler and Keller 2015). We present three general categories of theoretical contribution that can be derived from UD to create, communicate, and deliver value to customers. We draw on Shannon and Weaver's (1949) communication theory, which deals with the process of information encoding and decoding that takes place as a message flows from a sender to a recipient, as is typical in the course of marketing management activities. This process is marked by technical, semantic, and effectiveness challenges (Shannon and Weaver 1949). The technical problem refers to the accuracy of information transmitted; the semantic problem entails how the message recipient interprets or processes the information; and the effectiveness problem is related to the extent to which the decoded message aligns with the sender's intended meaning (Shannon and Weaver 1949). This view is well-suited to handle the multitude of conceptual approaches that drive various types of UD research in marketing management because transmission, processing, and extraction of information are fundamental processes that fuel value creation, communication, and delivery. We therefore review and discuss the main ways UD has been used to communicate value to customers through information transmission, deliver value through information processing, and create value from information extraction. This integrative review also emphasizes rapidly growing areas of interest in academia and practice.

Information transmission

The information transmission stage of the communication cycle encodes information to transfer it effectively (Shannon and Weaver 1949) and thus underlies the process of value communication to customers. Communication begins with creation of the message from an information source (e.g., marketer), flows through a transmitter (e.g., online banner), and adapts the message to establish an appropriate signal for the

Table 1 Characteristics of unstructured data vs. structured data

Characteristics		Advantage of UD	Descriptions and examples
Highly unstructured data	Highly structured data		
Nonnumeric	Numeric	Allows more flexibility for theoretical discovery	*Highly Unstructured Data*: No predefined, numerical representations of the constructs of interest, so the researcher must "assign" numeric values to the data, either manually (e.g., human coding) or with automated methods (e.g., unsupervised machine learning) before analysis can take place. Example: Highly unstructured data units are purely nonnumeric (e.g., video) requiring the researcher to assign values to the data. For example, a researcher interested in detecting customer satisfaction from UGC text data may choose to create a context-specific dictionary to determine the presence and degree of customer-expressed affect. *Highly Structured Data*: Predefined, numerical representations of the constructs of interest; in instances when it does not (e.g., categories), the researcher can easily assign numeric values to the data. Example: Extreme examples are purely numeric (e.g., net sales measured in USD), but other structured data units require just minimal researcher effort to translate the data into numeric values. For example, with surveys, the researcher determines the construct of interest (e.g., interfirm trust) and uses a scale on which participants indicate their degree of perceived interfirm trust, according to the given scale points. The participant's response will not be strictly numeric, but the researcher can easily translate the responses into numeric representations of interfirm trust.
Multifaceted	Unifaceted	Richer/deeper conceptual and managerial insights	*Highly Unstructured Data:* A single data unit has multiple facets each containing unique information. Example: A researcher analyzing highly unstructured data, such as voice data, could use acoustic software to extract unique information from the many facets the data possess including pitch, speech rate and pauses. The researcher determines which facet or combination of facets to include in the analysis, according to the research question because different prosody measures indicate distinct psychological responses according to psychology and communications literature. *Highly Structured Data*: A single data unit has one facet and therefore only contains one unique piece of information. Example: Consider the number of times customers click a banner advertisement on a website during a set period of time. Although the researcher can view the number of banner clicks from different perspectives (e.g., ad popularity, effective ad targeting), each data unit only possesses one unique facet – the number of clicks.
Concurrent Representation	Nonconcurrent Representation	Enables dynamic analysis at a given time through simultaneous capture of facets	*Highly Unstructured Data*: The unique information provided by the multiple facets of a single unstructured data unit are intertwined and thus represent different phenomena simultaneously. Example: A single data unit, such as text data, will contain many unique facets (e.g., syntax, semantics) that enable scholars to explore the presence of concurrent phenomena revealed through those facets (e.g., customer affect, persuasion). *Highly Structured Data*: Since a single unit of highly SD is unifaceted and contains only one unique piece of information, it cannot concurrently represent different phenomena. Example: Although structured data, such as weekly advertisement revenue, can flow fluidly over time, they have only one numeric facet at any given point in time.

audience (Shannon and Weaver 1949). Suitable message transmission is essential because effective information transfer requires the sender to ensure the message reflects recipients' "fields of experience" (Schramm 1954). Therefore, information transmission must be flexible because most information passes through several channels before reaching the target and rarely is delivered directly to the target or adapted in real time (e.g., personal encounter). Marketers understand the importance of approach and content adaptation during message transmission in that they communicate value through integrated marketing communications that comprise both traditional approaches (e.g., advertising, promotions) and novel methods (e.g., digital, social media, and mobile). Several studies analyze UD to investigate value communication to customers during information transmission, allowing us to identify (un)answered research questions (Table 2), gaps that warrant future investigation (Table 3), and the added value of analysis achieved through UD compared with traditional SD (Table 1).

Advertising and promotions Marketing scholars analyze UD to understand how the transmission of information can communicate value in advertisements and promotions with a "top-down" approach that assesses the overall structure of communication in relation to short and long-term objectives (Batra and Keller 2016). Empirical analyses of UD (Table 4) shed light on the dynamic nature of advertising and promotions in offline, online, and multimodal research settings (e.g., Aribarg, Pieters, and Wedel 2010; Brickman 1976; Derbaix 1995; Pieters and Wedel 2004; Teixeira and Stipp 2013; Treistman and Gregg 1979). Advertising is a powerful, controllable, firm-levered tool that can enhance the effect of positive sentiment that news media share about a firm on abnormal stock market returns, as Xiong and Bharadwaj (2013) find through natural language processing (NLP) and sentiment analysis of newspaper articles conducted with a high-speed text processing system (Lydia/TextMap). Bellman et al. (2016) use fully automated facial detection software that can distinguish "genuine smiling" to assess customer responses to advertisements and find that smile response is predictive of advertising success. In print advertising, encoded information can also be clouded by the model's appearance as shown through analysis of facial images in advertisements using supervised machine learning techniques revealing heterogeneity in consumers' face preferences, which in turn influence their brand attitudes and purchase intentions (Xiao and Ding 2014). Similarly, vocal cues in ads can determine the success of information transmission as Nelson and Schwartz (1979) find in their computer-assisted voice analysis of how actors' voice pitch affects customers' emotional responses (see Table 5). Computer-assisted text analysis[2] with linguistic inquiry and word count (LIWC) software (Pennebaker et al. 2001) that evaluates personal pronoun uses shows that message senders tend to use more negative emotion words when communicating with a single person, suggesting an other-oriented focus, but use more positive emotion words when communicating with multiple people and become self-focused (Barasch and Berger 2014). At a product campaign level, neither coupons nor rebates increase word of mouth (WOM), but product giveaways or offers of unrelated items prompt a strong positive response (Berger and Schwartz 2011). Due to the growing importance of videos in advertising, Li et al. (2018) employ supervised machine learning with convolutional neural networks (CNNs) coded in Python to analyze video data as frame by frame imagines to recognize the content and determine that video duration, visual variation, and video content all impact promotional video effectiveness. Such insights inform advertising and promotion designs for traditional communications.

In addition, paid search marketing has grown to $25.7 billion annually (Trusov et al. 2016), largely due to rising consumer trends to search for information online spurring development of research in digital marketing. With a computational linguistics text analysis approach, one study assesses the indirect effects of semantics in paid search text advertising using WordNet 2.1 and a stemming algorithm and it reveals that broader, branded search terms increase website visits through direct type-ins (Rutz et al. 2011). If they compare the potential of new versus old paid search advertisements, marketers can communicate value as gains in short-term performance. Rutz et al. (2017) demonstrate a unique approach to predict the performance and perceptions of new versus old advertisements using computer-assisted text mining (bag-of-words method) to uncover textual covariates and the latent Dirichlet algorithm (LDA), such that they identify higher-order structures in the text ads. Furthermore, advertisements that are both emotional and brand-forward can evoke positive brand-related outcomes, so firms should infuse emotions into their brand-centered online content to generate "valuable virality" (Apkinar and Berger 2017).

Extant literature also addresses the interplay of online and offline advertising on firm outcomes. Marketing scholars identify contextual advertisements with Google's AdSense algorithm, which analyzes website content on third-party websites, including image and text, and devise different campaign strategies (online banner, television ad) to reach different target markets (Lobschat et al. 2017). In their custom query text-mining algorithm designed to identify a complete list of product-related keyword searches, Joo et al. (2016) discover that television advertising improves customers' branded keyword searches, with a particularly strong effect for new companies. Research employing a novel text mining and computational linguistics approach on text data with KH Coder, a lexical software program, confirms the interaction of "traditional" and "new" media with the finding that advertising increases online user-generated content (UGC) prior to product launch but is less effective at stimulating UGC after the launch period (Onishi and Manchanda 2012). Synergy between traditional and new media can be a win–win option, considering research using human content coding of video data shows that both the advertised brand and the television show airing the ad benefit from "social TV" and the

[2] The term used by Humphreys and Wang (2018).

Table 2 Unstructured data in substantive areas in marketing

Unstructured data in substantive areas in marketing

Substantive domain	Examples of questions addressed by UD	Type of data used	Examples of questions yet to be addressed with UD	Data types warranting further study
Advertising and promotions	• Why are certain products talked about more than others and what product qualities lead to short and long-term product hype (Berger and Schwartz 2011)? • Can voice pitch used in advertisements inform customer purchases (Brickman 1980)? • Do certain types of televised advertising content (e.g., action-focused vs. emotional) influence customer online shopping (Liaukonyte et al. 2015)? • How do emotions evoked by advertisements affect brand evaluations (Pham et al. 2013)? • What is the ROI of paid online search advertising and what keywords are the most effective (Rutz et al. 2011)? • What emotions conveyed in televised advertisements encourage attention to the ad and reduce the likelihood to skip (Teixeira et al. 2012)? • How can firms predict the performance of new versus old advertisements prior to launch (Rutz et al. 2017)? • How does "new" media (e.g., UGC) interact with "traditional" media (e.g., television advertising) (Onishi and Manchanda 2012)? • Do video duration, visual variation, and video content impact the effectiveness of a promotional video (Li et al. 2018)?	• Eye tracking • Facial cues • Images • Text • Video • Voice	• How can firms effectively stimulate organic WOM for advertisements on social media platforms? • What components of an online advertising campaign on social media platforms are most essential for its virality? • What vocal characteristics are most desirable to have in advertisements for hedonic vs. utilitarian brands? • What are the consequences of having a spokesperson with vocal characteristics that align (contrast) with brand identity? • What are enabling and hindering boundary conditions that affect promotion redemption through mobile apps? • Do moderate/high levels of advertising hurt nonprofit organizations? Does it send the message that donor dollars are not being wisely spent? • Can large-scale text analysis of advertising labels determine the benefit, if any, of using influence tactics on product labels? • Can sophisticated image analysis techniques note graphic trends in the labels of popular brands (e.g., soft edges, texture)? • What language is most effective in promotional materials when attempting to win back a customer?	• Geographic • Gestural cues • Images • Text • Video • Voice
Managing channels	• How does the language on the recruitment websites of top performing franchises differ from that on sites of lower performing franchises (Zachary et al. 2011)?	• Text	• How should firms leverage knowledge of customer online activity to provide optimal product and service offerings? • Do dyad partners in successful B2B relationships adapt similar language and vocal patterns? Is it a subconscious effort to gain acceptance? • How can firms align messages sent through different channel segments to create a seamless experience for customers?	• Gestural cues • Images • Text • Video • Voice
Product	• Can analysis of UGC discussing one product's recall negatively affect other similar or related products (Borah and Tellis 2016)? • By implementing unstructured direct elicitation (UDE), can firms increase the accuracy of customers' consideration set (Ding et al. 2011)? • What is the optimum level of segment prototypicality, brand consistency and cross-segment mimicry in product design (Liu et al. 2017b)? • What is the optimum level of design complexity and prototypicality in product design (Landwehr et al. 2011)? • Do "copycat brands" try to mimic leading brands' product labels? How does it affect customers (Satomura et al. 2014)?	• Images • Text	• What components of customer praise or complaints about product features on UGC influence a firm's "next model" design? • How can framing product launches improve customer reception of unique products? • Can product attribute preferences be inferred through customers' facial and gestural cues? • How can firms use machine learning to glean insights from "unboxing" UGC videos to determine customer affect upon product launch? • What is the ROI of firm engagement of customers in product design (e.g., Dorito flavor contest)?	• Facial cues • Gestural cues • Images • Video • Voice

Table 2 (continued)

Unstructured data in substantive areas in marketing

Substantive domain	Examples of questions addressed by UD	Type of data used	Examples of questions yet to be addressed with UD	Data types warranting further study
Retail	• How can search engine data inform retailers of their brand position (Aggarwal et al. 2009)? • Can voice be used as an indicator of activation and how does activation in voice influence negotiations (Backhaus et al. 1985)? • Does an FLE's nonverbal mimicry of a customer increase desire to return to the store (Kulesza et al. 2014)? • Can nonverbal cues inform accurate clothing garment recommendations based on customer interaction with merchandise in real time (Lu et al. 2016)?	• Eye tracking • Facial cues • Gestural cues • Images • Text • Video • Voice	• How can firms maximize in-store signage to enhance current promotions? • What is the dark side of sending mobile promotions when customers are near a retailer? • Do elaborate (vs. unique) product displays in stores draw more positive customer attention? Or are simple, brand-focused displays more effective at capturing customer attention? • How can firms acquire a holistic view of customer retail preferences (e.g., products of interest, store perception) through UGC? • Can retailers create long-term value for customers by tracking individual-specific tastes online over time?	• Facial cues • Geographic • Images • Text • Video
Sales force	• Can nonverbal cues inform managers of a salesperson's ultimate sales performance (Chapple and Donald Jr. 1947)? • Can nonverbal cues provide insights into whether a sale will be made at the time of a meeting (Pennington 1968)? • Does a salesperson's voice pitch influence her or his efficacy (Peterson et al. 1995)? • Does a salesperson's display of nonverbal facial and gestural cues influence query handling effectiveness (Singh et al. 2018)? • How do social elements of a shopping experience affect customer product interactions and purchases (Zhang et al. 2014)?	• Facial cues • Gestural cues • Text • Video • Voice	• Do salespeople adapt voice to match customers in positive and neutral situations, but not in negative situations? • How does the language used in a written proposal affect customer perception of salesperson competence? • How can firms leverage existing information in CRM databases to systematically recognize opportunities for cross-selling or natural touchpoints? • How does a salesperson's appearance, which matches the brand image or not, influence efficacy? • How do customer displays of dominant vs. submissive body language influence bargaining power in a negotiation? • How can video data be used to inform salesperson efficacy during an ongoing sales interaction?	• Images • Text • Video • Voice
Service	• Can computer science techniques be used to improve customer service in online travel searches (Ghose et al. 2012)? • Can customer satisfaction be inferred from the first two minutes of a customer service call (Hall et al. 2014)? • Are service interventions on social media effective (Ma et al. 2015)? • How do nonverbal behaviors unfold during ongoing service interactions, and how do they affect service evaluations (Ma and Dubé 2011)? • Do customer nonverbal cues during a service interaction indicate overall service evaluation (Mattila and Enz 2002)? • Can virtual employees help acclimate new customers to a firm's offerings thereby increasing service delivery (Köhler et al. 2011)?	• Facial Cues • Gestural Cues • Images • Text	• How does a FLE's appearance match/mismatch with firm identity influence information encoding fluency? • Does the presence or absence of customers nearby influence FLE behavior when interacting with a customer? • Can nonverbal cues be used to gauge whether service expectations differ in for-profit vs. nonprofit contexts? • Is the interpretation of a FLE's spoken words combined with nonverbal cues different than either the words or nonverbal cues alone? • How can automated voice systems recognize customer sentiment and alter intonation accordingly in kiosks? • What FLE nonverbal facial and gestural cues enhance/hinder customer satisfaction in an international service context? • How would customers respond to the visible use of AI in a service interaction?	• Images • Text • Video • Voice

Table 2 (continued)

Unstructured data in substantive areas in marketing

Substantive domain	Examples of questions addressed by UD	Type of data used	Examples of questions yet to be addressed with UD	Data types warranting further study
User-generated content	• What are the managerial implications of "echoes" between social media and other online content (Hewett et al. 2016)? • How do content, content–user fit and user influence on a social media platform affect rebroadcasting behavior (Zhang et al. 2017)? • Can social tags in online content inform the firm's value and brand performance (Nam and Kannan 2014)? • Can a sentence-based topic model improve predictions of customer ratings online (Büschken and Allenby 2016)? • What role does expressed emotion play in the perceived helpfulness of online reviews (Yin et al. 2017)? • Can linguistic indicators in UGC accurately predict customer perceptions of product and service offerings (Tang and Guo 2015)? • Do varying levels of activation, sentiment expression and discourse patterns alter overall review sentiment (Ordenes et al. 2017)? • How do customer ratings influence subsequent product ratings and reviews (Sridhar and Srinivasan 2012)? • Does the person posing a question on an online forum drive the conversation? Or do the people responding have more influence on subsequent content (Hamilton et al. 2017)? • What impact do photos have on perceived helpfulness of an online review (Wang et al. 2018)?	• Text	• How valuable are "brand advocates" on social media platforms? Is there an interaction between influence and content shared? • Where are customers most likely to share UGC? Can firms use geographic information to target customers and encourage positive online WOM strategically? • How can firms interpret an individual's UGC from various online sources most efficiently to deploy advertising effort at the individual level? • How should firms target damage control for negative online WOM while avoiding freeriding? • What characteristics make up highly popular (viral) memes? • What is the ideal balance between edgy and politically correct content shared on social media for brands targeting younger customers? • How can firms interpret the content of images and videos that customers upload to social media websites to gauge trends in customer sentiment? • Can deep learning techniques detect patterns in voice of UGC videos to determine customer sentiment?	• Geographic • Images • Video • Voice

Select substantive areas are included in this table. Articles are from *Journal of Marketing, Journal of Marketing Research, Journal of the Academy of Marketing Science, International Journal of Research in Marketing*, and *Marketing Science*, published in 2002–2017, + select others. This table is not meant to be comprehensive

volume of online WOM increases for both elements (Fossen and Schweidel 2017).

Despite progress made in analyzing UD to answer research questions focused on communicating value during information transmission, many opportunities for continued contributions remain, as shown in Table 2 (column 4, row 1). Moving beyond simplistic analyses of UD, studies using advanced methods to analyze voice, video, and image data offer promise for both substantive and theoretical developments (Table 3, columns 3 and 4). Scholars might draw on insights from psychology that identify vocal features that influence the communication of value during message transmission, such as pitch, which listeners can use to differentiate speakers (Baumann and Belin 2010), or speech rate, which may increase attention when it is higher (Chattopadhyay et al. 2003), thereby affecting message absorption (Yokoyama and Daibo 2012). Methodological approaches from communication research using support vector machine (SVM) classifications of specific voice measures (Al-nasheri et al. 2017) could distinguish which vocal characteristics indicate memorable, distinguishable brand spokespeople. Marketing scholars also might leverage practitioner wisdom to advance research in promotions through analyses of video and image data. In 2016, the auto insurance company State Farm used customers' manual classification of images (e.g., safe driving, texting, talking on phone) taken with a dashboard camera to train two machine learning neural network models to identify distracted driving according to patterns of head and hand placement, which accurately identified the driver's actions roughly 91% of the time (Sennaar 2017). Such information ultimately should

Table 3 Conceptual framework: data, substantive and theoretical gaps

Stage of communication cycle	Substantive domain	Type of data	In the literature	Substantive focus	In the literature	Theoretical advancement by unstructured data type
Information transmission communicating value	Advertising and Promotions	Eye Tracking	✓	Advertising	✓	Eye Tracking: • Attention and Memory Theory (Wedel and Pieters 2000) • Dual Attitude Theory (Goodrich 2011) • Mere Exposure Theory (Goodrich 2011) • Moderate Incongruity Theory (Pieters and Wedel 2012) • Spreading Activation Theory (Brasel and Gips 2008) • Visual Complexity Theory (Pieters et al. 2010) Gestural Cues: • Dual Attitude Theory (Goodrich 2011) • Mere Exposure Theory (Goodrich 2011) Text: • Agency Theory (Kashmiri and Mahajan 2017) • Behavioral Consistency Theory (Kashmiri and Mahajan 2017) • Cognitive Response Theory (Xiong and Bharadwaj 2014) • Information Processing Theory (Xiong and Bharadwaj 2013) • Optimal Stimulation Level Theory (Li et al. 2018) • Persuasive Argumentation Theory (Xiong and Bharadwaj 2014) • Random Utility Theory (Rutz et al. 2017) • Social Impact Theory (Barasch and Berger 2014) • Associative Network Theory (Guitart and Hervet 2017) • Emotion Theory (Elpers et al. 2003) • Optimal Stimulation Level Theory (Li et al. 2018) • Utility Theory (Elpers et al. 2003)
		Facial Cues	✓	Financial		
		Gestural Cues	✓	International		
		Geographic		News Media		
		Images	✓	Online, General	✓	
		Text	✓	Products	✓	
		Video	✓	Salesforce		
		Voice	✓	Services		
				Retail		
				UGC	✓	
				Other	✓	
Information transmission communicating value	Digital, Social Media and Mobile	Eye Tracking		Advertising		Geographic: • Contextual Marketing Theory (Luo et al. 2014) • Characteristics Theory (Ghose et al. 2012) • Customer Utility Theory (Ghose et al. 2012) • Speech Act Theory (Ordenes et al. 2018) Text: • Characteristics Theory (Ghose et al. 2012) • Customer Utility Theory (Ghose et al. 2012) • Grounded Theory (Költringer and Dickinger 2015) • Social Identity Theory (Hewett et al. 2016) • Social Impact Theory (Colicev et al. 2018) • Speech Act Theory (Ordenes et al. 2018)
		Facial Cues		Financial	✓	
		Gestural Cues		International		
		Geographic	✓	News Media	✓	
		Images	✓	Online, General	✓	
		Text	✓	Products		
		Video		Salesforce		
		Voice		Services	✓	
				Retail		
				UGC	✓	
				Other	✓	

Table 3 (continued)

Stage of communication cycle	Substantive domain	Type of data	In the literature	Substantive focus	In the literature	Theoretical advancement by unstructured data type
Information processing delivering value	Managing Channels	Eye Tracking		Advertising		Text: • Organizational Identity Theory (Zachary et al. 2011)
		Facial Cues		Financial		
		Gestural Cues		International		
		Geographic		News Media		
		Images		Online, General	✓	
		Text	✓	Products		
		Video		Salesforce		
		Voice		Services		
				Retail		
				UGC		
				Other		
Information processing delivering value	Personal Selling and Sales Management	Eye Tracking		Advertising		Facial: • Social Cognition Theory (Singh et al. 2018) Gestural: • Social Cognition Theory (Singh et al. 2018) Text: • Social Cognition Theory (Singh et al. 2018) • Social Cognition Theory (Singh et al. 2018) • Social Proof Theory (Zhang et al. 2014) • Social Impact Theory (Zhang et al. 2014)
		Facial Cues	✓	Financial		
		Gestural Cues	✓	International		
		Geographic		News Media		
		Images		Online, General		
		Text		Products		
		Video	✓	Salesforce	✓	
		Voice	✓	Services		
				Retail	✓	
				UGC		
				Other		
Information processing delivering value	Retail Management	Eye Tracking		Advertising		Text: • Gatekeeper Theory (van Heerde et al. 2015) • Emotion Theory (Lu et al. 2016)
		Facial Cues		Financial		
		Gestural Cues	✓	International		
		Geographic		News Media		
		Images		Online, General	✓	
		Text	✓	Products		
		Video	✓	Salesforce		
		Voice	✓	Services		
				Retail	✓	
				UGC		
				Other	✓	
Information processing delivering value	Service Management	Eye Tracking		Advertising		Text: • Contrast Theory (Marinova et al. 2018) • Outsourced Regulation Theory (Marinova et al. 2018) • Role Theory (Marinova et al. 2018) • Contrast Theory (Marinova et al. 2018) • Outsourced Regulation Theory (Marinova et al. 2018) • Role Theory (Marinova et al. 2018)
		Facial Cues	✓	Financial		
		Gestural Cues	✓	International		
		Geographic		News Media		
		Images		Online, General		
		Text	✓	Products		
		Video	✓	Salesforce		
		Voice		Services	✓	
				Retail		
				UGC		
				Other		
Information extraction creating value	Marketing Intelligence for Value Creation	Eye Tracking		Advertising		Text: • Attribution Theory (Tang et al. 2014) • Consumer Information Search Theory (Marchand et al. 2017) • Communication Accommodation Theory (Ludwig et al. 2013) • Communication Theory (Song et al. 2018) • Diagnosticity of Information Theory (Hennig-Thurau et al. 2015) • Diffusion Theory (Marchand et al. 2017)
		Facial Cues		Financial	✓	
		Gestural Cues		International	✓	
		Geographic		News Media	✓	
		Images	✓	Online, General	✓	
		Text	✓	Products	✓	
		Video		Salesforce		
		Voice		Services		
				Retail		
				UGC	✓	
				Other		

Table 3 (continued)

Stage of communication cycle	Substantive domain	Type of data	In the literature	Substantive focus	In the literature	Theoretical advancement by unstructured data type
						• Human Communication Theory (Ludwig et al. 2013) • Information Theory (Godes and Mayzlin 2004) • Prospect Theory (Hennig-Thurau et al. 2015; Hsu and Lawrence 2016) • Social Influence Theory (Sridhar and Srinivasan 2012) • Speech Act Theory (Ordenes et al. 2017) • Theory of Cognitive Dissonance (Liu et al. 2016) • Theory of Information Accessibility and Influences (Liu 2006) • Theory of Weak Ties (Godes and Mayzlin 2009)
Information extraction creating value	Connecting with Customers	Eye Tracking Facial Cues Gestural Cues Geographic Images Text Video Voice	 ✓ ✓ ✓ ✓ ✓ 	Advertising Financial International News Media Online, General Products Salesforce Services Retail UGC Other	✓ ✓ ✓ ✓ ✓	Facial Cues: • Contingency Theory (Ma and Dubé 2011) • Emotion Regulation Theory (Teixeira et al. 2012) Gestural Cues: • Contingency Theory (Ma and Dubé 2011) Text: • Social Learning Theory (Köhler et al. 2011) • Emotion Regulation Theory (Teixeira et al. 2012) Voice: • Emotional Space Theory (Wang et al. 2015) • Signaling Theory (Cavanaugh et al. 2018)
Information extraction creating value	Brand Management	Eye Tracking Facial Cues Gestural Cues Geographic Images Text Video Voice	 ✓ ✓ ✓ 	Advertising Financial International News Media Online, General Products Salesforce Services Retail UGC Other	✓ ✓ 	Text: • Accessibility–Diagnosticity Theory (Borah and Tellis 2016) • Associative Network Theory (Borah and Tellis 2016) • Context Theory of Classified Learning (Nam and Kannan 2014)
Information extraction creating value	Product Management and Design	Eye Tracking Facial Cues Gestural Cues Geographic Images Text Video Voice	 ✓ ✓ 	Advertising Financial International News Media Online, General Products Salesforce Services Retail UGC Other	 ✓ ✓ 	• Categorization Theory (Liu et al. 2017b) • Two-Factor Theory (Landwehr et al. 2013) Text: • Behavioral Theory (Ding et al. 2011) • Search Theory (Dzyabura and Hauser 2018)

Articles in table from *Journal of Marketing*, *Journal of Marketing Research*, *Journal of the Academy of Marketing Science*, *International Journal of Research in Marketing* and *Marketing Science* published from 2002 to 2017 + select others

Table 4 Computational methods

Computational method	Method of analysis	In the literature	Method of analysis	In the literature	Type of data	In the literature	Example of computational method
Unsupervised machine learning	3rd Party Vendor Bag-of-Words Classification Cloud Computing Computational Linguistics Computer Vision Corneal Reflection Custom Dictionary Deep Learning Existing Dictionary Facial Recognition	 ✓ ✓	Geo-Targeting Human Coding Image Analysis Lexical Analysis LSM Neural Network NLP Semantic Analysis Sentiment Analysis Text Mining Topic Modeling Voice Analysis	 ✓ ✓ ✓ ✓ ✓	Eye Tracking Facial Cues Gestural Cues Geographic Text Voice	 ✓	Büschken and Allenby (2016) Approach: The authors extend the popular latent topic algorithm, LDA, to uncover clusters of co-occurring words within customer reviews. The authors do not provide the algorithm with input on how to form these clusters: the algorithm discovers patterns in the data on its own. One of the extended algorithms considers the fact that a single sentence is likely on one topic and constrains the LDA algorithm based on sentence. The second extended algorithm allows topics within a sentence to be “sticky” or carryover to other sentences in the review. This creates a “bag-of-sentences” approach that better reflects natural speech compared to the typical “bag-of--words” property. Hypotheses: No formal hypotheses. Propositions: The authors find that uncovering latent topics by sentence produces favorable results and can provide cleaner insights than traditional word-based approaches.
Semi-supervised machine learning	3rd Party Vendor Bag-of-Words Classification Cloud Computing Computational Linguistics Computer Vision Corneal Reflection Custom Dictionary Deep Learning Existing Dictionary Facial Recognition	 ✓	Geo-Targeting Human Coding Image Analysis Lexical Analysis LSM Neural Network NLP Semantic Analysis Sentiment Analysis Text Mining Topic Modeling Voice Analysis	 ✓ ✓	Eye Tracking Facial Cues Gestural Cues Geographic Text Voice	 ✓	Tirunillai and Tellis (2012) Approach: The authors used two algorithms to determine customer review valence. First, authors construct a training data set by manually labeling whether a word is positive or negative. Other sources (e.g., WordNet) supplement this training set. A naïve Bayesian algorithm is fed the training data and classifies reviews based on valence. A semi-parametric support vector machine (SVM) is fed the same training data. Semi-parametric means that some parameters are specified, while others are allowed to be added during the valence coding process. This mix of algorithm direction and freedom is indicative of semi-supervised machine learning. After both algorithms code valence, they “vote” on the review’s valence and researchers manually code any reviews with discrepant valence. Hypotheses: No formal hypotheses. Propositions: The volume of UGC is most predictive of stock market performance and negative UGC can negatively affect stock market performance, but positive UGC does not have much of an impact.

Table 4 (continued)

Computational method	Method of analysis	In the literature	Method of analysis	In the literature	Type of data	In the literature	Example of computational method
Supervised machine learning	3rd Party Vendor	✓	Geo-Targeting		Eye Tracking		Homburg et al. (2015) Approach: To determine customer sentiment from UGC text data, the authors first clean the data and then create training sets by manually coding the valence of a random sample of posts removing uncommon words identified. Weka machine learning software (version 3.6.6) trains a SVM based on the training set the authors developed manually. The authors conduct extensive algorithm training by using segments (10) of the training set to classify sentiment and using the estimated model to classify the "holdout sample." Ten iterations of this learning process took place. It is important to note that the training set provided the basis for the algorithm to learn what positive or negative sentiment is. Hypotheses: No formal hypotheses. Propositions: Active firm engagement on UGC is not always good since it can decrease customer sentiment at high levels. These findings are true for conversations handling functional needs and customers interested in product support, but not for conversations handling social needs or seeking inspiration/entertainment.
	Bag-of-Words		Human Coding		Facial Cues	✓	
	Classification	✓	Image Analysis	✓	Gestural Cues	✓	
	Cloud Computing		Lexical Analysis	✓	Geographic		
	Computational	✓	LSM		Images	✓	
	Linguistics		Neural Network	✓	Text	✓	
	Computer Vision	✓	NLP	✓	Video	✓	
	Corneal Reflection		Semantic Analysis	✓	Voice	✓	
	Custom Dictionary	✓	Sentiment Analysis	✓			
	Deep Learning		Text Mining	✓			
	Existing Dictionary		Topic Modeling	✓			
	Facial Recognition	✓	Voice Analysis				
Computer-assisted text analysis	3rd Party Vendor	✓	Geo-Targeting		Eye Tracking		Berger and Milkman (2012) Application: The authors collect information on news articles with a custom-built web crawler that scrapped information from the *New York Times* homepage. They then conduct automated sentiment analysis on the articles collected with LIWC, a preexisting dictionary, to count the number of positive and negative words in each article. The valence of each article is determined by difference between the percentage of positive and negative words in that article. This text mining approach searches for the exact words listed in the dictionary and does not learn from input. For example, if "happy" was in the dictionary but "happier" was not in the dictionary, "happier" would not be counted as a positive word because this variant of the keyword is not in the dictionary.
	Bag-of-Words	✓	Human Coding		Facial Cues		
	Classification	✓	Image Analysis		Gestural Cues		
	Cloud Computing		Lexical Analysis	✓	Geographic		
	Computational	✓	LSM	✓	Images	✓	
	Linguistics		Neural Network		Text		
	Computer Vision		NLP	✓	Video		
	Corneal Reflection		Semantic Analysis	✓	Voice		
	Custom Dictionary	✓	Sentiment Analysis	✓			
	Deep Learning		Text Mining	✓			
	Existing Dictionary	✓	Topic Modeling				
	Facial Recognition		Voice Analysis				

Table 4 (continued)

Computational method	Method of analysis	In the literature	Method of analysis	In the literature	Type of data	In the literature	Example of computational method
							Hypotheses: *(Not formal but listed in the text)* H_1: Positive content will be more viral than negative content. H_2: Content that evokes high arousal will be more viral than content that evokes low arousal content. Propositions: The authors find that positive content is more viral than negative content. However, high arousal content, regardless of valence (e.g., awe, anger), is more viral.
Computer-assisted facial, gestural, image & voice analysis	3rd Party Vendor Bag-of-Words Classification Cloud Computing Computational Linguistics Computer Vision Corneal Reflection Custom Dictionary Deep Learning Existing Dictionary Facial Recognition	✓ ✓ ✓	Geo-Targeting Human Coding Image Analysis Lexical Analysis LSM Neural Network NLP Semantic Analysis Sentiment Analysis Text Mining Topic Modeling Voice Analysis	 ✓ ✓	Eye Tracking Facial Cues Gestural Cues Geographic Images Text Video Voice	 ✓ ✓ ✓ ✓	Landwehr et al. (2011) Application: The authors determine objective design prototypicality and complexity of vehicles by taking professional photos of cars and using morphing software to create separate morphs of compact and premium cars. (In the interest of space, only calculation of design prototypicality is discussed). They identified 50 car features (e.g., grill, windshield) and the morphing software computed the mean position of these features across the sample of cars being studied. A warped image was created of the "most average" car in the sample. For each car in the sample, Euclidean distances are calculated between each feature of the car compared to the "most average" car—a higher score means that the car is more prototypical. Hypotheses: No formal hypotheses. Propositions: "Incorporating objective measures of design prototypicality and design complexity in sales forecasting models improves their prediction by up to 19%" (p. 416).
Fully automated	3rd Party Vendor Bag-of-Words Classification Cloud Computing Computational Linguistics Computer Vision Corneal Reflection Custom Dictionary Deep Learning Existing Dictionary Facial Recognition	 ✓	Geo-Targeting Human Coding Image Analysis Lexical Analysis LSM Neural Network NLP Semantic Analysis Sentiment Analysis Text Mining Topic Modeling Voice Analysis	✓	Eye Tracking Facial Cues Gestural Cues Geographic Text Voice	✓ ✓	Luo et al. (2014) Application: The authors use microcells located in customers' phones, with the help of the mobile industry, to determine customers' geographic location for targeting conditions (near, medium, far). Since the microcell transmits a geographic signal on its own, without any prompting, it is fully automated. Hypotheses: H_1: When targeting mobile users located at proximal distances, promotion lead time will have a negative effect on the likelihood of consumer purchases as a result of the mobile promotions.

Table 4 (continued)

<table>
<tr><th>Computational method</th><th>Method of analysis</th><th>In the literature</th><th>Method of analysis</th><th>In the literature</th><th>Type of data</th><th>In the literature</th><th>Example of computational method</th></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>H_2: When targeting mobile users located at nonproximal distances, promotion lead time will have an inverted U-shaped effect (one-day prior mobile promotions are more effective than same-day or two-day prior promotions) on the likelihood of consumer purchases as a result of the mobile promotions.
Propositions: For proximal customers, sending a mobile promotion the day of is more likely to increase sales. However, sending a mobile promotion to a nonproximal customer the day before increases sales dramatically, but this relationship is an inverted U.</td></tr>
<tr><td>Human coding</td><td>3rd Party Vendor
Bag-of-Words
Classification
Cloud Computing
Computational Linguistics
Computer Vision
Corneal Reflection
Custom Dictionary
Deep Learning
Existing Dictionary
Facial Recognition</td><td>

✓

✓</td><td>Geo-Targeting
Human Coding
Image Analysis
Lexical Analysis
LSM
Neural Network
NLP
Semantic Analysis
Sentiment Analysis
Text Mining
Topic Modeling
Voice Analysis</td><td>
✓

✓</td><td>Eye Tracking
Facial Cues
Gestural Cues
Geographic
Images
Text
Video
Voice</td><td>
✓
✓

✓
✓
✓</td><td>Ma and Dubé (2011)
Application: Trained observers sat two meters away from participants and coded dyadic verbal and nonverbal behaviors of the clients and customers across a series of two minute observation periods. The scales consisted of verbal and nonverbal behaviors that demonstrated dominance, submissiveness, agreeableness and quarrelsomeness.
Hypotheses: *(Select hypotheses included)*
H_1: In a frontline service encounter, a party's agreeable (quarrelsome) behavior elicits agreeable (quarrelsome) response and inhibits quarrelsome (agreeable) response from the other.
H_3: The co-occurrence of provider submissive behavior and client submissive behavior negatively affects the client's satisfaction.
Propositions: In general, client satisfaction is increased (decreased) when there are complimentary (anticomplementary) interactions between a frontline employee and a client during a service encounter.</td></tr>
</table>

Articles in table from *Journal of Marketing*, *Journal of Marketing Research*, *Journal of the Academy of Marketing Science*, *International Journal of Research in Marketing* and *Marketing Science* published from 2002 to 2017 + select others. * Denotes that software or vendor used is believed to conduct machine learning but it is not explicitly stated in manuscript

bolster State Farm's promotional program that rewards customers for their safe driving habits. Research also might seek more effective ways to communicate value through strategic uses of UD in customer acquisition or win-back promotions.

New theoretical insights also might emerge from well-established literature streams because the characteristics of UD allow scholars to capture new information that would be inaccessible with SD. First, UD's nonnumeric nature means that they can capture fine-grained adoption and reception trends for promotions that would be glossed over by SD-based analyses. For example, text analysis of promotional materials could indicate what types of

Table 5 Methodological techniques: definition, software and examples

Methodological techniques defined

Technique	Definition	Software examples in prior literature
Computer-assisted techniques		
Text analysis		
Bag-of-Words	A common simplifying representation in natural language processing that views a piece of text as a collection (or bag) of the words in the text that captures multiplicity while ignoring grammar and word order. This bag-of-words approach can be used for classification and subsequent machine training.	KH Coder (Onishi and Manchanda 2012)
Computational linguistics	An interdisciplinary technique drawing on insights from computer science, linguistics and psychology that employs both theoretical and applied aspects of language to analyze linguistic phenomena occurring in natural language through statistical or rule-based approaches.	KH Coder (Onishi and Manchanda 2012); WordNet 2.1 (Aggarwal et al. 2009; Rutz et al. 2011)
Custom dictionary	A list of words or phrases identified by the researcher and developed with the intention of analyzing a unique data set. Custom dictionaries can build on preexisting dictionaries, but the researcher adds additional context-specific words and phrases to capture the phenomenon of interest for that particular context (Marinova et al. 2018).	Software N/A, Examples Include: Joo et al. (2016); Marinova et al. (2018); Sridhar and Srinivasan (2012); Nam and Kannan (2014)
Lexicon-based sentiment analysis	A common method used to determine the sentiment of a given comment or piece of text that "compares words included in a comment with a labeled word list, in which each word has been scored for valence" (Tang et al. 2014, p. 47).	SentiStrength2 (Tang et al. 2014)
Linguistic Style Matching (LSM)	"The degree to which two people in a conversation coordinate by matching their word use" such that there is strong covariance between the words that each person uses on a turn-by-turn level as well as throughout the conversation (Niederhoffer and Pennebaker 2002, p. 338).	JavaScripts (Ludwig et al. 2013); LIWC (Ludwig et al. 2013)
Natural Language Processing (NLP)	A broad term for a computer science approach that seeks to interpret, recognize and understand natural language with an emphasis on interactions between humans and machines. Common NLP techniques analyze naturally occurring text data and examine syntax (e.g., part-of-speech tagging, parsing), semantics (e.g., lexical semantics) and discourse (e.g., automatic summarization). It also can be used to analyze voice data through speech recognition.	Sawtooth Software (Decker and Trusov 2010); Stanford CoreNLP (Liu et al. 2017a); Stanford Sentence and Grammatical Dependency Parser (Ordenes et al. 2017)
Ontology learning-based text mining	A text mining process frequently connected to the semantic web that identifies the core concepts in text data and uncovers "their terms, attributes, values, and relationships" within a given knowledge domain (Moon and Kamakura 2017, p. 266).	SAS Text Miner (Moon and Kamakura 2017)
Pre-existing dictionary	A list of words and/or phrases identified by someone other than the researcher used to analyze text data. Preexisting dictionaries generally are not context-specific and can be used to assess a variety of phenomena such as emotion in speech (e.g., LIWC, RDAL).*	LIWC (Berger and Milkman 2012; Hewett et al. 2016); Ludwig et al. 2013); PCNet (Ordenes et al. 2017); RDAL (Marinova et al. 2018)
Semantic text analysis	A form of text analysis in which different levels of the text (e.g., phrase, clause, sentence, paragraph) are related to the overall text, with recognition of the language-independent meanings that each level holds.	R Software (Toubia and Netzer 2017)
Sentiment analysis	Known as "opinion mining" or the "voice of the customer," sentiment analysis is an approach to determine a person's affective state with respect to a given topic through NLP, computational linguistics and other forms of text analysis. Marketing literature commonly assesses sentiment analysis by polarity, or the degree to which one expresses positive, negative or neutral affect.	Python (Hewett et al. 2016); SentiStrength (Ordenes et al. 2014; Ordenes et al. 2017; Tang et al. 2014)
Text mining	A broad term used to describe the process of deriving meaning from text data through tasks such as text categorization, clustering, summarization and concept extraction, which can be simple or more complex. Simple text mining approaches use dictionaries that count frequency distributions of the exact words in the dictionary. However, more complex forms employ machine learning to expand the list of provided words and add other relevant words, for example.	Custom dictionaries (e.g., Marinova et al. 2018); KH Coder (Onishi and Manchanda 2012); LIWC (e.g., Ordenes et al. 2017); SPSS Modeler (Ordenes et al. 2014)

Table 5 (continued)

Methodological techniques defined		
Image analysis		
Image analysis	A process to extract meaningful information from images. There are many techniques one can use to analyze (usually digital) images, which range in complexity. Each image analysis technique is well-suited for a finite set of tasks; in general, there is much work left to be done on quick and accurate image analysis.	R Software (Landwehr et al. 2013; Landwehr et al. 2013)
Image classification	A method that recognizes patterns in images and groups similar images through various techniques. For example, one approach uses contextual information in the images (relationship of nearby pixels) for categorization. Another, less granular image classification approach uses object-based image analysis (OBIA), such that groups of pixels of different shapes and scales classify the images.	Microsoft Virtual Earth Interactive SDK (Ghose et al. 2012)
Voice analysis		
Computer-assisted voice analysis	The extraction of information from speech sounds for extra-linguistic purposes. Voice analysis explores the nonverbal content of speech (e.g., pitch, speech rate) in the form of prosody and source measures. Such components of voice can be extracted with open-source software, such as Praat, or proprietary software such as LVA.	LVA (Cavanaugh et al. 2018; Duke and Amir 2018)
Video Analysis		
Computer vision	"An interdisciplinary field that deals with how computers can be made for gaining high-level understanding from digital images and videos." Computer vision employs many methods to gain information from an image or sequence of images with the goal of producing numeric or symbolic output.	MATLAB (Lu et al. 2016)
Technique	Definition	Examples in Prior Literature
Machine Learning Techniques		
Deep learning	A subset of machine learning modeled after the human brain using hidden layers in an artificial neural network that processes units of data nonlinearly to extract features. Deep learning can be supervised, semi-supervised or unsupervised.	Liu et al. (2017a)
Neural network	A class of machine learning inspired by neuron transmission in the human brain that consists of an input layer with one or more hidden layers that determine the output layer. Artificial neural networks and convolutional neural networks are common in computer science literature.	Li et al. (2018); Teixeira et al. (2012); Timoshenko and Hauser (2018); Wang et al. (2018)
Machine Learning (ML)	An umbrella term used in computer science to describe a computer's ability to learn something without explicitly being programmed to do so. It allows a computer to discover insights (e.g., patterns) that the researcher does not explicitly tell the computer to find. Various types of machine learning require more (supervised learning) or less (unsupervised learning) human guidance and offer different benefits and disadvantages.	Büschken and Allenby (2016); Lee and Bradlow (2011); Netzer et al. (2012); Tang et al. (2014); Timoshenko and Hauser (2018); Tirunillai and Tellis (2014); Zhang et al. (2017); Xiao and Ding (2014)
Supervised learning	A type of machine learning in which the computer is given labeled inputs and corresponding outputs (typically generated by the researcher) that form a "training set" which the computer uses to create a general rule for mapping inputs to outputs. The computer uses these inputs as a guide but learns to classify deviant cases on its own.	Ghose et al. (2012); Netzer et al. (2012); Tang et al. (2014); Timoshenko and Hauser (2018); Xiao and Ding (2014)
Semi-supervised learning	A type of machine learning in which the computer is given an incomplete training set that teaches it to recognize how inputs map to outputs. The researcher provides labels for some inputs/output but will not label some (or many) of the target outputs in the training set.	Tirunillai and Tellis (2012)
Unsupervised learning	A type of machine learning that does not provide any labeled examples of desired inputs/outputs in the training set and instead allows the computer to find its own structure with the data. Therefore, this method does not require human guidance to learn information from the data.	Büschken and Allenby (2016); Lee and Bradlow (2011); Liu et al. (2016); Nam et al. (2017); Timoshenko and Hauser (2018); Tirunillai and Tellis (2014); Zhang et al. (2017)

Table 5 (continued)

Methodological techniques defined		
Task-Specific Algorithms		
Machine Learning Algorithms		
Conditional Random Field Algorithm	An algorithm commonly used in machine learning and pattern recognition to structure predictions. Conditional Random Field algorithms structure predictions by considering the context (i.e., neighboring samples) to establish consistent interpretations of known relationships.	Text Data (Netzer et al. 2012)
Girvan-Newman Community Clustering Algorithm	An algorithm that employs a hierarchical method to identify communities, or network nodes that can be grouped together, in complicated systems. This algorithm clusters the communities by first removing edges of the original network, leaving connected components of the remaining network, known as communities.	Text Data (Netzer et al. 2012)
Latent Dirichlet Allocation (LDA)	An unsupervised topic modeling algorithm common in NLP that sifts through text data and identifies words and phrases indicative of a particular topic. Because LDA uses unsupervised machine learning, the researcher does not determine topics a priori and instead lets the algorithm uncover topics on its own.	Text Data (Büschken and Allenby 2016; Nam et al. 2017; Liu et al. 2017a; Tirunillai and Tellis 2014; Zhang et al. 2017)
Naïve Bayes Classifier	A popular machine learning algorithm for text categorization that assumes strong independence across features and indicates to which category each piece of text belongs.	Text Data (Chung et al., 2016; Tirunillai and Tellis 2012)
Support Vector Machine (SVM)	A supervised machine learning model used for classification and regression. The researcher provides the model with a labeled training set of examples indicating whether it belongs to a certain category. The SVM learns from the examples to categorize new and divergent cases.	Images/Facial (Ghose et al. 2012; Lu et al. 2016); Text Data (Homburg et al. 2015; Marchand et al. 2017; Tirunillai and Tellis 2012)
Viola-Jones Algorithm	A machine learning algorithm that can be trained to detect a variety of objects and that processes rapid images (e.g., from video data) to detect the object; it also can be trained to classify objects in rapid images, beyond mere detection.	Video Data (Lu et al. 2016)
Other Algorithms		
Pointwise Mutual Information Algorithm	An algorithm for feature extraction grounded in statistics and information theory that measures associations. This algorithm has been extended to computational linguistics and can classify text data on the basis of the semantic orientation of phrases, for example.	Text Data (Aggarwal et al. 2009)
Porter Stemmer Algorithm	An algorithm that reduces a word to its stem or root, such as changing “walking” to “walk.” Reducing text data to their root allows researchers to approximate word groupings according to on similar meanings.	Text Data (Rutz et al. 2011; Tirunillai and Tellis 2014; Toubia and Netzer 2017; Zhang et al. 2017

This table broadly defines select techniques mentioned in this manuscript and is not meant to be comprehensive. Unless otherwise indicated, Wikipedia served as the source for the definitions

*We constructed these definitions

language are most compelling for encouraging promotion redemption. Second, the multifaceted nature of UD can specify which combinations of nonverbal vocal cues (e.g., fast speech rate and high pitch, moderate speech rate and low pitch) are most effective for transmitting information in advertisements. Third, UD enables scholars to uncover trends across multiple facets of a data unit to understand how they dynamically interact over time. In a promotion context, researchers could analyze different facets of text data (e.g., semantics, sentiment) to model and conceptualize how UGC evolves within a thread in response to firm-generated promotional material.

Digital, social media and mobile

Lamberton and Stephen (2016) identify three phases of digital, social media and mobile (DSMM) communications in the past 15 years: initial facilitation of individual expression, shift to their use as a decision support tool and a market intelligence role. Firms increasingly are interested in communicating value by transmitting information through DSMM communications as illustrated by a recent survey that showed that CMOs anticipated spending 10.7% of their budget on social media in 2016 (Mochon et al. 2017). Understanding the interaction between firm-transmitted information on social media and other information sources is critical to adapt messaging proactively to maximize communicated value.

Traditional news media, UGC, firm-generated social media content, and press release data scraped by Python code form a complex "echoverse" that affects business outcomes as discovered with computer-assisted text analysis using LIWC software coupled with a supervised machine learning algorithm to capture consumer sentiment (Hewett et al. 2016). Part-of-speech tagging conducted with the Porter stemming algorithm focusing on noun roots and a popular unsupervised machine learning topic modeling algorithm, LDA, together provide a comparison of the content of firm-generated Tweets with each consumer's topic profile, finding that the content of the post, its fit with the user, and the user's influence on others on the platform all determine the communication of value and rebroadcasting transmission (Zhang et al. 2017). In addition to content fit, research using a custom-built web crawler to scrape text data from *The New York Times* website and analyze them with LIWC software find that consumers are more likely to engage in social transmission of positive firm-generated content and that the interaction between valence and arousal (e.g., awe) also affects content virality (Berger and Milkman 2012). Colicev et al. (2018) rely on supervised machine learning to detect sentiment on earned social media (ESM) posts with the naïve Bayes classifier algorithm, which reveals that the valence of ESM has the greatest impact on customer satisfaction. Firms also can communicate value through personalized product and service offerings online. Ghose et al. (2012) offer an improved hotel ranking system to increase customer-specific value by considering the utility gained from multiple hotel features by training a SVM classifier to categorize hotel location characteristics using satellite images, incorporating text data listing key hotel features identified with part-of-speech tagging and integrating a clustering algorithm with sentiment analysis of online customer review phrases coded by Amazon Mechanical Turk raters and a custom dictionary to handle negation phrases. Ordenes et al. (2018) are among the first to consider the dynamic interplay between firm-generated text and image content on social media through analysis of text data with supervised machine learning and NLP and human coding of images to find that certain combinations of text (e.g., expressive, directive) and image (e.g., information, action) content are more likely to encourage consumer sharing.

Mobile ad spending was predicted to reach $32.2 billion by 2017 (Danaher et al. 2015), which has sparked interest in using geographic data to understand how to communicate value to key prospects through mobile targeting. Although geographic data fall on the right panel of the unstructured–structured data continuum, we include them to reflect practitioners' push to "make the most of mobile moments" (Elliott 2014). A Forrester study conducted in June 2017 revealed that 70% of surveyed marketing decision makers plan to use mobile geographic data to enhance development of customer segments and of those individuals, 60% indicated a shifting focus to local tactics reflecting the customer's geographic location (Bernard 2017).

Luo et al. (2014) explore mobile promotion effectiveness using geographic data gathered through microcells in users' mobile phones to calculate their distance from a retailer. They find that geographical targeting increases sales, but its relationship with sales varies with temporal targeting. Sending mobile promotions to customers physically near a focal firm has strong face validity, but research also notes that a focal firm can avoid profit cannibalization by sending mobile promotions to customers near a competitor firm, thereby capturing additional customers and creating incremental sales (Fong et al. 2015). Other studies investigate the effect of the personalized adaptation of content shared with consumers on mobile devices using a modified naïve Bayesian algorithm programmed in Visual Basic that incorporates keywords from selected news articles to recommend other articles that may be of interest based on individual-specific reading behaviors (Chung et al., 2016).

Despite increased attention to DSMM in marketing literature, gaps remain and could benefit from UD analysis as shown in Table 2 (column 4, rows 1, 6 and 7). Analyses of video and voice data and further investigations of text, image, and geographic data merit scholarly attention to enhance theoretical developments (Table 3, columns 2 and 4). Batra and Keller (2016) note that research primarily has investigated the effects of firm-generated content on a single platform as opposed to assessing the overall impact of brand messaging on multiple social media sites, which may offer a more realistic evaluation of brand sentiment. Therefore, continued work could analyze firm-generated text and image data to better identify overarching brand sentiment and understand the drivers of message rebroadcasting behaviors across platforms. Communication literature using human coding and computer-assisted text analysis indicates that familiar content or content that elicits emotions is more commonly shared on social media (Kim 2015). For example, when the nonprofit ALS Association integrated emotional elicitation (calls to fight ALS) with familiarity (ice bucket) in its ALS Ice Bucket Challenge, the challenge quickly went viral, and on Instagram users uploaded 3.7 million videos, each with related social tags, which helped spur a dramatic increase in monetary support for the ALS Association compared with the prior year (Townsend 2014). Better understanding of the characteristics that create viral content can help marketing managers incorporate these qualities into their firm-generated content and thus realize financial gains.

Recent literature notes that "as the field embraces digital, social, and mobile strategies, the nature of metrics use needs to be reexamined to determine whether traditional metrics are replaced by more sensitive process measures" (Moorman and Day 2016, p. 18). In this sense, analyses of UD can unlock new insights in this domain because its characteristics enable more precise measurements. First, analysis of UD allows researchers to capture real-time phenomena that are nonnumeric, which is particularly important for research based on data that are typically in the form of text, images and video. Although academic insights into the impacts of augmented reality (AR) on customer experiences are still scarce (Hilken et al. 2017), technologically agile companies such as Apple are paving the way for innovative AR mobile apps by creating operating systems that support applications (e.g., ARKit) capable of a range of AR technologies (Chen 2017). Second, the unique information provided by UD's multiple facets enables scholars to understand how components of DSMM strategy, such as images used in mobile promotions (e.g., level of vividness, consumer–content fit), can be used to elicit customer response. Third, the concurrent representation of unique information from UD's multiple facets (e.g., vocal cues, facial cues) would allow marketing researchers to study how the flow of nonverbal cues conveyed in UGC video data influences their probability of going viral.

Information processing

Once the information reaches its target audience, cognitive and affective processing ensues and determines if the message progresses along the communication continuum or is disregarded. At this stage, the recipient has received the sender's message and must begin the process of decoding or processing the transmitted information (Shannon and Weaver 1949). Firms seek to deliver value during the information processing stage of the communication cycle and therefore need to interpret customer responses to maximize the perceived information value. Information that conflicts with a person's mental script rarely is considered due to the cognitive challenges associated with adjusting existing scripts (Shoemaker et al. 2004), whereas information that aligns and reinforces those scripts instead tends to receive a warm reception. The unpredictable nature of message transfer means that noise is introduced in the channel between the message sender and recipient, which can influence message reception and decoding (Shannon and Weaver 1949). Thus, a firm's delivery of value in this stage requires enhancing consumer information processing while reducing any introduced noise. This task can be especially challenging since novel, smart technologies are rapidly transforming customer interactions in retail, sales, and service settings (Marinova et al. 2017) and technology-mediated ad targeting enables ever-expanding consumer alternatives (Mims 2016), which increase the noise even further. Delivering value during the information processing stage of the communication cycle is particularly important for managing channels, personal selling and sales, retail, and service.

Managing channels Batra and Keller (2016) note that the customer journey has become shorter and more complex, rendering the traditional "purchase funnel" inadequate and thereby increasing the need for tight channel management. The effective management of channels can optimize the value delivered to customers during information processing because firms can "actively shape those decision journeys" and create competitive advantages through the journey itself (Edelman and Singer 2015). Branded websites seek to demonstrate value that franchisors can deliver to potential franchisees. Computer-assisted text analysis of content from franchisor recruitment websites with custom and preexisting dictionaries conducted with DICTION software reveals that high performing franchises use not only more charismatic language but also more language expressing market and entrepreneurial orientations compared with poorly performing franchises (Zachary et al. 2011).

As Table 2 (column 2, row 2) indicates, relatively little research uses UD to address questions about information processing in channel management contexts. Existing studies rely heavily on analyses of text data in select substantive areas leaving room for both substantive and theoretical development (Table 3, columns 2, 3, and 4). One avenue worthy of pursuit is user profiling, or the "summary of a user's interests and preferences revealed through the user's online activity," which may be critical for tapping the full potential of big data (Trusov et al. 2016, p. 406). Therefore, researchers might examine multiple types of UD and summarize them into user profiles to define the combination and sequence of touchpoints a customer receives through various channels. Communication accommodation theory also states that a person's motivations influence ongoing evaluations during an exchange, which alters communicative behavior, which can be calculated by the Zelig Quotient method. This method is similar to linguistic style matching (LSM) but accounts for a person's baseline use of particular words (Muir et al. 2016). Further research should investigate the impact of linguistic accommodation on the success of international business-to-business (B2B) negotiations between Western and non-Western salespeople, considering the ways that globalization has increased reliance on international channel partnerships. Recent research collects firms' 10-K text data with a webcrawler and conducts computer-assisted text analysis to show that multi-industry firms tend to operate in industries that use similar language when discussing product offerings but avoiding those with narrow, niche-centered language (Hoberg and Phillips 2018). Marketing scholars could expand on this work by investigating similarities in channel partner advertising through image content analyses.

UD's unique characteristics enable marketing scholars to address research questions pertaining to channel management in novel ways. First, the nonnumeric nature of UD allows scholars to track the customer journey dynamically in B2B relationships by analyzing text data over the course of several interactions (e.g., email exchange) to determine the presence and strength of LSM, which is indicative of conversational engagement (Niederhoffer and Pennebaker 2002). Second, as information flows to this stage of the communication cycle and is processed by the message recipient, the multifaceted nature of UD allows researchers to capture how unique facets of a data unit (e.g., syntax or semantics in firm-generated content) influence the path to purchase. Third, UD allow researchers to understand how the data facets interact with one another concurrently, such as the simultaneous adaption of syntax and semantics during an interaction to accommodate dyadic speech styles.

Personal selling and sales management Salespeople are critical boundary spanning agents charged with personally transmitting information to customers and as such, they play an important role in delivering value to customers. During an interaction, a salesperson's verbal and nonverbal cues collectively influence the customer's information processing. Scholars have long recognized the role of nonverbal cues in personal selling using machine-assisted analysis such as the Interaction Chronograph to assess department store salespeople's gestural and facial cues as gauges of future sales performance (Chapple and Donald Jr. 1947). In an effort to assess the interplay of verbal and nonverbal cues during a sales exchange, human coding of dyadic verbal and gestural cues during a business meeting predict whether the salesperson will make a sale at the time of the meeting or a later date (Pennington 1968). Human coding of a salesperson's verbal and nonverbal cues during industrial sales calls reveals that "nonverbal cues are most influential when verbal and nonverbal cues conflict" (Leigh and Summers 2002, p. 48). In addition, an examination of a salesperson's nonverbal vocal cues extends prior work by incorporating multiple nonverbal acoustic features such as pitch, pitch contour, speech rate, and pause duration to find that these measures can predict sales efficacy especially well for those with faster speech rates and more pitch contour variability (Peterson et al. 1995). Even subtle salesperson interactions with customers can influence their behavior. Human coding of video data finds that salesperson contact with customers in a store encourages interaction with the merchandise and that certain behavioral cues (e.g., walking speed) indicate whether the customer is a good sales prospect (Zhang et al. 2014). Dynamic analysis of verbal and nonverbal (facial, gestural, and body) cues of the salesperson combined with the customer's nonverbal cues reveal theoretical insights about the effectiveness of customer query handling, according to video data of business-to-consumer sales interactions (Singh et al. 2018). Therefore, UD can be used to understand customer processing and thereby inform salespeople about which actions deliver value.

As shown in Table 2 (column 2, row 5), scant research in the personal selling and sales management domain has used UD to address research questions pertaining to information processing. Therefore, analyses of video, voice, image, nonverbal (facial and gestural cues), and text data should be applied to examine remaining questions with a strong emphasis on expanding theoretical developments (Table 3, columns 2 and 4). Scholars can draw on insights from the finance literature regarding the benefits of custom dictionaries using the bag-of-words technique to capture specific phenomena through text data. For example, Loughran and McDonald (2011) find that the accuracy of valence classification of 10-Ks by pre-existing dictionaries (e.g., Harvard Psychological Dictionary) is inferior to that of custom dictionaries. Sales management could be improved with insights from unsupervised machine learning of text data from a firm's customer relationship management database, to provide guidance on optimal times for salespeople to promote cross-selling or contract renewals to individual customers. Research in psychology also might inform extant literature by providing automated dyadic analyses of nonverbal cues extracted from video data with software that detects specific body movements coded as gestural cues or adaptors in conjunction with TalkAnalyzer 1.2.5 software that detects the presence or absence of vocalizations (Fujiwara and Daibo 2014). Marketing scholars similarly might identify new boundary conditions in which certain nonverbal cues can help or hinder sales efficacy. For example, psychology literature using human coding of video data has identified patterns of speech rhythm and gaze to determine that synchronizing nonverbal cues in a dyadic interaction increases rapport (Reuzel et al. 2013) and this insight could be applied to investigate the impact of salesperson–customer nonverbal synchronization on sales. The critical role of voice in nonverbal communication (Bänziger et al. 2014), rivaled only by kinesics in communicative power (Burgoon et al. 2016), suggests the need for expanded research into voice data. Understanding ways to improve salesperson efficacy through voice could greatly influence a firm's bottom line because estimates suggest that only 1 in 50 calls that a salesperson makes results in a scheduled meeting (Olenski 2016).

The analysis of UD thus offers theoretical insights on how firms deliver value to customers during information processing in sales contexts that cannot be assessed by SD alone. Prior research notes the critical interaction between sales force efforts and other forms of marketing communications (Batra and Keller 2016), and the growing complexity and digitized nature of these interactions is not easily captured with SD. First, the nonnumeric nature of UD helps researchers understand dynamic customer message processing and emotional responses (e.g., coding facial cues) in a sales context, which can be leveraged to ensure that the salesperson continually monitors customer responses and adapts accordingly. Second, the multifaceted nature of UD allows scholars to gain more than one unique piece of information from the same data unit, such as salesperson voice, by quantitatively extracting various facets of voice data (e.g., pitch, speech rate). Third, UD analysis can capture the concurrent representation of a data unit's multiple facets, such as examination of nonverbal gestural and facial cues during sales interactions, which could shed light on whether matching or mismatching facial and gestural cues by a salesperson (e.g., smiling but with arms folded) enhances or diminishes customer information processing.

Retail management Despite the growing importance of online shopping for consumers, more than 91% of retail purchases occur offline (Gebeloff and Russell 2017), warranting more attention from marketing scholars to understand how firms deliver value through UD analyses in traditional retail settings. Early research examined voice in a capital goods bargaining context through computer-assisted analysis of voice pitch and pitch range during dyadic negotiations and find that these features of voice indicate speaker activation (Backhaus et al. 1985). Human coding of gestural cues also indicates that customers process transmitted information more positively, which results in increased satisfaction and sales, when the frontline employee (FLE) mimics the customer's gestural cues with a two-second delay (e.g., leaning forward) and when the FLE is attractive (Kulesza et al. 2014). Lu et al. (2016) employ cutting-edge computer vision techniques in MATLAB to develop a video-based automated recommender system for retail stores that personalizes clothing recommendations by comparing the customer's facial expressions and product interactions with those of other, similar customers in real time. To create this system, they used a SVM and the Viola-Jones algorithm for facial detection to detect the customer's face and recognize facial expression with an adapted facial feature extraction method that increases accuracy in a naturalistic setting. In an online retail context, brand managers can get a snapshot of their brand's overall position in the market by using computational linguistics paired with lexical text analysis to deduce the semantic orientation of online text data from search engines with a pointwise mutual information algorithm (Aggarwal et al. 2009).

Despite progress made with UD analysis to investigate research questions dealing with information processing in a retail context (Table 2, column 2, row 4), further work should explore the use of facial cues, text and images in a retail context and continue to build on past theoretical advancements (Table 3, columns 2 and 4). Women's apparel retailer Chico's FAS Inc. began tracking brand perceptions on social media platforms with text analyses of UGC that categorizes customer sentiment and identifies key influencers to determine which posts most critically require a response (Wyner 2013). Its proactive brand monitoring and attention to customer buzz could be one of the factors that helped the retailer achieve positive net income for three consecutive years (FY14–16) at a time when most retailers suffered deficits (Chico's FAS Inc. 2017). Since analysis of facial cues can provide effective monitors of customer sentiment, continued research might turn to deep learning algorithms to capture emotions expressed through facial cues in natural settings. Existing programs derived from psychology such as Ekman and Friesen's (1976) Facial Action Coding System (FACS) require a static, head-on facial image to analyze, which is unrealistic in many settings. Recent computer science contributions include a deep sharable and structural detector approach with supervised neural networks that can locate and align facial features in images despite the photograph's angle (Liu et al. 2017c). This computational innovation merits attention because despite industry's frequent use of deep neural networks, this method is rarely used in marketing research (Wedel and Kannan 2016).

UD analysis can reveal unique insights into how firms might deliver value to customers in retail contexts. First, the nonnumeric nature of UD allows researchers to assess valuable, previously unstudied aspects of information processing in retail settings. For example, analysis of video data enables researchers to objectively investigate how customers interact with the retail environment (e.g., attention to sensory information). Second, UD such as video data can capture the multifaceted characteristics of retail contexts including nonverbal cues (e.g., product consideration) and the impacts of social forces (e.g., group shopping, FLE interactions). Third, the concurrent representation of UD can offer insights about the dynamic interplay among its multiple facets (e.g., pathway movement, temporal information) to identify areas where traffic slows in a store. Such evidence could indicate ways to extend the time consumers spend in stores, which increases product interactions (Zhang et al. 2014).

Service management Delivering value to customers during information processing is crucial in service contexts, especially as the rise of smart technologies changes the nature of customized service delivery by substituting or complementing FLEs (Marinova et al. 2017), and UD analysis can unlock new insights in this area (e.g., Rafaeli, Ziklik, and Doucet 2008). In a computer-assisted text analysis approach with LIWC software that examined customer service phone call transcriptions, it was possible to infer customer satisfaction with the FLE within the first two minutes of the call (Hall et al. 2014). Dyadic human coding of FLE and customer nonverbal cues also has revealed that customers' expressed emotions during a service interaction and post-interaction mood indicate overall service evaluations, but FLEs inaccurately gauge how customers view their service performance (Mattila and Enz 2002). Human coding of video data to capture nonverbal cues in ongoing customer service interactions and computer-assisted text analysis of verbal cues provide empirical tests of custom dictionaries built off existing dictionaries (e.g., RDAL) and reveal that relational work in a service failure context should be balanced with problem solving since excessive relational work hinders complaint resolution (Marinova et al. 2018). Although firm-initiated service intervention strategies on social media are thought to improve customer relationships, a third-party proprietary text mining algorithm, which extracts and classifies the customer sentiment of firm-relevant Tweets shows that compensation for service failures actually can encourage more complaints (Ma et al. 2015). Ordenes et al. (2014) demonstrate how scholars can extract customer sentiment from online reviews of service experiences and provide an excellent roadmap for implementing linguistics-based text mining with SPSS Modeler to capture added richness in text data.

In Table 2 (column 2, row 6), we note that marketing literature has used UD in service contexts considerably more than other areas when it comes to the information processing stage of the communication cycle. Despite the relative volume of this work, most research relies on simpler methodologies (Table 3, column 2). Machine-learning–powered analyses of voice and text data offer promise for further theoretical developments (Table 3, column 4). For example, dynamic analyses of concurrent representations of UD can uncover patterns in the flow of acoustic voice features (e.g., spectral slope, decibels) during a service exchange, which can shed light on how nonverbal aspects of voice interact over time to maximize service performance. Communication research indicates that linguistic matching occurs not only in terms of the words spoken but also in with noncontent speech in which a person adjusts to the pitch and formant frequencies of a dyad partner (Hautamäki et al. 2015). This finding should be tested in a service context to determine if FLE vocal matching increases customer satisfaction. Language expectancy theory states that people develop standards for what language to use in given situations, and studies show that men are expected to use more intense language than women and that a woman's use of intense language decreases her persuasiveness (Burgoon et al. 1975). Marketing scholars might test this finding through text analysis of custom dictionaries to determine if a FLE's use of intense language during a service exchange is more or less persuasive depending on his or her gender. The firm Health Insurance Innovations seeks to improve customer satisfaction and call center efficiency by partnering with a vocal analytics company, which uses proprietary software to detect customer sentiment through voice data in real time and provide service agents with actionable feedback (CallMiner 2017).

Information extraction

The identification of information extracted from a message by a customer is beneficial because it enables marketers to detect value creation opportunities. Shoemaker et al. (2004, p. 131) note that "the context and the culture of the receiver dictate the meaning the message will have." In other words, the individual-specific perspectives that shaped the message processing stage also determine what information seems salient and therefore gets extracted by the recipient (i.e., customer). Attention is given to certain pieces of information while others are neglected in an effort to condense salient points. We build on Shannon and Weaver's (1949) theory of communication by extending the final, destination stage of the communication cycle, with what we call information extraction. The theory that drives research in this stage of the communication cycle considers the role of social influence and perspective for information extraction. Analysis of UD in this stage would provide marketing managers with unique theoretical insights into how social forces influence the information that gets extracted from the originally transmitted message. Understanding the information extraction stage of the communication cycle allows firms to create value through marketing intelligence, connecting with customers, brand management, and product management and design.

Marketing intelligence for value creation

Research seeking to create value through marketing intelligence by extracting insights from social media platforms reflects the most recent era in the digital revolution, indicative of "the connected consumer" (Lamberton and Stephen 2016). Firms create value by capturing marketing insights on social media through two primary approaches: mining UGC to "listen in" and using UGC for forecasting (e.g., Schweidel and Moe 2014; Sonnier, Mc Alister, and Rutz 2011; Toubia and Stephen 2013). Netzer et al. (2012) sparked substantial marketing research with their influential paper showcasing the value of listening in on social media to derive information about the competitive landscape through supervised machine learning of UGC text data using a conditional random field algorithm for text mining, a spring-embedded algorithm to create visual depictions of semantic networks and the Girvan-Newman community clustering algorithm to identify network clusters. Lee and Bradlow (2011) also recognize UGC's potential to assess competition through unsupervised machine learning with a K-means clustering algorithm that grouped similar topics in the text data to uncover insights on a brand's position in the competitive market. Other work has extracted customer valence and perceptions of quality expressed in UGC by extending the popular unsupervised Bayesian clustering algorithm, LDA, after applying part-of-speech tagging to the text data to hone in on certain words (e.g., nouns, verbs), which then were cleaned and refined with Porter's stemming algorithm to reduce words to their roots (Tirunillai and Tellis 2014). Yin et al. (2017) apply RDAL text analysis software to code emotional valence and arousal in online customer reviews and show that customers gain less utility from reviews with high levels of expressed emotion (see also Chen and Lurie 2013).

In addition to review text content, the presence and type of photos posted can also influence review helpfulness as Wang et al. (2018) discover through supervised machine learning of UGC images on online reviews. Further, computer-assisted text analysis with LIWC dictionaries indicates that linguistic indicators (e.g., optimism, negations) reveal the reviewer's perceptions of a product or service (Tang and Guo 2015). To examine customer sentiment, Ordenes et al. (2017) use a more fine-grained UGC text mining approach through the use of the Sanford Sentence and Grammatical Dependency, which identifies how words relate, such as an emotion-based word together with a "booster" (e.g., *very* good), to enhance existing dictionaries (LIWC, PCNet). By applying SentiStrength, they also determine that variations in language expressing activation, implicit sentiment and incoherence have differential effects on overall review sentiment. Recent enhancements of marketing insights captured from UGC using supervised and unsupervised machine learning techniques show that firms can identify customer needs at least as well, if not better, than they can through traditional approaches if they leverage text mining, NLP and CNNs that remove irrelevant information and cluster the remaining data to avoid information repetition (Timoshenko and Hauser 2018).

Along with these important marketing insights gleaned from analyses of valenced UGC, considerably less research has examined the impacts of neutral content, though prior studies note that most UGC is neutral (e.g., Ma et al. 2015). Some research considers neutral UGC language without expressed affect (e.g., Sridhar and Srinivasan 2012) whereas others consider it as having equal proportions of positive and negative language (e.g., Hsu and Lawrence 2016). Although this distinction may seem inconsequential, Tang et al. (2014) take a lexicon-based sentiment analysis approach with supervised machine learning of the existing affective dictionary from SentiStrength 2 and find that the impact of valenced UGC is enhanced by mixed neutral comments, but indifferent neutral comments reduce the consequences of valenced UGC. Researchers can use UD to capture marketing insights by considering not only *what* customers post but also *how* they interact with preexisting content. Human coding of text data from online forums shows that early responders have as much, if not more, influence on the direction of the conversation than the person who first posed the forum question because late responders tend to reiterate points made in the first few comments (Hamilton, Schlosser, and Chen 2017). Text analysis shows that reviews with explicit endorsements (e.g., overt product recommendations), as opposed to implicit recommendations (e.g., stating that the product has high quality), are more likely to result in purchase compliance because product review readers perceive the writer as having more expertise (Packard and Berger 2017). Sridhar and Srinivasan (2012) also find that customers tend to assimilate their message with previous posts in that prior product review ratings moderate subsequent product review content and ratings, according to a custom text mining algorithm that broke down each review into a bag of words and used custom dictionaries to extract UGC information.

In addition to creating value for customers by simply listening in on UGC, firms can extract valuable information to bolster their forecasting accuracy. The marketing literature primarily uses UGC to forecast specific financial outcomes, product ratings, and performance (e.g., McAlister, Sonnier, and Shively 2012; Netzer, Lemaire, and Herzenstein 2018). Liu (2006) offers an early assessment of the valence of online customer review text data through human coding and finds that the volume, but not the valence, predicts movie box office revenue. In a product recall context, Hsu and Lawrence (2016) obtain information on UGC volume, valence and growth rates from a third-party vendor (Alterian SM2 program) that tracks billions of social media interactions. They determine that the volume, valence and growth rate of online WOM surrounding a brand negatively affect firm value, but this negative effect is less pronounced for firms with strong brand equity. Analysis of text data through a semi-supervised machine learning approach improves sentiment classification, with naïve Bayesian and SVM machine learning algorithms, which in turn reveal that negative, but not positive, UGC is a strong indicator of abnormal stock returns (Tirunillai and Tellis 2012). Other research has collected data online with automated JavaScripts and investigated how consumers influence one another through LSM by employing automated text mining with LIWC dictionaries to determine that LSM increases customers' product conversion rates (Ludwig et al. 2013).

Other scholars extract information from UGC to capture marketing insights and create value by forecasting product ratings and performance. The expansion and comparison of LDA algorithm variants assessing UGC text data in an unsupervised machine learning approach extends prior work by introducing sentence-constrained LDA that allows topic "stickiness" and relates latent topic probabilities to ratings, which improves customer rating forecasts (Büschken and Allenby 2016). In a novel unsupervised text mining approach using Amazon Web Services to conduct cloud computing of billions of text files from multiple sources (e.g., Twitter, Wikipedia, Huffington Post) and Apache Mahout to reduce the topics identified in text data to manageable components, the content of Twitter posts could forecast television show ratings more accurately than other online data (Liu et al. 2016). Sentiment analysis using a supervised machine learning approach (SVM) conducted in Weka, an open-source collection of machine learning software written in Java, also supports the "Twitter effect" which suggests negative online WOM about a product considerably influences early adoption

(Hennig-Thurau et al. 2015). Additional analysis of Twitter text data from a third-party vendor (Crimson Hexagon) using a custom SVM learning algorithm to analyze Tweet sentiment indicates that the early stages of a product's lifecycle are particularly important because the timing of online WOM has varying impacts on product success (Marchand et al. 2017). As globalization shapes the world economy, analyses of UD can help marketers evaluate product performance prior to market entry. Song et al. (2018) combine human coding of movie characteristics with a supervised machine learning text mining technique to assess IMBD reviews of more than 250 movies in more than two dozen countries and find that when products are more congruent with the international market's culture, reviews are more positive.

UD analysis has focused on creating value by capturing marketing insights more so than other areas in marketing literature (Table 2, column 2, row 7). However, the emphasis on UGC text data and the scarcity of research investigating other forms of UD including images, video, and verbal and nonverbal facial and gestural cues represents a notable gap (Table 3, column 2). For example, marketing scholars could increase their use of unsupervised "fuzzy" algorithms, in which an unlabeled clustering method allows data points to belong to several clusters and is more effective than either purely supervised or unsupervised approaches (Hall et al. 1992). Such methods could recognize patterns and detect objects in UGC image data (e.g., frowning emojis) to determine brand sentiment trends. Marketing research also could explore methods to automate classification of UGC image data to inform firms about their brand's market position through approaches similar to Xiong et al.'s (2017), who use a multiclass data classifier with supervised machine learning to categorize images of faces and objects, which outperforms other common image classification methods. Despite our sophisticated understanding of fundamental concepts of image perception and communication (Arnheim 1954; Dondis 1974) and its application to digital images (Hashimoto and Clayton 2009), the implementation of these principles to analyzing image data in marketing has been limited. Continued analyses of UGC text data could employ a convolutional neural network (feed-forward deep learning) trained with multiple supervised layers, in conjunction with an extended version of the existing detector known as contrast-enhancement maximally stable extremal regions to identify text data in natural images (He et al. 2016). After the text in UGC images has been identified, scholars could develop approaches to extract text data and analyze them with common machine learning techniques. Other options to handle image data include assessments of objective visual features of an image according to local binary patterns and bag-of-words histograms as well as additional photo information (e.g., user, time) or the semantic similarity of social tags determined by WordNet (Xu et al. 2014).

Connecting with customers Connecting with customers is a critical component of value creation. Kumar and Reinartz (2016) describe value creation as a dual concept in which the firm creates value for the customer who in turn delivers value back to the firm. Analysis of UD offers a deeper understanding of how customers extract information during this stage in the communication cycle and how firms can transmit information that resonates with customers. Teixeira et al. (2012) find that firms can connect with customers through the use of emotional content (joy and surprise) in video advertisements, which increases attention and decreases their likelihood to skip the advertisement. They gleaned these insights from a Tobii 1750 infrared eye-tracker that captured eye movements, as well as existing emotion detection software that automatically categorizes emotional reactions conveyed in people's faces with a process related to Ekman's FACS using continuous video data and a Bayesian neural network classifier trained on an existing set of 50,000 facial expressions. A similar method analyzed facial cues extracted from video data to determine that video advertisements can decrease customer connection if the entertaining content appears prior to the brand information (Teixeira et al. 2014). Other research using human coding of gestural and facial cues offers evidence that boundary-spanning agents can connect with customers most effectively in a service encounter by engaging in complimentary behaviors that adjust in response to the customer's verbal and nonverbal cues (Ma and Dubé 2011). Initial research has sought to understand how firms can use voice to connect with customers. To the best of our knowledge, Wang et al. (2015) are the only marketing scholars to offer methodological innovations by analyzing voice data with custom voice analysis software, the Voice Emotion Response in Mandarin Chinese interphase, which extracts key prosody and source measures from voices and uses a weighted distance K-nearest neighbor algorithm to detect "full blown" emotions accurately 81.4% of the time.

The ubiquity of the internet allows firms to connect interactively with customers online and encourages customer engagement. The social media revolution has increased the importance of customer engagement by enabling customers to coproduce or destroy firm value (Lemon and Verhoef 2016). To engage customers online, roughly 83% of *Fortune* 500 firms used social media to connect with them by 2011 (Naylor et al. 2012). However, research into the benefits of such online engagement has produced mixed results. Analysis of UGC text data with custom dictionaries through supervised SVM learning with Weka machine learning software version 3.6.6 indicates that active engagement in customer forums initially increases customer sentiment, but experiences diminishing returns that can actually diminish customer sentiment at high levels (Homburg et al. 2015). On average, customer engagement programs negatively affect the firm because shareholders fear that these initiatives will backfire

(Beckers et al. 2018). With the growth of artificial intelligence, recent studies examine how virtual agents might connect with customers online. Research has investigated the impact of a virtual agent's interactions with online customers through custom machine learning algorithms that classify the type of interaction (e.g., functional, social, proactive) based on context-specific information and contact histories to determine that moderate levels of relational work by service agents that contains social content help newcomers acclimate to unfamiliar services (Köhler et al. 2011). The use of "virtual employees" is growing in popularity due to its potential for improving customer value and decreasing costs. In early 2017, Fukoku Mutual Life Insurance announced plans to replace 30 employees with artificial intelligence, powered by IBM Watson Explorer, a cognitive analysis platform that can identify trends in SD and UD, both internal and external to the firm. Top management anticipates that it will increase productivity by 30% (McCurry 2017).

As shown in Table 2 (column 2, rows 1 and 6), marketing scholars have used UD to investigate how firms can create value through connection with customers, but questions still remain. Studies have applied several types of UD in different substantive areas and built a strong theoretical foundation (Table 3, columns 3 and 4), but continued research could add computational rigor through newer video and image analysis techniques. Another notable gap is the persistent focus on value creation through shorter-term connections with customers as opposed to considerations of its role in customer lifetime value that accounts for individual-specific factors in a forward-looking approach (Kumar and Reinartz 2016).[3] Research also has not considered the role of the stage in the customer journey in terms of connecting with customers with UD (Lemon and Verhoef 2016), though practitioners already are making advances in this area. The home improvement retailer Lowe's sought to enhance customer experience through a partnership with Google Tango and Lenovo, creating a store-specific app that analyzes video data in real time to help customers find products on a 3D, depth-conscious navigation map that guides them to the products of interest (Lowe's 2017). Continuing work should build on prior research that manipulates sounds in an experimental setting revealing that customers associate higher pitch voices and music with small objects, but lower pitch voices and music with larger objects (Lowe and Haws 2017). An extension might find an objective measure of vocal characteristics and identify boundary conditions at which incongruent inferences about size create customer value, such as the popular "talking baby" E-Trade commercials.

Existing research has demonstrated the added depth of insights that UD analysis provides compared with SD analysis. First, understanding the antecedents and consequences of value creation through connection with customers is complex and the ability to measure subtler, nonnumeric data (e.g., emotions conveyed through facial cues) means that scholars can explore new ways to enhance customer experiences. Second, UD's multiple facets allow marketing researchers to capture numerous unique pieces of information from the same data unit, such as video data, which can reveal minute changes in different regions of the face (e.g., eyes, mouth). Marketing scholars could analyze the concurrent representation of these unique facets to explore uncovered phenomena such as how the flow and combinations of different facial cues can gauge customer affect.

Brand management Moorman and Day (2016) list brand management and the strategic leverage of brand equity as central marketing capabilities. An examination of image data with a color histogram descriptor to calculate the global color of brand labels, a dominant color descriptor to detect the dominant label color, and a texture histogram descriptor to determine shape orientation in labels reveals novel insights from these image data, including that copycat brands with packaging designs similar to the market leader's detract from the focal firm's value creation and confuse two-thirds of customers (Satomura et al. 2014). Another analysis of UD reveals dynamics of UGC that can detect brand threats and thereby prescribe brand management strategies. Using a third-party vendor that provided supervised machine learning to mine and code text data, these researchers find that the valence of UGC can create a "perverse halo" in which negative online chatter about a scrutinized brand adversely affects the brand's other product segments as well as rival brands' products in the same segment (Borah and Tellis 2016). However, negative WOM is not always harmful to brands, as Wilson et al. (2017) show with human coding of text responses that reveals that negative WOM can spur defensiveness and positive behavioral intentions if connection to the brand is high, because consumers need to affirm their identity.

Marketing research also uses UD to reveal how customers extract information from specific brand attributes, which means that practitioners can determine which attributes create value for consumers. Culotta and Cutler (2016) conduct text and data mining with custom algorithms that analyze Twitter data to determine how customers mentally cluster brand features and attributes. Liu et al.'s (2017a) sentiment analysis of Twitter text data through an unsupervised machine learning approach using the LDA algorithm in MALLET coupled with the Stanford CoreNLP sentiment toolkit indicates that despite the greater variation in customer brand sentiment, brand managers should not anticipate substantial expressions of positive sentiment on social media, because they constitute only a small fraction of brand-related Tweets (16.9%). Computer-assisted text mining with custom dictionaries of social tags adapted to the specific brands of interest suggests that strong

[3] Kumar and Reinartz (2016) provide a thorough review of studies that use SD to assess customer value, engagement, and experience, as well as key research questions moving forward.

brands should improve their category dominance, but weaker brands should work to improve their connectedness to manage brand equity (Nam and Kannan 2014). Extending this work, Nam et al. (2017) seek to define how customers evaluate brands through unsupervised machine learning with the LDA topic modeling algorithm, which depicts relationships between latent topics in customers' social tags (e.g., Apple–innovation) and the Fruchterman–Reingold graph algorithm, which creates social tag maps to inform marketers of brand perceptions. In some cases, a marketing manager may need to bring together multiple stakeholders to shape normative beliefs about brands that are radically new or controversial. Humphreys' (2010) computer-assisted text analysis of newspaper articles and press releases with LIWC software and custom dictionaries demonstrates the evolving normative conceptions guide the creation and legitimization of new markets.

Progress in this domain illustrates the enhanced quality of insights that marketing research can gain from analyses of UD compared with traditional SD. However, as Table 2 (column 3, rows 3 and 7) shows, research in this domain could benefit from more diverse types of UD such as voice and video, exploration of brand management in personal customer encounters (e.g., services) and enhanced theoretical developments (Table 3, column 4). Communication scholars create an automated synthetic vocal response system trained through unsupervised clustering of vocal source measures that adapts conveyed emotions in voice according to listener facial cues and text input to communicate an appropriate automated response (Székely et al. 2014). Similar approaches in marketing could test the feasibility of automated customer service kiosks that analyze vocal cues in real time and inform staff when FLE assistance is needed to improve customer satisfaction and maintain brand management efforts. Continued work also might extract key vocal measures with Praat software to reveal objective differences in customer emotional responses to brands belonging to the five personality types (Aaker 1997) because psychology literature reveals that vocal cues are among the most common ways people surmise the emotions of others (Juslin and Scherer 2005).

Product management and design Considerable research has analyzed UD to investigate how customers extract product information through subjective aesthetic preferences and evaluations of product attributes, which can inform practitioners about product management and design strategies. To determine product aesthetic preferences, computer-assisted image analysis conducted in R Software has computed the mean position of different points on cars and calculated the prototype similarity between morphed images to reveal that customers prefer typical designs at low exposure levels but atypical designs at high exposure levels (Landwehr et al. 2013). A similar methodological approach analyzing image data shows "that incorporating objective measures of car design prototypicality and design complexity in sales forecasting models improves their prediction by up to 19%" (Landwehr et al. 2011, p. 416). Liu et al. (2017b) adopt the same technique and conclude that consumer preferences peak for designs at average levels of brand consistency and prototypicality. These findings align well with psychology theory that posits that people prefer moderate incongruity (Mandler 1982).

Psychometric and ontology learning–based text mining techniques can achieve implicit information extraction through an approach that creates a "product positioning map" to easily identify latent topics in online product reviews (Moon and Kamakura 2017), which in turn can reveal attributes for consideration in future product designs. Marketers also can obtain aggregate consumer preferences for product attributes shared in online product reviews listing "pros" and "cons" through the creation of a custom classification algorithm that relies on text mining and NLP techniques (Decker and Trusov 2010). Recent work integrates two text analysis techniques: a computer-assisted text mining approach with the Porter stemming algorithm and a supervised machine learning method conducted in R software to build semantic networks that measure the reoccurrence of word pairs to assess the viability of generated ideas based on the balance between their novelty and familiarity (Toubia and Netzer 2017). UD analysis can also inform consideration set formation through a novel approach in an experimental setting that uses human coding of customers' desired mobile phone attributes and computer-assisted text analysis to predict their consideration set (Ding et al. 2011). Other scholars extend unsupervised machine learning recommendation system (RecSys) algorithms by considering the impact of product search on ongoing product preferences. It may be less optimal to recommend products with the highest expected utility and instead, the better choice might be to recommend products that encourage customer learning through exposure to undervalued products or products with unique attributes (Dzyabura and Hauser 2018).

As demonstrated in Table 2 (column 2, row 3), critical product management and design questions have been explored using UD (e.g., Kashmiri and Mahajan 2017), though continued work could consider additional types to expand the focus of substantive areas in which product management and design are particularly relevant (Table 3, column 3). One such gap identified in recent work refers to the need to determine the role of sophisticated versus simple product designs in customer value creation (Kumar and Reinartz 2016). Analysis of UGC video data also could offer insights on product design preferences by extracting audio from UGC in "unboxing" videos and applying IBM Watson's Speech to Text (IBM 2017) program to convert the audio into text transcriptions, which then could be classified and summarized according to the topic content with the LDA algorithm (Blei et al. 2003). UD's multiple facets flow concurrently and permit researchers to understand interactive movements over time. For example, features of UGC text data online (pronoun use, sentiment) might change before, during, and after a rival firm releases a similar product. Further work could build on Landwehr et al.'s (2011, 2013) efforts to determine boundary conditions for customers' preferences for brand typicality through analyses of product image data.

Conclusion

In 2013 IBM estimated that 2.5 quintillion bytes of data were being created each day (Bodell 2014) and 80% of business contributions to this vast amount of data take the form of UD (Nelson 2013). This vast growth of UD fuels predictions that data creation and storage will grow by 4300% between 2010 and 2020 (Hessman 2013). Analysis of UD also is reshaping business practices in many industries including insurance (Collins 2016), retail (Karolefski 2015), transportation (Stringer 2013), energy (Bodell 2014), healthcare (Appold 2017), and banking (Welsh 2017). Therefore, UD analysis and implementation eventually may have a prominent presence in every department of organizations (Hodgson 2015). In addition, UD enables practitioners to examine phenomena at more granular levels to understand the *process* by which behaviors shape business outcomes. Despite the promise that UD hold for improving value creation and exploring new business opportunities (Bodell 2014), their increased volume remains mostly untapped by firms (Collins 2016). According to a Forrester Research study, firms are only currently analyzing 12% of their available data (Smith 2015). A notable reason for this lag is that only one-quarter of firms surveyed claimed to have the internal competencies to analyze UD (Klie 2013).

The integrative framework proposed in this study addresses the nature of UD and reveals how theoretical richness and computational advancements can be gained from other disciplines. We thus make three main contributions to prior literature by (1) offering a unifying definition and conceptualization of UD in marketing; (2) bridging disjoint literature in an organizing framework that conceptualizes and synthesizes multiple subsets of UD relevant for marketing management through a review of publications in marketing and other relevant literature; and (3) identifying substantive, computational, and theoretical gaps in the literature as well as ways to leverage interdisciplinary knowledge to advance marketing research with UD in underdeveloped areas.

References

Aaker, J. L. (1997). Dimensions of brand personality. *Journal of Marketing Research, 34*(3), 347–356.

Aggarwal, P., Vaidyanathan, R., & Venkatesh, A. (2009). Using lexical semantic analysis to derive online brand positions: An application to retail marketing research. *Journal of Retailing, 85*(2), 145–158.

Al-nasheri, A., Muhammad, G., Alsulaiman, M., & Ali, Z. (2017). Investigation of voice pathology detection and classification on different frequency regions using correlation functions. *Journal of Voice, 31*(1), 3–15.

Apkinar, E., & Berger, J. (2017). Valuable Virality. *Journal of Marketing Research, 54*(2), 318–330.

Appold, K. (2017). Turn data into insight: How predictive analytics can capture revenue. *Managed Healthcare Executive, 27*(7), 16–21.

Aribarg, A., Pieters, R., & Wedel, M. (2010). Raising the BAR: Bias adjustment of recognition tests in advertising. *Journal of Marketing Research, 47*(3), 387–400.

Arnheim, R. (1954). *Art and visual perception: A psychology of the creative eye*. Berkley: University of California Press.

Backhaus, K., Meyer, M., & Stockert, A. (1985). Using voice analysis for analyzing bargaining processes in industrial marketing. *Journal of Business Research, 13*(5), 435–446.

Bänziger, T., Patel, S., & Scherer, K. R. (2014). The role of perceived voice and speech characteristics in vocal emotion communication. *Journal of Nonverbal Behavior, 38*(1), 31–52.

Barasch, A., & Berger, J. (2014). Broadcasting and narrowcasting: How audience size affects what people share. *Journal of Marketing Research, 51*(3), 286–299.

Bashir, N. Y., & Rule, N. O. (2014). Shopping under the influence: Nonverbal appearance-based communicator cues affect consumer judgments. *Psychology & Marketing, 31*(7), 539–548.

Batra, R., & Keller, K. L. (2016). Integrating marketing communications: New findings, new lessons, and new ideas. *Journal of Marketing, 80*(6), 122–145.

Baumann, O., & Belin, P. (2010). Perceptual scaling of voice identity: Common dimensions for different vowels and speakers. *Psychological Research, 74*(1), 110–120.

Beckers, S. F. M., van Doorn, J., & Verhoef, P. C. (2018). Good, better, engaged? The effect of company-initiated customer engagement behavior on shareholder value. *Journal of the Academy of Marketing Science*, In-Press. https://doi.org/10.1007/s11747-017-0539-4.

Bellman, S., Nenycz-Thiel, M., Kennedy, R., McColl, B., Larguinat, L., & Varan, D. (2016). What makes a television commercial sell? Using biometrics to identify successful ads: Demonstrating neuromeasures' potential on 100 Mars brand ads with single-source data. *Journal of Advertising Research, 57*(1), 53–66.

Berger, J., Sorensen, A. T., & Rasmussen, S. J. (2010). Positive effects of negative publicity: When negative reviews increase sales. *Marketing Science, 29*(5), 815–827.

Berger, J., & Schwartz, E. M. (2011). What drives immediate and ongoing word of mouth? *Journal of Marketing Research, 48*(5), 869–880.

Berger, J., & Milkman, K. L. (2012). What makes online content viral?? *Journal of Marketing Research, 49*(2), 192–205.

Bernard, J. (2017). Local and location-based: Combining strategies for mobile marketing maturity. In *Forbes*. Retrieved on January 24, 2018 from https://www.forbes.com/sites/forbesagencycouncil/2017/09/25/local-and-location-based-combining-strategies-for-mobile-marketing-maturity/#24356c2aae90.

Blei, D. M., Ng, A. Y., Jordan, M. I., & Lafferty, J. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 3*(4/5), 993–1022.

Bodell, T. (2014). How big data is becoming a bigger deal for the power sector. In *Electric Light & Power*. Retrieved on September 8, 2017 from http://www.elp.com/articles/print/volume-92/issue-3/columns/economic-inquiry/how-big-data-is-becoming-a-bigger-deal-for-the-power-sector.html.

Borah, A., & Tellis, G. J. (2016). Halo (spillover) effects in social media: Do product recalls of one brand hurt or help rival brands? *Journal of Marketing Research, 53*(2), 143–160.

Brasel, A. S., & Gips, J. (2008). Breaking through fast-forwarding: Brand information and visual attention. *Journal of Marketing, 72*(6), 31–48.

Brickman, G. A. (1976). Voice analysis. *Journal of Advertising Research, 16*(3), 43–48.

Brickman, G. A. (1980). Uses of voice-pitch analysis. *Journal of Advertising Research, 20*(2), 69–73.

Briggs, B., & Hodgetts, C. (2017). Tech trends 2017: An overview. In *Wall Street Journal*. Retrieved on September 5, 2017 from http://deloitte.wsj.com/cio/2017/02/08/tech-trends-2017-an-overview/.

Burgoon, M., Jones, S. B., & Stewart, D. (1975). Toward a message-centered theory of persuasion: Three empirical investigations of language intensity. *Human Communication Research, 1*(3), 240–256.

Burgoon, J. K., Guerrero, L. K., & Floyd, K. (2016). Vocalics. In J. K. Burgoon, L. K. Guerrero, & K. Floyd (Eds.), *Nonverbal Communication* (p.132–144). New York: Routledge.

Büschken, J., & Allenby, G. M. (2016). Sentence-based text analysis for customer reviews. *Marketing Science, 35*(6), 953–975.

CallMiner. (2017). Health Insurance Innovations selects CallMiner Interaction Analytics to build consumer and regulatory confidence. In *CallMiner Eureka*. Retrieved on September 11, 2017 from https://callminer.com/company/news/health-insurance-innovations-selects-callminer-interaction-analytics-build-consumer-regulatory-confidence/.

Cavanaugh, L. A., Nunes, J. C., & Han, Y. J. (2018). *Please process the signal, but don't praise it: How compliments on identity signals result in embarrassment*. Los Angeles: Working Paper, University of Southern California.

Chapple, E. D., & Donald Jr., G. (1947). An evaluation of department store salespeople by the interaction chronograph. *Journal of Marketing, 12*(2), 173–185.

Chattopadhyay, A., Dahl, D. W., Ritchie, R. J. B., & Shahin, K. N. (2003). Hearing voices: The impact of announcer speech characteristics on consumer response to broadcast advertising. *Journal of Consumer Psychology, 13*(3), 198–204.

Chen, Z., & Lurie, N. H. (2013). Temporal contiguity and negativity bias in the impact of online word of mouth. *Journal of Marketing Research, 50*(4), 463–476.

Chen, B. X. (2017). The smartphone's future: It's all about the camera. In *New York Times*. Retrieved on October 15, 2017 from https://www.nytimes.com/2017/08/30/technology/personaltech/future-smartphone-camera-augmented-reality.html.

Chico's FAS Inc. (2017). Interactive Analyst Center. In *Chico's FAS Inc*. Retrieved on September 10, 2017 from http://apps.indigotools.com/IR/IAC/?Ticker=CHS &Exchange=NYSE#.

Chung, T. S., Wedel, M., & Rust, R. T. (2016). Adaptive personalization using social networks. *Journal of the Academy of Marketing Science, 44*(1), 66–87.

Colicev, A., Malshe, A., Pauwels, K., & O'Connor, P. (2018). Improving customer mindset metrics and shareholder value through social media: The different roles of owned and earned media. *Journal of Marketing, 82*(1), 37–56.

Collins, S. (2016). Harnessing the transformative power of big data. *Milliman Inc.*, 1–6.

Coughlin, T. (2017). Analysis of dark data provides market advantages. In *Forbes*. Retrieved on September 5, 2017 from https://www.forbes.com/sites/tomcoughlin/2017/07/24/analysis-of-dark-data-provides-market-advantages/#4fe9af29872b.

Culotta, A., & Cutler, J. (2016). Mining brand perceptions from twitter social networks. *Marketing Science, 35*(3), 343–362.

Danaher, P. J., Smith, M. S., Ranasinghe, K., & Danaher, T. S. (2015). Where, when and how long: Factors that influence the redemption of mobile phone coupons. *Journal of Marketing Research, 52*(5), 710–725.

Davies, A. (2015). Why unstructured data holds the key to understanding the customer. In *My Customer*. Retrieved February 27, 2017 from http://www.mycustomer.com/. marketing/data/why-unstructured-data-holds-the-key-to-understanding-the-customer.

Decker, R., & Trusov, M. (2010). Estimating aggregate consumer preferences from online product reviews. *International Journal of Research in Marketing, 27*(4), 293–307.

Derbaix, C. M. (1995). The impact of affective reactions on attitudes toward the advertisement and the brand: A step toward ecological validity. *Journal of Marketing Research, 32*(4), 470–479.

Ding, M., Hauser, J. R., Dong, S., Dzyabura, D., Yang, Z., Chenting, S. U., & Gaskin, S. P. (2011). Unstructured direct elicitation of decision rules. *Journal of Marketing Research, 48*(1), 116–127.

Dondis, D. A. (1974). *A primer of visual literacy*. Cambridge: The MIT Press.

Duke, K., & Amir, O. (2018). *Guilt accounting theory: Consequences of temporally separating decisions and their enactment*. San Diego: Working Paper: University of California.

Dzyabura, D., & Hauser, J. R. (2018). Recommending products when customers learn their preferences. *Marketing Science*, forthcoming.

Edelman, D., & Singer, M. (2015). The new consumer journey. In *McKinsey & Company*. Retrieved on September 26, 2017 from http://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/the-new-consumer-decision-journey.

Ekman, P., & Friesen, W. V. (1976). Measuring facial movement. *Journal of Nonverbal Behavior, 1*(1), 56–75.

Elliott, S. (2014). Targeting customers on mobile during the holiday shopping season. In *New York Times*. Retrieved January 24, 2018 from https://www.nytimes.com/2014/12/03/business/media/targeting-customers-on-mobile-during-holiday-shopping-season.html.

Elpers, J., Wedel, M., & Pieters, R. (2003). Why do consumers stop viewing television commercials? Two experiments on the influence of moment-to-moment entertainment and information value. *Journal of Marketing Research, 40*(4), 437–453.

Fong, N. M., Fang, Z., & Luo, X. (2015). Geo-conquesting: Competitive locational targeting of mobile promotions. *Journal of Marketing Research, 52*(5), 726–735.

Fossen, B. L., & Schweidel, D. A. (2017). Television advertising and online word-of-mouth: An empirical investigation of social TV activity. *Marketing Science, 36*(1), 105–123.

Fujiwara, K., & Daibo, I. (2014). The extraction of nonverbal behaviors: Using video images and speech-signal analysis in dyadic conversation. *Journal of Nonverbal Behavior, 38*(3), 377–388.

Gebeloff, R., & Russell, K. (2017). How the growth of E-commerce is shifting retail jobs. In *The New York Times*. Retrieved on October 2, 2017 from https://www.nytimes.com/interactive/2017/07/06/business/ecommerce-retail-jobs.html?mcubz=1.

Ghose, A., Ipeirotis, P. G., & Li, B. (2012). Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Marketing Science, 31*(3), 493–520.

Godes, D., & Mayzlin, D. (2004). Using online conversations to study word-of-mouth conversations. *Marketing Science, 23*(4), 1–44.

Godes, D., & Mayzlin, D. (2009). Firm-created word-of-mouth communication: Evidence from a field test. *Marketing Science, 28*(4), 721–739.

Goodrich, K. (2011). Anarchy of effects? Exploring attention to online advertising and multiple outcomes. *Psychology & Marketing, 28*(4), 417–440.

Guitart, I. A., & Hervet, G. (2017). The impact of contextual television ads on online conversations: An application in the insurance industry. *International Journal of Research in Marketing, 34*, 480–498.

Hall, L. O., Bensaid, A. M., Clarke, L. P., Velthuizen, R. P., Silbiger, M. S., & Bezdek, J. C. (1992). A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain. *IEEE Transactions on Neural Networks, 3*(5), 672–682.

Hall, J. A., Verghis, P., Stockton, W., & Goh, J. X. (2014). It takes just 120 seconds: Predicting satisfaction in technical support calls. *Psychology & Marketing, 31*(7), 500–508.

Hamilton, R. W., Schlosser, A., & Chen, Y. J. (2017). Who's driving this conversation? Systematic biases in the content of online consumer discussions. *Journal of Marketing Research, 54*(4),1–16.

Hashimoto, A., & Clayton, M. (2009). *Visual design fundamentals: A digital approach* (third ed.). Boston: Course Technology CENGAGE Learning.

Hautamäki, R. G., Kinnunen, T., Hautamäki, T., & Laukkanen, A. (2015). Automatic versus human speaker verification: The case of voice mimicry. *Speech Communication, 72*, 13–31.

He, T., Huang, W., Qiao, Y., & Yao, J. (2016). Text-attentional convolutional neural network for scene text detection. *IEEE Transactions on Image Processing, 25*(6), 2529–2541.

Hennig-Thurau, T., Wiertz, C., & Feldhaus, F. (2015). Does twitter matter? The impact of microblogging word of mouth on consumers' adoption of new movies. *Journal of the Academy of Marketing Science, 43*(3), 375–394.

Hessman, T. (2013). Putting big data to work. *Industry Week, 262*(4), 14–18.

Hewett, K., Rand, W., Rust, R. T., & van Heerde, H. J. (2016). Brand buzz in the echoverse. *Journal of Marketing, 80*(3), 1–24.

Hilken, T., de Ruyter, K., Chylinski, M., Mahr, D., & Keeling, D. I. (2017). Augmenting the eye of the beholder: Exploring the strategic potential of augmented reality to enhance online service experiences. *Journal of the Academy of Marketing Science, 45*(6), 884–905.

Hoberg, G., & Phillips, G. (2018). Conglomerate industry choice and product language. *Management Science-In Press*. https://doi.org/10.1287/mnsc.2016.2693.

Ho-Dac, N. N., Carson, S. J., & Moore, W. L. (2013). The effects of positive and negative online customer reviews: Do brand strength and category maturity matter? *Journal of Marketing, 77*(6), 37–53.

Hodgson, K. (2015). What's the big deal about big data? SDM*: Security Distributing & Marketing*, 69–75.

Homburg, C., Ehm, L., & Artz, M. (2015). Measuring and managing consumer sentiment in an online community environment. *Journal of Marketing Research, 52*(5), 629–641.

Howatson, A. (2016). How to unlock the power of unstructured data. In *Marketing Tech News.* Retrieved February 27, 2017 from https://www.marketingtechnews.net/news/2016/dec/13/how-unlock-power-unstructured-data/.

Hsu, L., & Lawrence, B. (2016). The role of social media and brand equity during a product recall crisis: A shareholder value perspective. *International Journal of Research in Marketing, 33*(1), 59–77.

Huang, M., & Rust, R. T. (2017). Technology-driven service strategy. *Journal of the Academy of Marketing Science, 45*(6), 906–924.

Humphreys, A. (2010). Megamarketing: The creation of markets as a social process. *Journal of Marketing, 74*(2), –1, 19.

Humphreys, A., & Wang, R. J. H. (2018). Automated text analysis for consumer research. *Journal of Consumer Research*, *In-Press*. https://doi.org/10.1093/jcr/ucx104.

IBM. (2017). IBM Watson Speech to Text. Retrieved on July 30, 2017 from https://www.ibm.com/watson/services/speech-to-text/.

Joo, M., Wilbur, K. C., & Zhu, Y. (2016). Effects of TV advertising on keyword search. *International Journal of Research in Marketing, 33*(3), 508–523.

Juslin, P. N., & Scherer, K. R. (2005). Vocal expression of affect. In J. A. Harrigan, R. Rosenthal, & K. R. Scherer (Eds.), *The New Handbook of Methods in Nonverbal Behavior Research* (pp. 65–135). New York: Oxford University Press.

Karolefski, J. (2015). Accepting the BIG DATA challenge. *Progressive Grocer, 94*(7), 142–146.

Kashmiri, S., & Mahajan, V. (2017). Values that shape marketing decisions: Influence of chief executive officers' political ideology on innovation propensity, shareholder value, and risk. *Journal of Marketing Research, 54*(2), 260–278.

Kashmiri, S., Nicol, C. D., & Arora, S. (2017). Me, myself, and I: Influence of CEO narcissism on firms' innovation strategy and the likelihood of product-harm crises. *Journal of the Academy of Marketing Science*, In-Press. https://doi.org/10.1007/s11747-017-0535-8.

Kim, H. S. (2015). Attracting views and going viral: How message features and news-sharing channels affect health news diffusion. *Journal of Communication, 65*(3), 512–534.

Klie, L. (2013). Getting closer to customers tops big data agenda. *CRM Magazine, 17*(1), 15–15.

Köhler, C. F., Rohm, A. J., de Ruyter, K., & Wetzels, M. (2011). Return on interactivity: The impact of online agents on newcomer adjustment. *Journal of Marketing, 75*(2), 93–108.

Költringer, C., & Dickinger, A. (2015). Analyzing destination branding and image from online sources: A web content mining approach. *Journal of Business Research, 68*(9), 1836–1843.

Kostyra, D. S., Reiner, J., Natter, M., & Klapper, D. (2016). Decomposing the effects of online customer reviews on brand, price, and product attributes. *International Journal of Research in Marketing, 33*(1), 11–26.

Kotler, P., & Keller, K. L. (2015). *Marketing Management*. Upper Saddle River: Prentice Hall.

Kozinets, R. V., de Valck, K., Wojnicki, A.C., & Wilner, S. J. S. (2010). Networked narratives: Understanding word-of-mouth marketing in online communities. Journal of Marketing, 74(2), 71–89.

Kulesza, W., Szypowska, Z., Jarman, M. S., & Dolinski, D. (2014). Attractive chameleons sell: The mimicry-attractiveness link. *Psychology & Marketing, 31*(7), 549–561.

Kumar, V., & Reinartz, W. (2016). Creating enduring customer value. *Journal of Marketing, 80*(6), 36–68.

Lamberton, C., & Stephen, A. T. (2016). A thematic exploration of digital, social media, and mobile marketing: Research evolution from 2000 to 2015 and an agenda for future inquiry. *Journal of Marketing, 80*(6), 146–172.

Landwehr, J. R., Labroo, A. A., & Herrmann, A. (2011). Gut liking for the ordinary: Incorporating design fluency improves automobile sales forecasts. *Marketing Science, 30*(3), 416–429.

Landwehr, J. R., Wentzel, D., & Herrmann, A. (2013). Product design for the long run: Consumer responses to typical and atypical designs at different stages of exposure. *Journal of Marketing, 77*(5), 92–107.

Lee, T. Y., & Bradlow, E. T. (2011). Automated marketing research using online customer reviews. *Journal of Marketing Research, 48*(5), 881–894.

Leigh, T. W., & Summers, J. O. (2002). An initial evaluation of industrial buyer's impressions of salespersons' nonverbal cues. *Journal of Personal Selling and Sales Management, 22*(1), 41–53.

Lemon, K. N., & Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. *Journal of Marketing, 80*(6), 69–96.

Li, X., Shi, M., & Wang, X. (2018). Video Mining: Measuring Visual Information Using Automatic Methods. *Working paper, Ivey Business School.*

Liaukonyte, J., Teixeira, T., & Wilbur, K. C. (2015). Television advertising and online shopping. *Marketing Science, 34*(3), 311–330.

Liu, Y. (2006). Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing, 70*(3), 74–89.

Liu, X., Singh, P. V., & Srinivasan, K. (2016). A structured analysis of unstructured big data by leveraging cloud computing. *Marketing Science, 35*(3), 363–388.

Liu, X., Burns, A. C., & Hou, Y. (2017a). An investigation of brand-related user-generated content on Twitter. *Journal of Advertising, 46*(2), 236–247.

Liu, Y., Li, K. J., Chen, H., & Balachander, S. (2017b). The effects of products' aesthetic design on demand and marketing-mix effectiveness: the role of segment prototypicality and brand consistency. *Journal of Marketing, 81*(1), 83–102.

Liu, H., Lu, J., Feng, J., & Zhou, J. (2017c). Learning deep shareable and structural detectors face alignment. *IEEE Transactions on Image Processing, 26*(4), 1666–1678.

Lobschat, L., Osinga, E. C., & Reinartz, W. J. (2017). What happens online stays online? – Segment-specific online and offline effects of banner advertisements. *Journal of Marketing Research, 54*(6), 901–913.

Lohr, S. (2012). The age of big data. In *New York Times*. Retrieved on September 6, 2017 from http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html?mcubz=1.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance, 66*(1), 35–65.

Lowe, M. L., & Haws, K., L. (2017). Sounds big: The effects of acoustic pitch on product perceptions. Journal of Marketing Research, 54(2), 331–346.

Lowe's (2017). Lowe's introduces in-store navigation using augmented reality. In *Lowe's*. Retrieved on September 10, 2017 from https://newsroom.lowes.com/news-releases/lowesintroducesin-storenavigationusingaugmentedreality/.

Lu, S., Xiao, L., & Ding, M. (2016). A video-based automated recommender (VAR) system for garments. *Marketing Science, 35*(3), 484–510.

Ludwig, S., de Ruyter, K., Friedman, M., Brüggen, E. C., Wetzels, M., & Pfann, G. (2013). More than words: The influence of affective content and linguistic style matches in online reviews on conversion rates. Journal of Marketing, 77(1), 87–103.

Luo, X., Andrews, M., Fang, Z., & Phang, C. W. (2014). Mobile Targeting. *Marketing Science, 60*(7), 1738–1756.

Ma, Z., & Dubé, L. (2011). Process and outcome interdependency in frontline service encounters. *Journal of Marketing, 75*(3), 83–98.

Ma, L., Sun, B., & Kekre, S. (2015). The squeaky wheel gets the grease–An empirical analysis of customer voice and firm intervention on Twitter. *Marketing Science, 34*(5), 627–645.

MacInnis, D. (2011). A framework for conceptual contributions in marketing. *Journal of Marketing, 75*(4), 136–154.

Mandler, G. (1982). The structure of value: Accounting for taste. In H. Margaret, S. Clarke &. S.T. Fiske (Eds). *Affect and Cognition: The 17th Annual Carnegie Symposium on Cognition*, pp. 3–36. Hillsdale: Lawrence Erlbaum.

Marchand, A., Hennig-Thurau, T., & Wiertz, C. (2017). Not all digital word of mouth is created equal: Understanding the respective impact of consumer reviews and microblogs on new product success. *International Journal of Research in Marketing, 34*(2), 336–354.

Marinova, D., de Ruyter, K., Huang, M. H., Mueter, M. L., & Challagalla, G. (2017). Getting smart: Learning from technology-empowered frontline interactions. *Journal of Service Research, 20*(1), 29–42.

Marinova, D., Singh, S., & Singh, J. (2018). Frontline problem-solving interactions: A dynamic analysis of verbal and nonverbal cues. *Journal of Marketing Research, 55*:178–192. https://doi.org/10.1509/jmr.15.0243.

Marr, B. (2017). The complete beginner's guide to big data in 2017. In *Forbes*. Retrieved on. September 5, 2017 from https://www.forbes.com/sites/bernardmarr/2017/03/14/the-complete-beginners-guide-to-big-data-in-2017/#1ce04b97365a.

Mattila, A. S., & Enz, C. A. (2002). The role of emotions in service encounters. *Journal of Service Research, 4*(4), 268.

McAlister, L., Sonnier, G., & Shively, T. (2012). The relationship between online chatter and firm value. *Marketing Letters, 23*(1), 1–12.

McCurry, J. (2017). Japanese company replaces office workers with artificial intelligence. In *The Guardian*. Retrieved on October 22, 2017 from https://www.theguardian.com/technology/2017/jan/05/japanese-company-replaces-office-workers-artificial-intelligence-ai-fukoku-mutual-life-insurance.

Mims, C. (2016). Why are there more consumer goods than ever before? In *Wall Street Journal.* Retrieved August 27, 2017 from https://www.wsj.com/articles/why-there-are-more-consumer-goods-than-ever-1461556860.

Mochon, D., Johnson, K., Schwartz, J., & Ariely, D. (2017). What are likes worth? A Facebook page field study. *Journal of Marketing Research, 54*(2), 306–317.

Moon, S., & Kamakura, W. A. (2017). A picture is worth a thousand words: Translating product reviews into a product positioning map. *International Journal of Research in Marketing, 34*(1), 265–285.

Moorman, C., & Day, G. S. (2016). Organization for marketing excellence. *Journal of Marketing, 80*(6), 6–35.

Muir, K., Joinson, A., Cotterill, R., & Dewdney, N. (2016). Characterizing the linguistic chameleon: Personal and social correlates of linguistic style accommodation. *Human Communication Research, 42*(3), 462–484.

Nair, R., & Narayanan, A. (2012). Benefitting from big data: Leveraging unstructured data capabilities for competitive advantage. In *Booz & Company.* Retrieved April 19, 2017 from https://www.strategyand.pwc.com/media/file/Stategyand_Benefiting-from-BigData.pdf.

Nam, H., & Kannan, P. K. (2014). The informational value of social tagging networks. *Journal of Marketing, 78*(4), 21–40.

Nam, H., Joshi, Y. V., & Kannan, P. K. (2017). Harvesting brand information from social tags. *Journal of Marketing, 81*(4), 88–108.

Naylor, R. W., Lamberton, C. P., & West, P. M. (2012). Beyond the "like" button: The impact of mere virtual presence on brand evaluations and purchase intentions in social media settings. *Journal of Marketing, 76*(6), 105–120.

Nelson, M. (2013). The big picture on big data. *Intermedia, 41*(1), 12–16.

Nelson, R. G., & Schwartz, D. (1979). Voice-pitch analysis. *Journal of Advertising Research, 19*(5), 55–59.

Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market- structure surveillance through text mining. *Marketing Science, 31*(3), 521–543.

Netzer, O., Lemaire, A., & Herzenstein, M. (2018). When words sweat: Identifying signals for loan default in the text of loan applications. *Working Paper*, Columbia Business School, Columbia University.

Niederhoffer, K. G., & Pennebaker, J. W. (2002). Linguistic style matching in social interaction. *Journal of Language and Social Psychology, 21*(4), 337–360.

Olenski, S. (2016). How inbound marketing killed cold calling. In *Forbes*. Retrieved on July 2, 2017 from https://www.forbes.com/sites/steveolenski/2016/06/30/how-inbound-marketing-killed-cold-calling/#4979d86da71f.

Onishi, H., & Manchanda, P. (2012). Marketing activity, blogging and sales. *International Journal of Research in Marketing, 29*(3), 221–234.

Ordenes, F. V., Theodoulidis, B., Burton, J., Gruber, T., & Zaki, M. (2014). Analyzing customer experience feedback using text mining: A linguistics-based approach. *Journal of Service Research, 17*(3), 278–295.

Ordenes, F. V., Ludwig, S., de Ruyter, K., Grewal, D., & Wetzels, M. (2017). Unveiling what is written in the stars: Analyzing explicit, implicit, and discourse patterns of sentiment in social media. *Journal of Consumer Research, 43*(6), 875–894.

Ordenes, F. V., Grewal, D., Ludwig, S., de Ruyter, K., Mahr, D., & Wetzels, M. (2018). Cutting through content clutter: How speech and image acts drive consumer sharing of social media brand messages. *Journal of Consumer Research*. https://doi.org/10.1093/jcr/ucy032.

Packard, G., & Berger, J. (2017). How language shapes word of mouth's impact. *Journal of Marketing Research, 54*(4), 572–588.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic Inquiry and Word Count: LIWC 2001. [computer software]*. Mahwah: Lawrence Erlbaum Associates.

Pennington, A. L. (1968). Customer-salesman bargaining behavior in retail transactions. *Journal of Marketing Research, 5*(3), 255–262.

Peterson, R. A., Cannito, M. P., & Brown, S. P. (1995). An exploratory investigation of voice characteristics and selling effectiveness. *Journal of Personal Selling & Sales Management, 15*(1), 1–15.

Pham, M. T., Geuens, M., & De Pelsmacker, P. (2013). The influence of ad-evoked feelings on brand evaluations: Empirical generalizations from consumer responses to more than 1000 TV commercials. *International Journal of Research in Marketing, 30*(4), 383–394.

Pieters, R., & Wedel, M. (2004). Attention capture and transfer in advertising: Brand, pictorial, and text-size effects. *Journal of Marketing, 68*(2), 36–50.

Pieters, R., Wedel, M., & Batra, R. (2010). The stopping power of advertising: Measures and effects of visual complexity. *Journal of Marketing, 74*(5), 48–60.

Pieters, R., & Wedel, M. (2012). Ad gist: Ad communication in a single eye fixation. *Marketing Science, 31*(1), 59–73.

Rafaeli, A., Ziklik, L., & Doucet, L. (2008). The impact of call center employees' customer orientation behaviors on service quality. *Journal of Service Research, 10*(3), 239–255.

Reuzel, E., Embregts, P. J. C. M., Bosman, A. T. M., Cox, R., van Nieuwenhuijzen, M., & Jahoda, A. (2013). Conversational synchronization in naturally occurring settings: A recurrence-based analysis of gaze directions and speech rhythms of staff and clients with intellectual disability. *Journal of Nonverbal Behavior, 37*(4), 281–305.

Rizkallah, J. (2017). The big (unstructured) data problem. In *Forbes*. Retrieved on September 5, 2017 from https://www.forbes.com/sites/forbestechcouncil/2017/06/05/the-big-unstructured-data-problem/#274fefa9493a.

Rutz, O. J., Trusov, M., & Bucklin, R. E. (2011). Modeling indirect effects of paid search advertising: Which keywords lead to more future visits? *Marketing Science, 30*(4), 646–665.

Rutz, O. J., Sonnier, G. P., & Trusov, M. (2017). A new method to aid copy testing of paid search text advertisements. *Journal of Marketing Research, 54*(6), 885–900.

Satomura, T., Wedel, M., & Pieters, R. (2014). Copy alert: A method and metric to detect visual copycat brands. *Journal of Marketing Research, 51*(1), 1–13.

Schramm, W. (1954). *How communication works*. In W. Schramm (Ed.), *The process and effects of mass communication*. Urbana: The University of Illinois Press.

Schweidel, D. A., & Moe, W. W. (2014). Listening in on social media: A joint model of sentiment and venue format choice. *Journal of Marketing Research, 51*(4), 387–402.

Sennaar, K. (2017). How America's top 4 insurance companies are using machine learning. In *Tech Emergence*. Retrieved on September 11, 2017 from https://www.techemergence.com/machine-learning-at-insurance-companies/.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: The University of Illinois Press.

Shoemaker, P. J., Tankard, J. W., & Lasorsa, D. L. (2004). *How to build social science theories*. Thousand Oaks: Sage Publications, Inc..

Singh, J., Marinova, D., & Singh, S. (2018). Customer query handling in sales interactions. *Journal of the Academy of Marketing Science*, forthcoming.

Smith, K. (2015). Big data discoveries. *Best's Review, 7*, 53–56.

Song, R., Moon, S., Chen, H. A., & Houston, M. B. (2018). When marketing strategy meets culture: The role of culture in product evaluations. *Journal of the Academy of Marketing Science, In-Press*. https://doi.org/10.1007/s11747-017-0525-x.

Sonnier, G. P., McAlister, L., & Rutz, O. J. (2011). A dynamic model of the effect of online communications on firm sales. *Marketing Science, 30*(4), 702–716.

Sridhar, S., & Srinivasan, R. (2012). Social influence effects in online product ratings. *Journal of Marketing, 76*(5), 70–88.

Stringer, B. (2013). Big data in the chemical industry. *ICIS Chemical Business, 284*(20), 2–2.

Székely, E., Ahmed, Z., Hennig, S., Cabral, J. P., & Carson-Berndsen, J. (2014). Predicting synthetic voice style from facial expressions. An application for augmented conversations. *Speech Communication, 57*, 63–75.

Tang, T., Fang, E., & Feng, W. (2014). Is neutral really neutral? The effects of neutral user- generated content on product sales. *Journal of Marketing, 78*(4), 41–58.

Tang, C., & Guo, L. (2015). Digging for gold with a simple tool: Validating text mining in studying electronic word-of-mouth (eWOM) communication. *Marketing Letters, 26*(1), 67–80.

Teixeira, T., Wedel, M., & Pieters, R. (2012). Emotion-induced engagement in internet video advertisements. *Journal of Marketing Research, 49*(2), 144–159.

Teixeira, T. S., & Stipp, H. (2013). Optimizing the amount of entertainment in advertising: What's so funny about tracking reactions to humor? *Journal of Advertising Research, 53*(3), 286–296.

Teixeira, T., Picard, R., & el Kaliouby, R. (2014). Why, when, and how much to entertain consumers in advertisements? A web-based facial tracking field study. *Marketing Science, 33*(6), 809–827.

Timoshenko, A., & Hauser, J. R. (2018). Identifying customer needs from user-generated content. *Marketing Science*, forthcoming.

Tirunillai, S., & Tellis, G. J. (2012). Does chatter really matter? Dynamics of user-generated content and stock performance. *Marketing Science, 31*(2), 198–215.

Tirunillai, S., & Tellis, G. J. (2014). Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent Dirichlet allocation. *Journal of Marketing Research, 51*(4), 463–479.

Toubia, O., & Stephen, A. T. (2013). Intrinsic vs. image-related utility in social media: Why do people contribute content to Twitter? *Marketing Science, 32*(3), 369–392.

Toubia, O., & Netzer, O. (2017). Idea generation, creativity, and prototypicality. *Marketing Science, 36*(1), 1–20.

Townsend, L. (2014). How much has the ice bucket challenge achieved? In *BBC*. Retrieved on July 1, 2017 from http://www.bbc.com/news/magazine-29013707.

Treistman, J., & Gregg, J. P. (1979). Visual, verbal, and sales responses to print ads. *Journal of Advertising Research, 19*(4), 41.

Trusov, M., Ma, L., & Jamal, Z. (2016). Crumbs of the cookie: User profiling in customer-base analysis and behavioral targeting. *Marketing Science, 35*(3), 405–426.

van Heerde, H. J., Gijsbrechts, E., & Pauwels, K. (2015). Fanning the flames? Flow media coverage of a price war affects retailers, consumers, and investors. *Journal of Marketing Research, 52*(5), 674–693.

Wang, W. C., Chien, C. S., & Moutinho, L. (2015). Do you really feel happy? Some implications of voice emotion response in Mandarin Chinese. *Marketing Letters, 26*(3), 391–409.

Wang, X., Li, X., Goldenberg, J., Muchnik, L. (2018). One picture is worth 253 characters: Using photo mining to understand the role of a camera in online word of mouth. *Working paper, Ivey Business School.*

Wedel, M., & Pieters, R. (2000). Eye fixations on advertisements and memory for brands: A model and findings. *Marketing Science, 19*(4), 297–312.

Wedel, M., & Kannan, P. K. (2016). Marketing analytics for data-rich environments. *Journal of Marketing, 80*(6), 97–121.

Welsh, A. (2017). Unstructured content: An untapped fuel source for AI and machine learning. In *Developer*. Retrieved on September 10, 2017 from http://www.developer.com/db/unstructured-content-an-untapped-fuel-source-for-ai-and-machine-learning.html.

Wikipedia. (2017). Unstructured data. Retrieved October 29, 2017 from https://en.wikipedia.org/wiki/Unstructured_data.

Wilson, A., Giebelhausen, M., & Brady, M. (2017). Negative word of mouth can be a positive for consumers connected to the brand. *Journal of the Academy of Marketing Science, 45*(4), 534–547.

Wyner, G. (2013). Data, data everywhere. *Marketing News, 47*(3), 18–19.

Xiao, L., & Ding, M. (2014). Just the faces: Exploring the effects of facial features in print advertising. *Marketing Science, 33*(3), 338–352.

Xiong, G., & Bharadwaj, S. (2013). Asymmetric roles of advertising and marketing capability in financial returns to news: Turning bad into good and good into great. *Journal of Marketing Research, 50*(6), 706–724.

Xiong, G., & Bharadwaj, S. (2014). Prerelease buzz evolution patterns and new product performance. *Marketing Science, 33*(3), 401–421.

Xiong, H., Yu, W., Yang, X., Swamy, M. N. S., & Yu, Q. (2017). Learning the conformal transformation kernel for image recognition. *IEEE Transactions on Neural Networks and Learning Systems, 28*(1), 149–163.

Xu, Z., Zhang, Y., & Cao, L. (2014). Social image analysis from a non-IID perspective. *IEEE Transactions on Multimedia, 16*(7), 1986–1998.

Yadav, M. S., Prabhu, J. C., & Chandy, R. K. (2007). Managing the future: CEO attention and innovation outcomes. *Journal of Marketing, 71*(4), 84–101.

Yin, D., Bond, S. D., & Zhang, H. A. N. (2017). Keep your cool or let it out: Nonlinear effects of expressed arousal on perceptions of consumer reviews. *Journal of Marketing Research, 54*(3), 447–463.

Yokoyama, H., & Daibo, I. (2012). Effects of gaze and speech rate on receivers' evaluations of persuasive speech. *Psychological Reports, 110*(2), 663–676.

Zachary, M. A., McKenny, A. F., Short, J. C., Davis, K. M., & Wu, D. (2011). Franchise branding: An organizational identity perspective. *Journal of the Academy of Marketing Science, 39*(4), 629–s645.

Zhang, X., Li, S., Burke, R. R., & Leykin, A. (2014). An examination of social influence on shopper behavior using video tracking data. *Journal of Marketing, 78*(5), 24–41.

Zhang, Y., Moe, W. W., & Schweidel, D. A. (2017). Modeling the role of message content and influencers in social media rebroadcasting. *International Journal of Research in Marketing, 34*(1), 100–119.