© *Academy of Management Annals*
2019, Vol. 13, No. 2, 586–632.
https://doi.org/10.5465/annals.2017.0099

TOPIC MODELING IN MANAGEMENT RESEARCH: RENDERING NEW THEORY FROM TEXTUAL DATA

TIMOTHY R. HANNIGAN
University of Alberta

RICHARD F. J. HAANS
Erasmus University

KEYVAN VAKILI
London Business School

HOVIG TCHALIAN
Claremont Graduate University

VERN L. GLASER
MILO SHAOQING WANG
University of Alberta

SARAH KAPLAN
University of Toronto

P. DEVEREAUX JENNINGS[1]
University of Alberta

Increasingly, management researchers are using topic modeling, a new method borrowed from computer science, to reveal phenomenon-based constructs and grounded conceptual relationships in textual data. By conceptualizing topic modeling as the process of rendering constructs and conceptual relationships from textual data, we demonstrate how this new method can advance management scholarship without turning topic modeling into a black box of complex computer-driven algorithms. We begin by comparing features of topic modeling to related techniques (content analysis, grounded theorizing, and natural language processing). We then walk through the steps of rendering with topic modeling and apply rendering to management articles that draw on topic modeling. Doing so enables us to identify and discuss how topic modeling has advanced management theory in five areas: detecting novelty and emergence, developing inductive classification systems, understanding online audiences and products, analyzing frames and social movements, and understanding cultural dynamics. We conclude with a review of new topic modeling trends and revisit the role of researcher interpretation in a world of computer-driven textual analysis.

We would like to thank the editors of the Academy of Management Annals for their support and helpful comments. We also thank the participants in our various topic modeling presentations and reviewers and division organizers (specifically Peer Fiss and Renate Meyer) at the Academy of Management meetings. In addition, we would like to recognize Marc-David Seidel and Christopher Steele from the Interpretive Data Science (IDeaS) group for their role in germinating these ideas, Mike Pfarrer for his comments on a later draft of the paper, and Kara Gehman for her fine-grained edits on next-to-final drafts. Finally, we wish to express our appreciation to our life partners for not only putting up with but also actively discussing this paper as it evolved.

[1] Corresponding author.

INTRODUCTION

New methods can have profound impacts on management scholarship (Arora, Gittelman, Kaplan, Lynch, Mitchell, & Siggelkow, 2016), as they enable scholars to take fresh approaches to theory and reexamine previously intractable problems and old questions (Timmermans & Tavory, 2012). For example, the introduction of event history analysis helped advance both population ecology (Hannan & Carroll, 1992) and institutional analysis (Tolbert & Zucker, 1996) research; the introduction of the case comparison method aided the development of strategy process research (Eisenhardt, 1989); and the introduction of set theoretic methods and qualitative comparative analysis led to renewed investigations of configurations (Fiss, 2007; Ragin, 2008). Recently, the management field's understandings of cognition, meaning, and interpretation have been dramatically reshaped by the emergence of new computer-based language processing techniques (DiMaggio, 2015), which have amplified and sharpened the linguistic turn in management research (Alvesson & Kärreman, 2000a, 2000b). In our review, we focus on one of the most commonly used new techniques: topic modeling.

During the last decade, social scientists have increasingly used topic modeling to analyze textual data. Borrowed from computer science, this method involves using algorithms to analyze a corpus (a set of textual documents) to generate a representation of the latent topics discussed therein (Mohr & Bogdanov, 2013; Schmiedel, Müller, & vom Brocke, 2018). It has helped scholars unpack conundrums in management theory, such as how critics' framings of corporate activities simultaneously affect and are affected by their audiences (Giorgi & Weber, 2015), and how knowledge recombination is a double-edged sword with opposite impacts on an innovation's degree of novelty and its usefulness (Kaplan & Vakili, 2015). Similarly, topic modeling has been used to generate new conceptual linkages, such as how a particular topic appearing in media statements impacted departures of British parliament members (Hannigan, Porac, Bundy, Wade, & Graffin, 2019), and to refine older constructs such as strategic differentiation (Haans, 2019). Because of its features, topic modeling can serve as a bridge in the social sciences, for it sits at the interfaces between case studies and big data, unstructured and structured analysis, and induction and deduction (DiMaggio, Nag, & Blei, 2013; Grimmer & Stewart, 2013; Mützel, 2015). Not surprisingly, its use in social science, and in management theory more specifically, has increased greatly over the last decade.

As with all new methods, topic modeling techniques continue to be refined. In the current emergent phase of its employment, scholars are still learning the best ways to reveal constructs and develop theory (Evans & Aceves, 2016; Grimmer & Stewart, 2013)—which implies a need for deeper insights into how topic modeling can inform new theories. There are also many technical issues to resolve around topic modeling, such as how to collect and prepare data (Evans & Aceves, 2016), how much supervision should be involved in topic creation (DiMaggio, 2015; Schmiedel et al., 2018), which algorithms are most useful (Bail, 2014), and how new constructs and conceptual linkages can be derived when developing theories from big data (Nelson, 2017; Timmermans & Tavory, 2012). This review addresses these questions with the aim of expanding its use and effectiveness.

We begin by comparing topic modeling's technical and theory-building features with those of close methodological cousins: content analysis, grounded theorizing, and general natural language processing (NLP) of text.[2] Topic modeling's attractive features and ease of use are generating increased interest across the social sciences—raising the disconcerting possibility that the method will become a technical "black box" without an appropriate appreciation of topic modeling's statistical and theoretical underpinnings and implications. In this review, we show that topic modeling is best conceptualized as a "rendering process," which can be understood as a means to juxtapose data and theory (Charmaz, 2014) to generate new theoretical artifacts such as constructs and the links between them (Whetten, 1989). This process involves the rendering of corpora (preparing the sets of texts to be analyzed), the rendering of topics (making analytical choices that determine how topics are identified within those texts), and the rendering of theoretical artifacts (crafting

[2] Topic modeling can be seen both as a specific NLP approach and as something distinct from NLP. Topic modeling relies on interpretation and language-oriented rules but is also unique in its emphasis on the role of human researchers in generating and interpreting specific groups of topics based on the social contexts in which they are embedded. Recent developments have also moved topic modeling further away from NLP, as researchers have applied it to images (Cao & Fei-Fei, 2007) and music (Hu & Saul, 2009) rather than natural language.

topics into constructs, causal links, or mechanisms). By articulating this rendering process, we show that using the machine learning algorithms of topic modeling do not reduce textual analysis to a mechanistic process but actually foreground and inform the analyst's interpretive decisions and theory work.

Our own topic modeling analysis of topic modeling articles created or routinely used by management researchers reveals five theoretical subject areas to which the technique has contributed: detecting novelty and emergence, developing inductive classification systems, understanding online audiences and products, analyzing frames and social movements, and understanding cultural dynamics. For each subject area, we review key concepts and theoretical relationships that have surfaced from the use of topic modeling and identify articles that exemplify its application. We then turn to new trends in topic modeling in the rendering of corpora, topics, and theoretical artifacts. Our review demonstrates that topic modeling not only appeals to diverse management audiences—those interested in topic, content, and category models as well as mixed methods—but also can play a part in cultural structuralism (Lounsbury & Ventresca, 2003), new archivalism (Ventresca & Mohr, 2002), and interpretative data science (Breiger, Wagner-Pacifici, & Mohr, 2018; Mattmann, 2013).

SITUATING TOPIC MODELING AS A TECHNIQUE

Thanks to widespread availability of digitized textual data from a variety of sources and significant increases in computational power, it is now possible for social scientists to study large collections of text (Alvesson & Kärreman, 2000a, 2000b; Langley & Abdallah, 2011; Vaara, 2010). Not surprisingly, a variety of methods for textual analysis—often from neighboring disciplines—have appeared as part of this "linguistic turn." To distinguish the key characteristics of topic modeling and situate it among this wider set of techniques, we first briefly examine three closely related methods: content analysis (Duriau, Reger, & Pfarrer, 2007; Krippendorf, 1980, 2004, 2012; Lasswell, 1948), grounded theorizing with textual data (Gioia, Corley, & Hamilton, 2013; Locke, 2001), and interpretive analysis using the broad class of NLP approaches. These three are particularly useful for elucidating topic modeling's features because they capture the extremes from highly contextualized, careful assessment of smaller batches of selected texts to broader, more algorithmic, and systematic assessment of text from large corpora.

Content Analysis

Social scientists have long been interested in using texts to understand social phenomena [see Krippendorf (1980) for a review]. Content analysis, "a research technique for the objective, systematic, and quantitative description of the manifest content of communication" (Berelson, 1952: 18), represents arguably the most prominent and mainstream approach in this domain (Nelson, 2017; Tirunillai & Tellis, 2014). It relies on the creation of dictionaries or indices comprising mutually exclusive lists of words that can then be applied to texts to isolate meanings and systematically measure specific constructs of interest to the researcher (Krippendorff, 2004). Since its introduction to management theory, scholars have used content analysis in flexible ways, using a range of data sources in areas as varied as the study of management fads (Abrahamson & Fairchild, 1999), industry categories and CEO compensation (Porac, Wade, & Pollock, 1999), corporate reputation (Pfarrer, Pollock, & Rindova, 2010), and technology strategy (Kaplan, 2008a).

From its inception, content analysis scholars have been particularly concerned with the reliability and validity of its various methods (Weber, 1990), advocating the use of protocols and multiple coders to guide text selection and analysis. In recent years, those who use content analysis have increasingly relied on computer-aided text analysis using software and general dictionaries such as General Inquirer and Linguistic Inquiry and Word Count (LIWC) (Pennebaker, Boyd, Jordan, & Blackburn, 2015) to further improve its scalability and systematic nature. At the same time, the mutually exclusive nature of dictionaries precludes "polysemy" (DiMaggio et al., 2013: 578)—an important concept in linguistics where the same word may have a different meaning based on the context in which it appears. A common critique of content analysis has, therefore, been that it yields decontextualized results by reducing complex theoretical constructs into overly general and simple indices (Dey, 1995; Prein & Kelle, 1995).

Grounded Theorizing with Textual Data

To develop theory, scholars often use a highly contextualized approach whereby they gather and engage intensively with texts and then use comparative coding to identify higher order constructs (Charmaz, 2014). By engaging in such grounded theorizing with textual data, a researcher demonstrates a commitment to " 'discovery' through direct contact with the social world studied coupled with a rejection of *a priori* theorizing" (Locke, 2001: 34). Proponents of this approach urge researchers to start with a loosely scoped research question and phenomenon of interest, with the researcher subsequently identifying recurring patterns, ideas, or elements that emerge directly from the data. Doing so often requires culling primary observations and key points and then using axial coding to identify constructs or relationships (Denzin & Lincoln, 2011). Researchers then iteratively group codes into higher order categories to develop general theory. Rather than measurement, grounded theorizing is thus fundamentally concerned with identifying deeper structures embedded in data to attain a rich understanding of social processes.

During the last two decades, grounded theorizing has been used by many groups of management scholars (Charmaz, 2014), including those interested in analyzing language in organizations (Alvesson & Kärreman, 2000a, 2000b), organizational processes and routines (Langley, 1999; Pentland & Feldman, 2005), and culture and identity (Hatch & Schultz, 2017; Nelsen & Barley, 1997). Its theoretical flexibility also makes it the target of some critiques because the role and primacy of meaning, discourse, and understanding typically are not made explicit in research studies (Locke, 2001). Practically speaking, the method also requires great knowledge of context and expertise to apply; it can be not only time- and resource-intensive but also difficult to use with large-scale textual data (Baumer, Mimno, Guha, Quan, & Gay, 2017; Gehman, Glaser, Eisenhardt, Gioia, Langley, & Corley, 2018).

Interpretive Analysis Using NLP

Researchers in linguistics have long used computerization to enable systematized analysis of natural language informed by linguistic rules, with NLP emerging in the 1980s as a way to combine dictionary-based data processing with semantic use to map out likely interpretations of text (Manning & Schütze, 1999). Early versions of NLP relied heavily on grammatical rules from language structure but have given way to more flexible, stochastic approaches to language use (especially as machine learning–based approaches evolved with increased computing power). In management research, scholars often leverage NLP tools to perform semantic parsing on big data and then interpret emerging patterns using computer-aided recognition tools. Kennedy (2005, 2008) was one of the first to analyze media data and sort through evaluations of firms using these tools. Recently, Mollick and others have studied linguistic patterns in crowdfunding and other contexts involving pitches (Kaminski, Jiang, Piller, & Hopp, 2017; Mollick, 2014).

Consistent with its roots in computer science, NLP has been developed to optimize specific tasks or solve particular problems, such as part-of-speech tagging, word segmentation, machine translation, and automatic text summarization. This has resulted in a rich and varied toolkit that is deeply informed by linguistic rules and a firm appreciation for the complexities underpinning human language. At the same time, a single unifying theory does not link the various NLP tools, nor are there standard practices or rules about engaging in NLP-based work. This has created certain challenges for management researchers in applying technical or descriptive tools for theoretically informed purposes. Indeed, scholars have noted that "cooperation between linguistics and the social sciences with regard to text analysis has always been meager" (Pollach, 2012: 264); however, this does not imply that NLP approaches are, by definition, unable to inform management theory.

Topic Modeling

In the early 2000s, topic modeling was developed as a unique NLP-like approach to information retrieval and the classification of large bodies of text (Blei, Ng, & Jordan, 2003). Topic modeling uses statistical associations of words in a text to generate latent topics—clusters of co-occurring words that jointly represent higher order concepts—but without the aid of predefined, explicit dictionaries or interpretive rules. In a pivotal article, Blei et al. (2003) introduced a Bayesian probabilistic model using latent Dirichlet allocation (LDA) to uncover latent structures in texts. LDA is a "statistical model of language" (DiMaggio et al., 2013: 577) and is the simplest of several possible generative models available for topic modeling (Blei, 2012). It focuses on words that co-occur in documents, viewing documents as random mixtures of latent topics, where each topic is itself a distribution among words (Blei et al., 2003). Importantly, an assumption of topic modeling is that documents are "bags of words" without syntax, which defines meaning as relational (Saussure, 1959) and emerging from co-occurrence patterns independent of syntax, narrative, or location within the documents (Mohr, Wagner-Pacifici, Breiger, & Bogdanov, 2013).

Generating topics using statistical probabilities has three key benefits. First, researchers do not have to impose dictionaries and interpretive rules on the data. Second, the method enables the identification of important themes that human readers are unable to discern. Third, it allows for polysemy because topics are not mutually exclusive; individual words appear across topics with differing probabilities, and topics themselves may overlap or cluster (DiMaggio et al., 2013: 578).

Comparison of Text Analysis Techniques in Management Research

Figure 1 compares the use of topic modeling in social science and management research to the use of grounded theory, content analysis, and general NLP approaches in articles listed in the Web of Science and Scopus published between 2003 (the year Blei and colleagues' foundational article was published) and 2017. We included articles for topic modeling if

FIGURE 1
A Comparative Assessment of Topic Modeling's Use

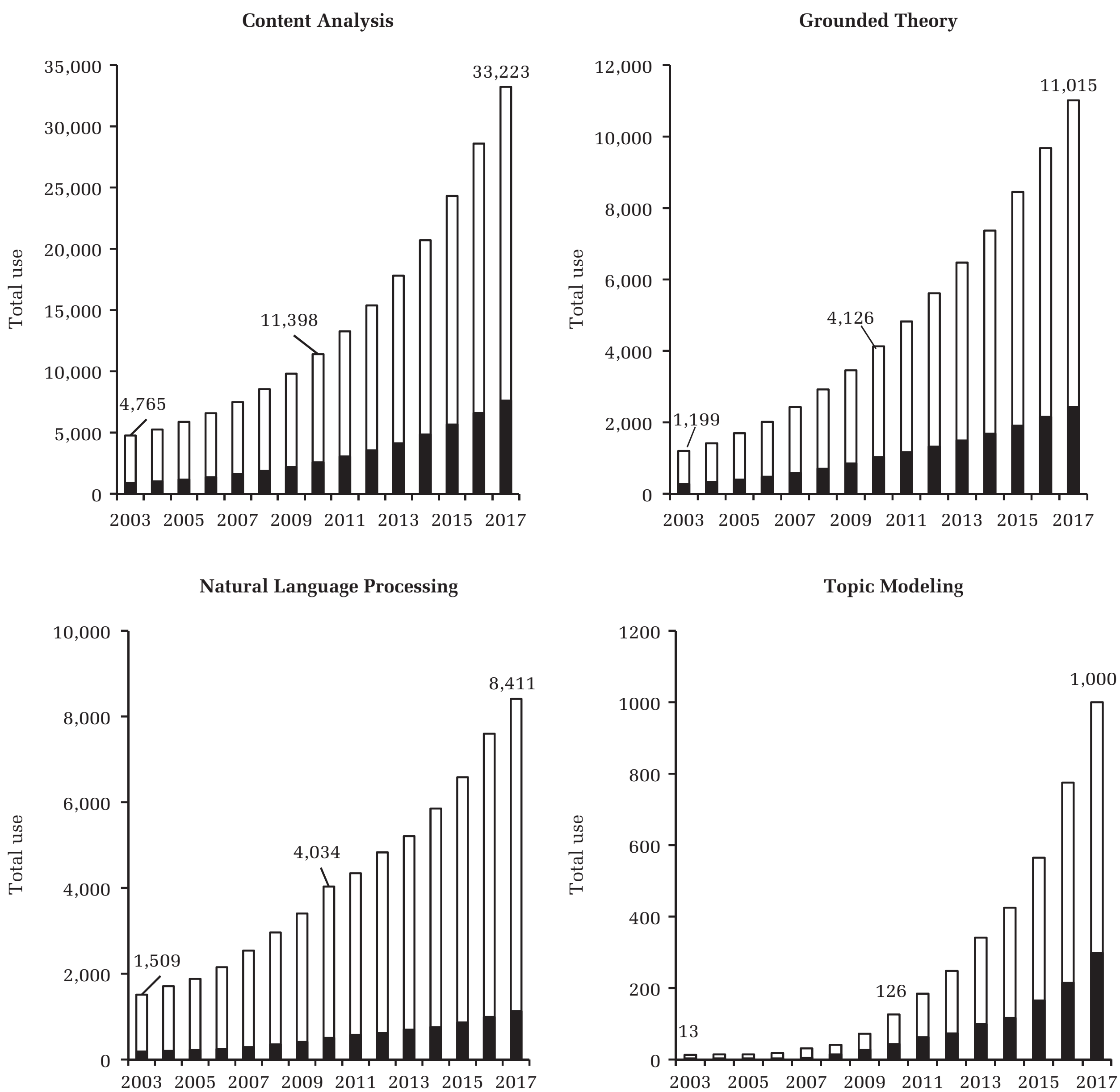
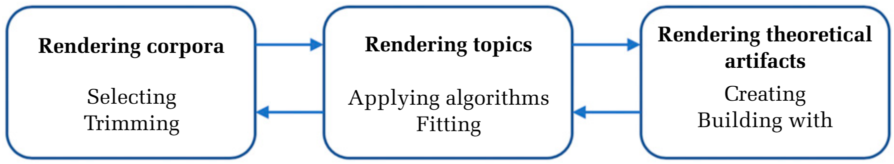


FIGURE 2
Topic Modeling Rendering in Theory-Building Spaces



"topic mod*" appears in their titles, abstracts, keywords, or automated indexed keywords. We included articles for grounded theorization, content analysis, and NLP if they contain "ground theor*", "content analys*", and "natural language process*", respectively.[3] The bar charts in each panel represent the cumulative number of articles in each year, with black bars showing the number of articles in business and economics specifically and white bars showing articles in the social sciences more generally.

As a group, the four panels highlight the linguistic turn in social science, with increased use of all of these approaches reflecting the increasing appetite in the field to study the structure and meaning underpinning collections of text. By 2017, 1,000 topic modeling articles had been published, with around 300 in the management domain specifically. Although this is just a fraction of the literature relative to studies based on more established approaches, Figure 1 does suggest that the use of topic modeling has been particularly high in the management domain. Indeed, 29.8 percent of all articles based on topic modeling published between 2003 and 2017 fall within the management domain, compared with 13.4 percent, 22.0 percent, and 22.9 percent for NLP, grounded theorization, and content analysis, respectively. Figure 1 also reveals that topic modeling has been adopted at an exceptionally rapid rate in recent years, with a compound annual growth rate of 34.4 percent since 2010, versus 11.1 percent for NLP, 15.1 percent for grounded theory, and 16.5 percent for content analysis. We suggest that topic modeling's appeal primarily lies in its unique position at the intersection of the other three approaches, a point that we elaborate in the conclusion.

[3] Although these may undercount articles that do not mention the methodologies and overcount articles without textual data, we suspect that these issues are equally salient for each approach. For illustration, adding "Linguistic Inquiry and Word Count" and "LIWC" adds just 271 articles to the set of more than 20,000 for content analysis.

RENDERING THEORY FROM DATA IN TOPIC MODELING

Given its increasing importance in the social sciences and its unique location between human-based and machine-learned analysis of discourse, a more careful consideration of the nature of topic modeling and the topic modeling process is useful for management researchers. To date, much of the work on topic modeling has focused on issues of algorithm selection (Blei et al., 2003; Schmiedel et al., 2018) and its application to curated texts. We think it is important to discuss the use of topic modeling from the preprocessing to theorization stages to illustrate its possibilities for theory building.

We use the term "rendering" to describe the iterative creation of theory from corpora through topic modeling. In the social sciences, Charmaz (2014: 216, 369) used the term rendering to describe the process of "juxtaposing data and concept" and "categorizing data" for interpretation, whereas computer scientists use rendering to create photorealistic or nonphotorealistic images in two or three dimensions via automated analysis and specific algorithms (Strothotte & Schlechtweg, 2002). Drawing on these descriptions for inspiration, we define rendering in topic modeling as a three-part process of generating provisional knowledge by iterating between selecting and trimming raw textual data, applying algorithms and fitting criteria to surface topics, and creating and building with theoretical artifacts, such as processes, causal links, or measures. These three steps are displayed in Figure 2. To provide readers with background information, we present definitions of common terms used in topic modeling in Table 1.

Rendering Corpora

In the first process—rendering corpora—an analyst, guided by theoretical and empirical considerations, selects types of textual data. As with any form of empirical analysis, selection of the sample (in our

TABLE 1
Topic Modeling Conceptual Terms

Conceptual Terms	Definition in the Context of Rendering with Topic Modeling
Algorithm	A process or unambiguous set of rules to be followed, usually by a computer. An automated processing technique for distilling data inputs into topic modeling elements (clusters, weights, and similarities).
Big textual data	Data characterized by large volume (a million or more words), high variety (diverse sources), and high temporality (many periods).
Coherence	A quantitative metric for topic quality. Clear and well-bounded topic(s) with evident criteria for classification of other text or topics within it. Based on pairs of words in a topic that have high co-document frequencies.
Dictionary	The set of meaningful key words to be used to assess the content and meaning of a corpus. The basis for annotating words in a text as a code category.
Disambiguation	A process of using the context to adjudicate between different meanings (or readings) of a word beyond its literal definition.
Fit	Criteria for how many topics are derived, how they are related, and what they might mean.
Heteroglossia	Multiple styles of word-use in a single text reflecting different perspectives or styles of expression.
hLDA	Hierarchical latent Dirichlet allocation—a form of structured LDA.
KWIC	Key words in context; embedding or considering words in their relationship with other words in a corpus and in a sociocultural condition.
Lemmatizing	Transforming a word into its dictionary form. In practice, different lemmatizing methods convert words to their singular forms or by using a higher level synonym from a linguistic thesaurus.
LDA	Latent Dirichlet allocation, in which documents are assumed to draw content from a latent set of topics with probability-based parameters that can be adjusted to determine those topics.
LIWC	Linguistic inquiry and word count (aka "Luke") is a dictionary-based, positive- and negative-affect word frequency program designed to capture content and affective meaning.
LSI	Latent semantic indexing is an algorithm which uses linear algebra to perform dimensionality reduction and convert texts to a matrix form.
LSVDs	Lasswell value dictionary tags.
Perplexity	A quantitative metric for the quality of a topic model based on the number of topics selected. In general, perplexity is a statistical measure of how well a model fits based on splitting data into a training set and test set. In LDA topic modeling, it is a relative measure of topic fit; better models have lower perplexity scores.
Polysemy	Words that have multiple meanings or uses.
Relationality	Words whose meanings are contextually dependent.
Rendering	The process of generating provisional knowledge by iterating between selecting and trimming raw textual data, applying algorithms and fit criteria to surface topics, and creating and building with theoretical artifacts, such as processes, causal links or measures.
Selecting	Selecting documents (e.g., using sampling) and forms of text to be assessed.
Smoothing	Applying LDA-related algorithms to reduce the number of and disparity among topics, normally through iteration.
Stemming	The conversion of text segments (words) to their root word forms.
Stop words	Words that serve a less important role in meaning construction (i.e., articles such as "the" or "a").
Theoretical artifact	A construct, conceptual association, process, causal linkage, mechanism, or measure.
Token	The smallest, disaggregated, distinct bit of textual data (normally a noun) used in analysis.
Topic	A bag of words that frequently appear together across documents; the derived word(s) from a topic in topic modeling representing word tokens.
Trimming	Reducing textual data and specific words into useful tokens, normally by lemmatizing and/or stemming; a form of text normalization.

context, texts) is a crucial step that fundamentally shapes all subsequent steps. For textual data in particular, selection needs to account for language, authoring, and document sources—ensuring a logical fit with the research question being investigated while simultaneously considering common issues such as representativeness, levels of analysis, and temporal considerations (e.g., longitudinal vs. cross-sectional data). The analyst then compiles such data for further preprocessing and cleaning. If the data are from one primary source, the compiled text is considered a corpus; if from different sources, corpora.

On the whole, topic modeling tends to be applied more frequently to sampled corpora than to a single, homogenous corpus (Borgman, 2015; Kitchin & McArdle, 2016). As a result, topic modeling relies on a great deal of preprocessing with various techniques and rules of practice to prepare texts for analysis (Nelson, 2017; Schmiedel et al., 2018). During preprocessing, the texts are sorted, disassembled, and then trimmed according to broader content analysis principles such as ignoring "stop words" (e.g., "the" and "a") and focusing on nouns rather than verbs, adjectives, or adverbs. Topic modelers also often standardize word forms, using stemming and lemmatizing (see Table 1) to transform words into their roots (Kobayashi, Mol, Berkers, Kismihók, & Den Hartog, 2018). Recently, more refined techniques such as WordNet have been developed to convert words to their singular forms or to use higher level synonyms (Miller, Beckwith, Fellbaum, Gross, & Miller, 1990). These considerations are all crucial, as most topic modeling algorithms analyze words based on how they appear, letter-by-letter (e.g., "firm" is not the same as "firms"). As such, these cleaning steps represent a form of systematic, normatively guided trimming to standardize words to allow the capture of constellations of words that represent deeper sociocultural structures (Mohr, 1998).

Rendering Topics

During the second process—rendering topics—the analyst applies an algorithm to identify appropriate topics. An algorithm provides an analyst with the ability to use a pre-programmed set of rules to automatically reduce the dimensions of the corpora (Mohr, 1998). The most well-known algorithm, as discussed previously, is LDA. According to Blei et al. (2003: 994), the key assumption in LDA is that "each word in a document [is modeled] as a sample from a mixture model, where the mixture components are multinomial random variables that can be viewed as representations of 'topics.' " The major theoretical and methodological insight here is that documents are assumed to draw content from a latent set of topics with probability-based parameters that can be adjusted to determine those topics. This implies that words are generated from a topic yet can also be used in different topics with different probabilities. Because documents belong to the same corpus, the algorithm assumes that they were generated from the same process, and thus, each document constitutes a mixture of the same set of "topics" in different proportions. Topics are a weighted vector of words and each topic corresponds to a distinct concept (Grimmer & Stewart, 2013). However, unlike the dictionaries used in content analysis, which are comprised of mutually exclusive lists of words (Krippendorff, 2004: 132), in topic modeling, the same words can appear in different topics (DiMaggio et al., 2013: 578), although likely in very different proportions and juxtaposed with different words.

The inputs to the LDA algorithm include (a) a set of documents that can be represented as a document–word matrix—with rows representing each document in the corpora, columns representing each unique word in the corpus, and cells indicating the number of times each word occurs in each document—and (b) the number of topics to be estimated by the algorithm. Importantly, most topic modeling algorithms (such as LDA) require probability draws for each document, such that each document is considered "a bag of words" with no syntax. The outputs from LDA include a topic–word matrix (vectors of the weights of words in each topic) and a topic–document matrix (vectors of the weights of topics in each document). In subsequent analyses, math (i.e., vector space calculations) can be applied to these outputs to classify texts into categories, analyze themes, or compare corpora based on similarities.

Each successfully computed model is based on different parameters (e.g., number of topics) and generates a distribution of topics over documents and/or words, which can be used by the researcher to identify the eventual model that will be used in the study. The notion of fit is typically invoked to decide how many topics are derived, how they are related, and what they might mean. A researcher can focus on one of two notions of fit—rooted in a logic of either accuracy or validity—and this focus has important implications for which topic model is judged to provide the most appropriate fit given the research question.

One version of fit is based on a logic of accuracy, a central focus of computer scientists who rely on metrics such as perplexity, log-likelihood and coherence (defined in Table 1) to determine the number of topics and their salience (Azzopardi, Girolami, & van Risjbergen, 2003; Chang, Boyd-Graber, Gerrish, Wang, & Blei, 2009; Mimno, Wallach, Talley, Leenders, & McCallum, 2011). However, Chang et al. (2009) pointed to disparities between some quantitative metrics and how people interpret topics: topic models that perform better on quantitative metrics tend to infer topics that humans judge to be semantically less meaningful. Indeed, DiMaggio et al. (2013: 582) suggested that "there is no statistical test for the optimal number of topics or for the quality of a solution" and that "the point is not to estimate population parameters correctly, but to identify the lens through which one can see the data most clearly."

Therefore, social scientists tend to focus more on the logic of fit as validity (DiMaggio, 2015). DiMaggio et al. (2013) identified two key forms of validity: semantic or internal validity, and predictive or external validity. To demonstrate internal validity, the researcher must confirm that the model meaningfully discriminates between different senses of the same or similar terms. To demonstrate external validity, the researcher must determine whether particular topics correspond to information external to the topic model (e.g., by confirming that certain topics became more salient when an external event relevant to those topics occurred) (DiMaggio et al., 2013). For example, Kaplan and Vakili (2015) identified models with 50, 75, and 100 topics for a corpora of nanotechnology patent abstracts and then used three expert evaluators to determine that the 100-topic model was the most semantically meaningful. Jointly, these two forms of validity are concerned with confirming that the topic model's outputs are semantically meaningful—a process that entails substantial interpretive uncertainty (DiMaggio, 2015). Because of the uncertainty involved in the rendering of topics, most scholars in the social sciences attempt to locate the optimal balance between the two logics of accuracy and validity to identify the "best" topic model to be used in further theorizing.

In sum, topic modeling has advanced how we think about and interpret topics in textual data by enabling researchers to uncover latent topics rather than imposing pre-established categories on the data. It is superior to word-count techniques because it identifies ideas or concepts based on constellations of words used across documents in a corpus. It is thus sensitive to semiotic principles of polysemy (words with multiple meanings or uses), heteroglossia [uses predicated on audiences and authors, as described by Bakhtin (1982)], and the relationality of meaning (which is contextually dependent) (DiMaggio et al., 2013). As a result, topic model outputs, after some interpretation and theoretical defense, are useful in generating theoretical artifacts, especially in large and otherwise unmanageable data sets.

Rendering Theoretical Artifacts

In the third process—rendering theoretical artifacts—researchers iterate between theory and the topics that emerge from the chosen model to create new theoretical artifacts or to build theory with them (Whetten, 1989). The word- and topic-vectors offer a wide range of opportunities for the researcher to build artifacts. The artifacts may be multidimensional constructs, such as novelty (Kaplan & Vakili, 2015) or differentiation (Haans, 2019), captured by a set of topics clustered or scaled around words or concepts. The artifacts may also be relational (correlational, causal, or process-based), thereby allowing researchers to uncover mechanisms.

For instance, Croidieu and Kim (2018: 11) used an "iterative, multistep process" to interpret the outputs of the topic model to discover concepts related to lay expertise legitimation and the mechanisms underpinning it. They described their process for creating theoretical artifacts from their algorithmic output in detail.

First, we started with the raw topics as descriptive codes. Second, we labeled these topics as first-order concepts. We coded all labels separately and together as an author team, extensively discussed the results, and recoded the topics when necessary. Third, we grouped these topics into more abstract and general second-order themes. Fourth, we analyzed the distribution of these second-order themes per year and iteratively developed four aggregate dimensions, which we present in the following sections as the mechanisms for expertise legitimation. Fifth, we refined the labeling and theorizing of these aggregate dimensions by dividing our analysis into two periods. . .We chose these periods both for their historical significance and because they are anchored by a central empirical puzzle related to our theoretical framework. . .Last, we repeated this procedure multiple times to ensure tight correspondence between our raw-topic data and our coding interpretations. From this iterative coding work, we produced our findings and constructed our process model (Croidieu & Kim, 2018: 11).

The inherent flexibility of the rendering process has enabled topic modeling researchers to develop better measures and clever extensions of existing theoretical constructs and relationships and to induce novel concepts, processes, and mechanisms. As such, topic modeling can be used for either deductive or inductive theorizing. Indeed, during the rendering process, different choices arise (e.g., around selection, fit, and the form of artifact) based on whether one uses more deductive versus inductive theorizing. The many paths defined by these choices provide further evidence of topic modeling's flexibility and potential. Not surprisingly, topic modeling is contributing to a wide array of management theory subjects, some arising from more mature theory and some from emerging areas.

BUILDING MANAGEMENT KNOWLEDGE THROUGH TOPIC MODELING

During the 15 years since topic modeling was first used in management research, its use through rendering has enabled management scholars to explore subjects in new ways, thereby building management knowledge. To systematically identify the subjects enhanced by such rendering, we applied the topic modeling rendering process depicted in Figure 2 to topic modeling articles in the literature [for similar meta-theorizing moves, see Mohr and Bogdanov (2013) or Wang, Bendle, Mai, and Cotte (2015)]. Although our rendering process was iterative and recursive, we present our methodological approach as a series of sequential steps, as outlined in Figure 1 (e.g., rendering our corpus, topics, and theoretical artifacts).

We began our analysis by curating a corpus consisting of all relevant topic modeling articles from the Web of Science and Scopus. We winnowed those articles down by focusing on management journals (e.g., *Administrative Science Quarterly* [*ASQ*] *and Strategic Management Journal* [*SMJ*]) and other journals that management scholars read. We identified these journals based on both our first-hand experience and citations of articles that have influenced management scholars. Following the procedure used by Mohr and Bogdanov (2013), we divided the articles into paragraphs to form 5,362 documents and used the Stanford CoreNLP software (Manning, Surdeanu, Bauer, Finkel, Bethard, & McClosky, 2014) to lemmatize the words, yielding 351,786 distinct words for analysis. During our analysis, we sharpened our criteria for including and excluding particular articles in our analysis as we interpreted the output of topic modeling algorithms. Our final corpus contained 66 articles (for details, consult Table A1 in the Appendix). We organized these procedures using the Jupyter Notebook software in Python, which enabled us to track and visually annotate our process.

We continued our analysis by applying a collapsed Gibbs sampler with the LDA algorithm to our corpus to render topics. Collapsed Gibbs sampling (Griffiths & Steyvers, 2004) is an approach from the Markov Chain Monte Carlo framework that iteratively steps through configurations to estimate optimal model fit. When combined with the LDA algorithm (Blei et al., 2003), topics can be estimated with minimal configuration by the user. As is a common practice (Jha & Beckman, 2017; Mohr & Bogdanov, 2013), we used the MALLET software tool (McCallum, 2002) to conduct this procedure. We approached the critical task of determining the optimal number of topics by computing a variety of topic models. For each model, we graphed the average coherence score across topics (Mimno et al., 2011), which revealed a plateau value; we used this evidence as guidance and observed several models (i.e., those with 30, 35, 40, 45, and 50 topics) more closely from an interpretive perspective. Fligstein, Stuart Brundage, and Schultz (2017) followed a similar procedure, moving from collapsed Gibbs sampling through various models, using coherence and interpretability to narrow in on stable sets of topics. Finally, following Mohr and Bogdanov (2013), we applied our 35-topic model (derived from separate paragraphs) to each document to generate a distribution of topic weights (i.e., the topic–document matrix where each row is a document and each column is a topic weight, with all weights adding up to 1). We then sorted topics for salience based on average topic weights and word relevance to identify 35 ordered topics.

Three coauthors then independently used the algorithmic output of the topic models to render theoretical artifacts. Specifically, we each created a summary document for each topic that contained three visualizations generated by the topic modeling algorithm: a weighted word list, a weighted document list, and a multidimensional scaling visualization (Sievert & Shirley, 2014) that showed each topic in relation to other topics (see Figure A2, for an example of this theoretical artifact). The three authors then independently analyzed these documents to generate first- and second-order codes (Bansal & Corley, 2014; Denzin & Lincoln, 2011; Gioia et al., 2013; Pratt, 2009; Strauss & Corbin, 1998). Through a series of independent coding exercises and interactive conversations, the authors then aggregated these first- and second-order codes into broader management subject areas (Gioia et al., 2013). In other words, in keeping with rendering practice, we tried not to impose too much meaning on the set of topics; instead, we let the insights and themes for management theorizing emerge from them.

Our bottom-up, inductive analysis suggests that topic modeling has enhanced our management theory knowledge in five subject areas: detecting novelty and emergence, developing inductive classification systems, understanding online audiences and markets, analyzing frames and social movements, and understanding cultural dynamics.[4] This specific ordering of subjects is not determined by topic weights; moreover, the timing of their identification in the model's convergence does not reflect a strict ordering. In fact, our preliminary analyses of

[4] In addition, some topics corresponded specifically to the method of performing topic modeling, and given our interest in the rendering of management theory, we purposefully backgrounded these topics (see Table A2 for details).

the wider corpora in the field and understanding of the field's evolution reveal how analyses of novelty, classification, and online audiences developed in parallel with analyses of framing and cultural dynamics. In the sections that follow, we focus on how theoretical knowledge in each subject area has been extended by rendering with topic modeling. Subject areas, topic-based themes, exemplary articles, and theoretical contributions are summarized in Table 2.

Detecting Novelty and Emergence

Management researchers are interested in topics of novelty and emergence because they apply to a variety of research streams, such as categories (Durand & Khaire, 2017; Hannan, Pólos, & Carroll, 2007; Kennedy & Fiss, 2013), cultural entrepreneurship (Lounsbury & Glynn, 2001, 2019), innovation (Fleming, 2001; Sørensen & Stuart, 2000), organizational forms (Rao, Monin, & Durand, 2003), and changes in managerial cognition and attention (Ocasio, 1997). Novelty is a key concern within innovation studies (Kline & Rosenberg, 1986; Trajtenberg, 1990), but measures typically are indirect. For instance, as noted by Kaplan and Vakili (2015), many studies identify emergence based on the successful introduction of new innovations, thus raising concerns of endogeneity and lack of causal identification.

Topic modeling offers a solution to fundamental challenges faced in these broad research streams. Specifically, topic modeling can be applied to documents to generate theoretical insights because (a) the language used in documents represents their cognitive content (Whorf, 1956); and (b) actors use vocabularies to describe similar ideas (Loewenstein, Ocasio, & Jones, 2012). Thus, topic modeling can be used to discern the cognitive content of documents that describe cases of novelty and emergence (i.e., innovation contexts) and assess the extent to which such content is similar or different across documents. Topics rendered in our analysis include explaining shifts in patent citations (#25), understanding innovation (#24), managerial cognition (#1), understanding knowledge dynamics (#14), and emerging organizational forms (#10).

The first topic in this subject area relates to the use of topic modeling to measure the novelty of ideas in patents—an arena in which novelty has been heavily studied under the rubric of recombination and innovation (Fleming, 2001). For instance, Kaplan and Vakili (2015) applied topic modeling techniques to create representations of ideas in documents that can be compared using mathematical distance to determine cognitive novelty. This measure of novelty based on the actual cognitive content of documents provides several advantages over more traditional measures of novelty based on citations in subsequent patents or publications (Trajtenberg, 1990). In the popular citation-based approach, a patent is flagged as a breakthrough if it has a substantial impact on subsequent technologies. However, citation-based measures of technological novelty often confound novelty and impact (Momeni & Rost, 2016); consequently, novel ideas may not be recognized as important precursors because of the processes by which citations are produced (false negatives), and incremental ideas may be incorrectly identified as novel when they generate substantial impact for reasons other than novelty (false positives).

In contrast to simple counts of citations or patent classes, a measure based on the cognitive content of a document enables researchers to gauge the novelty of the idea(s) presented, independent of their ex-post economic value. Kaplan and Vakili (2015) used topic modeling to distinguish cognitive novelty from economic value. In their analysis of nanotube patents, they reported a very small correlation between topics identified by LDA and patent classes assigned by the U.S. Patent and Trademark Office. Often, truly novel ideas are assigned to classes that may not reflect their actual cognitive content. Their study has implications for teasing out longstanding debates in management around contrasting theories of creative processes surrounding the sources of innovative breakthroughs. In a related study, Ruckman and McCarthy (2017) used topic modeling to analyze patents in an attempt to explain why some patents are licensed over others. Their goal was to address conflicting findings in prior research: some scholars have advocated a "status model" (Podolny, 1993), whereas others have supported organizational learning explanations based on optimizing knowledge transfer in licensing contracts (Arora, 1995). Ruckman and McCarthy used topic modeling to directly measure cognitive content, enabling them to construct a set of "alternate patents" that could have been licensed based on content, but were not. Thus, by controlling for cognitive content, they were able to isolate other variables such as the licensor's technological prestige and experience at licensing, and characteristics of the patent itself such as combined technological breadth and depth. Using better controls when comparing similar patents enabled them to produce a contingent model of patent licensing likelihood based on licensor attributions and the

TABLE 2
Management Subject Areas Enhanced by Topic Modeling Research

Subject Area	Topics	Exemplars	Key Contributions
Detecting novelty and emergence	Understanding shifts in patent citations (#25: patent, technology, knowledge, technological, citation, identify, path, base, cite, and highly)	Kaplan and Vakili (2015)	Provides a means to disentangle the cognitive content of novel innovations from the outcomes associated with innovations
	Measuring topics to understand innovation (#24: idea, weight, distribution, edge, measure, base, node, combination, average, and semantic)	Toubia and Netzer (2016)	Provides a means of empirically measuring different theoretical dimensions of creativity to develop new understandings of idea generation
	Using topic models to understand managerial cognition through technology problems, search, and attention (#1: problem, search, structure, attention, concept, process, exist, unit, create, and general)	Wilson and Joseph (2015)	Provides a way for researchers to understand the dynamics of managerial attention relative to background knowledge
	Understanding knowledge dynamics (#14: scientific, impact, focus, app, knowledge, article, content, find, rhetorical, and attribute)	Antons et al. (2018)	Provides a means to theorize how latent knowledge structures undergird innovative activities
	Understanding emerging organizational forms (#10: form, identity, community, logic, organizational, actor, institutional, application, distinct, and school)	Jha and Beckman (2017)	Provides a method for theorizing the relationships between constructs at different levels of analysis, such as organizational identity and institutional logics
Developing inductive classification systems	Understanding dynamics of meanings and networks in knowledge fields (#34: article, journal, field, publish, year, citation, scholar, papers, author, and paper)	Wang et al. (2015)	Provides a means to discover emerging trends in knowledge fields by enabling researchers to identify different dimensions of knowledge and connect these dimensions with other theoretical constructs
	Understanding how categories affect competitive dynamics (#18: firm, category, industry, performance, position, distinctiveness, competitor, show, level, and competitive)	Haans (2019)	Provides a means to measure differentiation associated with cultural concepts in strategic action
	Understanding the relationships between risk and investment (#31: information, analyst, report, investor, risk, discovery, interpretation, manager, role, and find)	Huang et al. (2017)	Provides a way for researchers to compare disparate forms of data such as written reports and transcripts of conference calls
	Inducing underlying meanings associated with cultural events (#32: major, rebellion, job, event, state, report, case, crime, level, and related)	Miller (2013)	Provides a way to overcome human biases associated with interpreting cultural events
	Classifying sets of data and consumers (#4: make, pile, task, datum, set, summary, consumer, sort, propose, and item)	Blanchard, Aloise, and Desarbo (2017)	Introduces a new technique that can be used to address a classic consumer behavior problem of sorting
Understanding online audiences and products	The nature of online consumer profiles (#12: user, content, message, social-media, consumer, influence, individual, role, activity, and platform)	Trusov et al. (2016)	Provides a means for conceptualizing customers as click groups, networks, and online communities
	Online brand recognition and preference (#23: brand, approach, car, text-mining, map, keyword, association, mention, tag, and consumer)	Netzer et al. (2012)	Helps capture brand network attributes and evolving brand linkages
	Online customer evaluations and responses to them (#29: review, response, rating, health, restaurant, post, hotel, regulation, find, and treatment)	Wang and Chaudry (2018)	Maps the co-occurrence of reviews and responses in real time to understand performance adjustment effects
	Improving topic modeling of online audiences and products (#13: product, dimension, customer, consumer, attribute, purchase, market, prediction, review, and online)	Jacobs et al. (2016)	Refines topic selection and supervision criteria, as well as fit criteria (e.g., smoothing, correlation, and hierarchy across topics)

TABLE 2
(Continued)

Subject Area	Topics	Exemplars	Key Contributions
Analyzing frames and social movements	Understanding how frames influence political processes (#27: financial, fomc, economy, price, market, hypothesis, macroeconomic, primary, discussion, and real)	Fligstein et al. (2017)	Provides a means to identify and measure the deployment of different frames in political activities
	The relationship between frames, context, and audience (#6: frame, context, audience, important, framing, make, process, give, individual, and part)	Levy and Franklin (2014)	Enables researchers to identify distinct discursive frames
	Understanding field-level relationships between organizations, discourse, and strategies (#17: organization, theme, individual, effort, people, comment, strategy, day, term, and field)	Bail et al. (2017)	Provides a means to capture sentiment and bias in normalized spaces
	Social movement strategies, networks, and actions (#11: group, network, identify, radical, movement, pair, environmental, action, strategy, and finding)	Almquist and Bagozzi (2017)	Provides a means to map unseen or hidden ties
Understanding cultural dynamics	Understanding the professionalization of a field (#2: amateur, field, professional, public, space, radio, actor, theme, expertise, and expert)	Croidieu and Kim (2018)	Provides a method for inductively analyzing a corpus as part of a longitudinal case study
	Using topic modeling to analyze big data to understand cultural trends (#5: social, conversation, big-data, language, theory, cognitive, public, shift, meaning, and emotional)	Wagner-Pacifici et al. (2015)	Articles that explicitly describe and illustrate how to use topic modeling to extract meanings from large corpora
	Understanding dynamics associated with literary meanings (#9: work, author, write, literary, passage, read, corpus, series, gender, and stm)	Tangherlini and Leonard (2013)	Enables researchers to identify and compare meanings across different subcorpora over time
	Understanding how cultural meanings change over time (#19: art, support, term, percent, view, recombination, newspaper, assign, agency, and grant)	DiMaggio et al. (2013)	Enables researchers to analyze shifts in cultural meanings over time
	Understanding the evolution of cultural trends (#28: time, period, trend, change, fertility, population, country, context, British, and demographic)	Marshall (2013)	Uses methods such as correlated topic modeling to connect changes in cultural meaning over time with quantitative data

combination of technological breadth and depth as an attractive signal. Topic modeling has thus enabled researchers who study patents and innovation to not only increase the precision of their analyses but also develop new theory about the role of knowledge dynamics on economic outcomes.

A second topic in this subject area that is closely related to explaining shifts in patent citations is the use of texts more generally as a means to measure innovation and creativity. Toubia and Netzer (2016) proposed that creative and novel ideas should have some type of structural signature that can be found in cognitive representations. Drawing on literature related to cognitive creative processes in science (i.e., Rothenberg, 2014; Uzzi, Mukherjee, Stringer, & Jones, 2013), they explored this proposition as an optimal balance of familiarity and novelty. Toubia and Netzer (2016) primarily adopted a semantic network analysis approach to explore the structural argument of familiarity, showing how co-occurrences of word stems can constitute a common substructure, what they called a "structural prototype." In turn, they argued that creativity is a function of a semantic network structure with a core substructure corresponding to a familiar prototype, and novelty dimensions reflected as sufficient semantic distance in the overall structure. They demonstrated this argument empirically across eight studies and 4,000 different ideas in multiple domains that were coded by expert judges. They used LDA as a robustness check to show that creativity was not simply a function of semantic distance. Interestingly, both Toubia and Netzer (2016) and Kaplan and Vakili (2015) featured in this topic: in different domains, the authors leveraged topic modeling techniques to theorize how to identify innovation in documents through the direct measurement of cognitive representations.

The third and fourth topics—using topic models to understand managerial cognition and knowledge dynamics—relate to actors detecting novelty within a body of knowledge. The core idea of using topic modeling to study knowledge dynamics is based on two related insights: first, the language used in documents represents their cognitive content (Whorf, 1956), and second, actors use similar vocabularies to describe similar ideas (Loewenstein et al., 2012). In our analysis, the third topic reveals that topic models can be used to understand changing cognition over time through varying managerial attention (Ocasio, 1997). When a corpus covers the body of knowledge in a specific domain (e.g., scientific articles or patents in the technology field), topic modeling can reveal an accurate depiction of the idea space in that body of knowledge. However, topic modeling can also reveal how actors, as producers of documents, can attend to ideas in the latent idea space. As Kaplan and Vakili (2015) demonstrated, to the extent that describing a truly novel (or disruptive) idea requires using a new vocabulary, one can identify the level of cognitive novelty in a document by measuring how much it conforms to or deviates from previously established topics and their constitutive vocabularies in the corresponding body of knowledge. Wilson and Joseph (2015: 417) used topic modeling to render the "patent background" as a "representation of a technical problem" at a particular point in time. Because managerial attention is scarce, it is allocated across a small set of technological problems, particularly at the level of a business unit (Argote & Greve, 2007). Thus, the rise and fall of topics as technological problems reflect not only managerial attention within a firm but also novelty within the broader field or patent class.

Topic modeling has also been used to study knowledge dynamics in science by tracking the novelty of ideas in journals over time. Conceptualizing scientific communities as "thought collectives with distinct thought styles," Antons, Joshi, and Salge (2018: 1) used topic modeling to break down articles in terms of topical and rhetorical attributes. They demonstrated that topical newness is not only associated with an article "citation premium" in a scientific community, but also significantly increases with a rhetorical stance of tentativeness rather than certainty. Similarly, Wang et al. (2015) used topic modeling to discover emerging trends in knowledge fields, noting that citation analyses and LDA together can be used to narrate a story about novelty and progress against a broader backdrop of social structure, including niche topical areas and author status dynamics. Both articles in this topic contextualize traditional citation-based measures of article impact against cognitive dynamics in topic analyses.

A final topic revealed by our analysis of this subject area reflects the use of topic modeling to understand emerging organizational forms. This approach provides a method to trace how meanings of organizational forms emerge longitudinally. Jha and Beckman (2017) used topic modeling to show how field-level logics moderated actors' attempts to carve out organizational identities around charter schools. Topic modeling enabled the authors to connect two traditionally distinct theoretical concepts—institutional logics and organizational identities—and explain the relationships between them. Given how meaning has typically been studied in organizational theory using concepts such as identity, institutional logics, and frames, studying the emergence of meanings in spaces such as organizational fields and categories may become an increasingly relevant application of topic modeling methods.

Topic modeling has increased precision and enabled deeper insights in studies of novelty and knowledge dynamics, thereby facilitating the generation of new theory in a variety of innovation-related contexts. Topic modeling provides considerable advantages over traditional methods such as counts of patent filings or subsequent citations, which rely on existing classification methods that were not designed to capture novel and emergent ideas. By directly leveraging the cognitive content of texts (such as patents or articles), topic modeling augments traditional measures of impact in knowledge fields. Furthermore, by separating measures of impact from those of knowledge itself, topic modeling has advanced theory by empowering researchers to invent more precise means to empirically test competing theoretical mechanisms. In the bigger picture, these uses of topic modeling may help scholars address longstanding questions in the management literature by conceptualizing the role of novelty with institutional logics (Thornton, Ocasio, & Lounsbury, 2012), or delineating the roles of innovation and boundaries with paradigms (Kuhn, 1996).

Developing Inductive Classification Systems

Management researchers routinely use topic modeling to develop inductive classification systems. Such systems are particularly important in a variety of theoretical research streams, including studies of competitive dynamics and optimal distinctiveness (Deephouse, 1999; Zhao, Fisher, Lounsbury, & Miller, 2017), and the evaluation of risk factors in corporate disclosures to investors (Fama & French, 1993). More generally, these research streams are exploring classification as shared structures of meaning that are not formally materialized. For example, studying institutional logics (Thornton et al., 2012) or implicit understandings of early industry structure (Forbes & Kirsch, 2011) requires researchers to develop inductive understandings of shared meanings that have categorical imperatives. Researchers in each of these traditions who seek to identify categories of meaning in text face challenges of analyzing large quantities of data without introducing researcher bias. Our analysis reveals six topics in this subject area: understanding dynamics of meanings and networks in knowledge fields (#34), understanding how categories affect competitive dynamics (#18), understanding the relationships between risk and investment (#31), inducing underlying meanings associated with cultural events (#32), and classifying sets of data and consumers (#4).

The first topic reveals how researchers use topic modeling to compare hidden meaning structures in knowledge fields with networks of relationships among articles, journals, scholars, and citations. One approach has been to track the development of a journal or field by combining historical topic modeling analyses with bibliometrics and authorship networks (Cho, Fu, & Wu, 2017; Wang et al., 2015) to confirm field-level insights using patterns of dominant topics while rendering "hidden structures and development trajectories" (Antons, Kleer, & Salge, 2016: 726). This approach has been applied in science to track the rise and fall of meanings within a journal (Antons et al., 2016; Wang et al., 2015). For instance, Antons et al. (2016) used a semi-automated topic model combining both inductive (machine) analysis and abductive (human) labeling and generalization to add fine-grained detail to prior reviews of literature in the *Journal of Product Innovation Management*. Their topic model revealed latent meaning structures not identified in earlier reviews because the journal's interdisciplinary character made it difficult to identify and properly assess the breadth of articles published during its 30-year history.

A major benefit of the approach by Antons et al. (2016) is the ability to compare and contrast content according to classification schemes in the field and then induce categories of topics. They first applied the topic model analysis using LDA. After using methodological best practices and ensuring interrater reliability across 14 researchers, they clustered related topics into 6 semantically meaningful groups, including new ones the authors identified and labeled (once again, inductively) in correspondence with the interpretation and theory-generation stages depicted in our Figure 2. The authors then made an abductive, conceptual link to disciplinary trends—that is, they modeled "topic dynamics" by creating a weighting scheme. Finally, the authors combined this human-centered approach with a final and more automated deductive move, regressing topics that appeared more frequently than the median topics (those with a topic loading greater than 10 percent) for each year of their analysis, tracing topic development by comparing each of the topics against the mean, and in a final abductive iteration, classifying them according to trajectory shape ("hot," "cold," "revival," and "evergreen"). The result is a large-scale, many-to-many classification scheme across the entire study period that serves as a comprehensive semi-automated literature review, balancing meaningful knowledge categories with abductively rendered topics.

In another form of rendering in the classification of science, scholars have used topics as intermediate artifacts to perform social network analyses of authorship behavior. Cho et al. (2017) used topic modeling to augment coauthorship network data from 25 marketing journals over a 25-year period. Building on the work of Wang et al. (2015), who used topic modeling to map topic usage over time in the *Journal of Consumer Research* to predict promising research topics for the future, Cho et al. (2017) showed that social network analysis revealed two major communities of coauthors, whereas topic modeling analysis revealed three. They then used these intermediate analyses to show that communities of highly cited articles corresponded to heterogeneous clusters of related topics, but that the communities identified by each method had different features. In combining topic modeling with network analysis, Cho et al. (2017) showed how journals comprise the ecology of a field, but the structures constituting it (communities) can be seen at the levels of both citations and topics. Management scholars are not alone in using topic modeling analysis to advance field-level bibliometric studies, as it is being adopted in psychology (Oh, Stewart, & Phelps, 2017) and the humanities (Mimno, 2012) as well. Topic modeling has thus provided scholars with a way to both develop new understandings of cultural meanings and to connect those understandings with network and other structural features of fields.

A second topic relates to the role of categories in shaping competitive dynamics. Questions around optimal distinctiveness have long been of interest to management scholars (Deephouse, 1999; Navis & Glynn, 2011; Zhao et al., 2017), but this line of research is contingent upon the ability to measure coherence and variation of strategic action against the backdrop of a category. How to delineate categorical boundaries is thus a key concern. Haans (2019) explored the optimal distinctiveness of firm positioning relative to industry categories. He used topic modeling on texts from organizational Web sites to uncover the strategic positioning of firms in Dutch creative industries. The method enabled him to calculate both industry average and distinctiveness measures for individual firms. By using topic modeling to induce bottom-up, positioning-based classifications, Haans (2019) was able to generate new theoretical insights that diverged from prior research by suggesting that optimal distinctiveness for organizations depends on the distinctiveness of other organizations. Thus, positioning-based classification, as identified through topical analysis, has strategic implications. In related work, scholars have used topic modeling to develop important conceptual infrastructure in the form of inductive classifications for research on industry intelligence and competitive dynamics (Guo, Sharma, Yin, Lu, & Rong, 2017; Shi, Lee, & Whinston, 2016).

A third topic in this area identifies topic modeling as a means to derive categories of risk perception in finance. Such studies build on a long history of debates about the impact of corporate disclosures on investor behavior (Fama & French, 1993). Researchers have struggled to classify how risk factors are communicated and perceived by companies, analysts, and investors. In contrast to the established method of using predefined dictionaries for content analysis to quantify risk types [e.g., Campbell, Chen, Dhaliwal, Lu, and Steele (2014) using the schema: idiosyncratic, systematic, financial, tax, and litigation], researchers have applied unsupervised learning methods to financial texts to inductively classify risk factors. For example, Bao and Datta (2014) applied LDA to induce risk types from corporate 10-K forms, and then tested these against risk perceptions of investors, advancing theory by showing that the topic modeling-induced risk meanings better predicted investor perceptions of risk. Huang, Lehavy, Zang, and Zheng (2017) were able to extend this analysis to inductively identify risk factors and other economically interpretable topics within analyst reports and corporate conference calls, providing additional insights into how analysts both discover relevant information and interpret it on behalf of investors. In both of these articles, scholars used topic modeling to extend textual analyses of corporate financial disclosures by moving beyond the "how" (i.e., volume, sentiment, and length) to the level of topical meaning in terms of "what is the meaning of what is being said." Topic modeling thus has enabled researchers to develop better classification systems based on the textual data being sampled.

Another topic focuses on meanings associated with cultural events that are not captured by formal documents and artifacts. Miller (2013) used topic modeling to capture meanings around the nature of violence during the Qing Dynasty in China. Instead of relying on a fixed set of categories, the method enabled him to induce an original typology of violence based on the Qing administrator's perceptions of unrest. Similarly, Ahonen (2015) applied topic modeling techniques to challenge existing theory by inductively identifying the sources of legal traditions across countries. The author considered differences in legal language in government budgeting legislation as a basis for distinguishing between legal traditions. Both studies offer an approach to overcome biases associated with interpreting cultural events.

In similar articles, scholars have used topic modeling to study topic-based classifications in patent data (Kaplan & Vakili, 2015; Suominen, Toivanen, & Seppänen, 2017; Venugopalan & Rai, 2015). The practice of mapping knowledge structures in science is in its infancy, and the use of topic modeling has the potential to change how scientific fields are classified (Song, Heo, & Lee, 2015; Song & Kim, 2013; Yau, Porter, Newman, & Suominen, 2014) because topic modeling analyses do not perfectly correspond to formal systems of classification (Cho et al., 2017; Kaplan & Vakili, 2015). Topic modeling analyses also may reveal insights when used in conjunction with other forms of analysis such as citation and coauthorship patterns. As such, topic modeling can yield more fine-grained classifications and extend classic bibliometric and content analysis methods.

The articles we reviewed in this section map the knowledge spaces and dynamics of academic fields. Topic modeling enables scholars to compare latent topics in particular documents with preexisting bodies of knowledge and quantitatively measure broad trends in meaning, thus providing a counterpoint or corroboration of coding performed exclusively by humans. Because topic modeling is a rendering process based on human and algorithmic efforts, using it to map knowledge spaces uncovers latent classification systems that may or may not overlap with more formal classifications. Our review of articles in this subject area has resulted in the discovery of new concepts that can be used to better understand phenomena in a variety of management research streams.

Understanding Online Audiences and Products

For the last two decades, management theorists have been particularly interested in understanding how audiences evaluate firms and products in research on cultural entrepreneurship (Martens, Jennings, & Jennings, 2007; Navis & Glynn, 2010, 2011), status (Podolny, 1993), categories (Hannan et al., 2007; Zuckerman, 1999), and now, with the expansion of the Internet, understanding how these dynamics may change in online contexts (Mollick, 2014). These scholars have sought to understand the deeper patterns and meanings of producer communications and theorize audiences' reactions (Cornelissen, Durand, Fiss, Lammers, & Vaara, 2015). Nevertheless, isolating nuances both in the meanings of sensegiving communications (e.g., about products) and the responses of heterogeneous audiences remains difficult.

Topic modeling has been taken up by researchers—particularly in marketing—to analyze the cognitive content of online discourse about products and the behavior of online consumers as audiences. This subject area of understanding online audiences and products has emerged out of four topics: the nature of online consumer profiles (#12), online consumer brand recognition and preferences (#23), online customer evaluations and responses to them (#29), and enhanced topic modeling techniques on products and audiences (#13).

The first topic, the nature of online consumer profiles, has been advanced by conceptualizing consumers based on the clicking patterns of different online groups (Trusov, Ma, & Jamal, 2016), the network of related brands and brand tags clicked on by consumers (Netzer, Feldman, Goldenberg, & Fresko, 2012), and communities of consumers defined based on common virtual market participation (e.g., portals) or similar patterns of geolocation markers (Zhang, Moe, & Schweidel, 2017). In these studies, topics were rendered not just from a "bag of words" across a corpus of documents but from a "bag of behaviors" across a corpus of activities. This conceptual pivot maps roles to "topics" of behaviors. For example, click patterns for a group across diverse products/services during a particular time period offer unobtrusive measures of both a latent set of consumer profiles and their associated behaviors. Marketing studies using topic modeling have also uncovered evaluations by consumers in new ways. For instance, the work by Zhang et al (2017) on elite universities revealed that the willingness to tweet—and, even more importantly, retweet—about topics associated with a university reinforces the elite university status hierarchy. Ironically, the most elite of the elites receive more tweet outs and retweets, not only from their own members but also from members of other universities. Management scholars interested in categories (Durrand & Paolella, 2015; Vergne & Wry, 2014) and communities (Marquis, Glynn, & Davis, 2007) might use these reconceptualized online consumer communities to broaden theorization and measures of their core constructs. Scholars might also use online endorsements (clicks and tweets) to complement other forms of analyst assessments (Giorgi & Weber, 2015; Zuckerman, 1999).

A second topic is online brand recognition and preference. Here, scholars conceptualize brands not just as specific offerings with cachet, but as the associated networks of audiences linked to those products along with the sets of user-generated tags used by audiences to identify brand groups. For example, Nam, Joshi, and Kannan (2017) used topic modeling to render representative topics based on user-generated "social tags" from the shared bookmarking service Delicious. They then examined how Apple customers linked and endorsed Apple products via product tags, such as, "mac," "phone," and "Apple," all of which were linked to "Apple Corporation." The brand in its fullest form (Apple), then, was the overall network of linked tags used by customers. Similarly, Netzer et al. (2012) used car brand clicks on the online forum Edmonds.com to identify co-occurring words in topics about different car brands. The clusters of words (topics) revealed overlaps, evolving brand clusters, and "semantic networks" (i.e., meaningful text-based attributes) that differentiated brands. In addition, Netzer et al. (2012) were able to anticipate brand switches within and across these topic-based networks. They did so by studying changes in discussions about and associations among brands in these topic networks [also see Tirunillai and Tellis (2014)]. These rendering moves do not differ significantly from management theory approaches to fashion and design (Dalpiaz, Rindova, & Ravasi, 2016) and exemplar categories (Zhao, Ishihara, Jennings, & Lounsbury, 2018); management scholars working in this vein might broaden their understandings of how meaning is associated with brands and use topic modeling to augment their measures of templates and categories. In addition, given the association of brand and identity (Navis & Glynn, 2010; Raffaelli, 2018), management scholars might use group brand identification (as measured by topic preferences) to track identity formation and evolution.

A third topic focuses on the dynamics of influencing online consumers, or in other words, how agency is exercised online and with what effects. Marketing scholars, by and large, believe that online consumers are more difficult to understand and influence because they are decentralized, diverse, and switch often. Research identified as related to the topic of online consumer responses suggests that learning adjustment is due to latent structural modifications around topics captured by analyzing online forum data. For example, Puranam, Narayan, and Kadiyali (2017) used topic modeling to analyze all New York City restaurant reviews before and after the implementation of a regulation that required posting calorie counts; their results demonstrate a shift in online consumer evaluations, and in their view, food consumption patterns in New York City. More recently, Wang and Chaudhry (2018) examined online hotel ratings, and the effects of managers' responses to positive and negative customer reviews. They used LDA to generate a measure of response tailoring by comparing the content of managers' responses with a baseline value. Highly tailored managerial responses to negative reviews were considered by customers to be a form of high-quality complaint management; in contrast, tailored responses to positive reviews were considered to be overly promotional (hence, backfired on management). The use of topic modeling techniques to capture consumer evaluations and adjustments is of interest to management scholars engaged in cultural analysis and neostructuralism research (DiMaggio, 2015; Lounsbury & Ventresca, 2003; Mohr & Bogdanov, 2013) because a bedrock assumption in these culture-oriented approaches is that agency is less observable and more distributed. Topic modeling of online reviews across audiences can also help capture actor adjustments around latent structures [see Hannigan et al. (2019) and Heugens and Lander (2009)]. In addition, longitudinal, affect-based topic modeling might enrich studies of performance adjustment (Greve, 2003), anchoring (Ballinger & Rockmann, 2010), and event analysis (Morgeson, Mitchell, & Liu, 2015).

A final topic in this subject area is focused on improving topic modeling of online audiences and products to capture nuances of communication and audience responses (#13). The groundbreaking and oft-cited work by Lee and Bradlow (2011) regarding automated online reviews has several features that have become norms for rendering with topic modeling, such as using triangulation (e.g., with k-means clustering and multidimensional scaling [MDS]), mapping structures, thinking about "fit" with algorithms, and examining change over time. Recently, Guerreiro, Rita, and Trigueros (2016) and Jacobs, Donkers, and Fek (2016) introduced correlational topic models, sentence-based models, and hierarchical topic models to demonstrate the utility of using some supervision and structure in topic model rendering. Along similar lines, Büschken and Allenby

(2016) used sentences and phrases rather than words as inputs for LDA to show that topics based on them might exhibit less change (i.e., be "sticky") over time. Because management researchers are currently interested in understanding the interface of such methods and derived topics and meaning (DiMaggio, 2015; Schmeidel et al., 2018), The work of Büschken and Allenby (2016) poses an interesting rendering question for management researchers: Is stickiness a product of using sentences (the method) or is it due to linguistic meaning being constructed at the sentence- (rather than word-) level by online consumers?

To summarize, using topic modeling to analyze online audiences and products enables management scholars to think more deeply about the nature of online audiences (e.g., as click-based profiles, virtual networks, and computer-mediated communities); to reconceptualize products as distributed brands tied to evolving individual and category identities; and to capture the more subtle means by which audiences evaluate online products, and correspondingly understand how organizations might adjust in real time to those evaluations. In addition, the refinement of topic models of online audiences creates modeling standards for other topic modeling research, and encourages scholars to think more deeply about the meaning given to products by online audiences.

Analyzing Frames and Social Movements

Topic modeling also has been used to analyze frames and understand the dynamics of social movements. Management scholars have long been interested in symbolic management (Zajac & Fiss, 2006; Zajac & Westphal, 1994; Zott & Huy, 2007), such as understanding how investors respond to organizational framing efforts (Giorgi & Weber, 2015; Rhee & Fiss, 2014), theorizing the political dynamics associated with different framing strategies within firms (Kaplan, 2008b) and understanding the dynamics of social movements (Benford & Snow, 2000). This research requires scholars to identify frames—epistemological devices that actors use to organize experiences by answering the question posed by Goffman (1974: 8): "What is it that's going on here?"

Topic modeling methods have helped scholars expand theoretical boundaries in this area by providing an empirical method for inductively uncovering latent frames and then understanding the dynamics associated with frame proliferation and effectiveness. Our topic modeling analysis revealed four topics in this subject area: understanding how frames influence political processes (#27); the relationship between frames, context, and audience (#6); understanding field-level relationships between organizations, discourses, and strategies (#17); and social movement strategies, networks and actions (#11).

The first topic relates to how frames influence political processes. Frames enable actors to "render what would otherwise be a meaningless aspect... into something that is meaningful" (Goffman, 1974: 21). Scholars are particularly interested in the often political and contested dynamics associated with framing (Fiss & Hirsch, 2005; Kaplan, 2008b). An exemplar article showing how topic modeling can contribute to this research stream is the study by Fligstein et al. (2017) on the Federal Open Market Committee's decision-making processes in public meetings. Specifically, they sought to develop a theory to explain how the committee failed to appropriately perceive the risks to the economy in the months leading up to the financial crisis. In addition to confirming the existence of macroeconomics as a master frame, their topic modeling approach revealed the existence and application of a banking frame and a finance frame. By focusing on the specific events—the housing bubble and the financial crisis—the researchers were able to track which frames came to dominate Fed committee discussions at the time of each event. The authors thus used topic modeling to develop a theory that explains how a predominant frame can blind actors involved in decision-making processes.

A second topic explores the relationship between frames, context, and audience. Actors use distinct frames to advance their interests (Kaplan, 2008b) and seek to create effective frames through mechanisms such as frame alignment (Snow, Rochford, Word, & Benford, 1986) or frame resonance (Snow & Benford, 1988). In an exemplar article, Levy and Franklin (2014) used topic modeling as a means of identifying distinct discursive frames. Specifically, they used a study of political contention in the U.S. trucking industry regarding hours of service to inductively analyze the frames that emerged from a study of comments on a public Web site. They were able to use topic modeling to uncover distinct differences between individual and organizational uses of frames in the debate, showing how different parties used different frames to promote their interests. Uncovering nuanced distinctions in framing content deployed by different parties over time can help researchers generate new theory about the influence of communication content and techniques on political processes.

The third topic relates to research on field-level relationships between organizations, discourse, and strategy. Specifically, to understand framing effects, it is often necessary to move beyond the content of a specific frame. To illustrate, Bail, Brown, and Mann (2017) explored the relationship between conversational and emotional styles in advocacy work—seeking to incorporate sentiment analysis into our understanding of frames. The authors used topic modeling to classify the types of topics raised by autism advocates and used LIWC to capture sentiment and bias in normalized spaces. This unique combination of topic modeling and LIWC sentiment analysis enabled them to reveal the cognitive and emotional "currents" running through advocacy groups and to show how the ability to "dispatch messages that contribute to a phase shift [between emotional and cognitive-focused communication]" ultimately leads to more effective results (Bail et al., 2017: 1205). Thus, topic modeling has enhanced our understanding of frame effectiveness in the context of broad field-level relationships between organizations, discourse, and strategy.

Similarly, the fourth topic relates to researchers' attempts to understand the relationship between social movement strategies, networks, and actions. For example, Almquist and Bagozzi (2017) sought to understand the network relationships between radical environmental activists in the United Kingdom. Based on a longitudinal corpus of a radical social movement's texts, they identified the centrality of network ties and then used structural topic modeling (STM) to locate the groups and the positions they took on various radical issues, thereby enabling them "to evaluate whether the presence of a given group tie (or cluster member) significantly increases the attention dedicated to a given topic" (Almquist & Bagozzi, 2017: 26). By combining STM and network analysis, the authors were able to classify subnetworks of actors to develop a better theoretical account of the discursive actions and network relationships of social movements by mapping unseen or hidden ties. Put another way, topic modeling generates theoretical artifacts that facilitate researchers' efforts to connect the content of communications with other theoretical constructs.

In summary, topic modeling provides several benefits that have led to significant theoretical advancements related to frames and framing. First, topic modeling has helped researchers strengthen their understanding of frames. For example, scholars can use topic modeling to track the prominence of researcher-derived high-level frames for large corpora over an extended period of time. In addition, the algorithmic nature of topic modeling approaches ensures the replicability of identified frames. Second, the inductive nature of many topic modeling techniques enables the discovery of unanticipated frames and audiences that use them, providing a powerful opportunity for scholars to generate new theory. Specifically, topic modeling methods enable researchers to understand the dynamics associated with the copresence of competing voices within a single text (i.e., heteroglossia, Bakhtin, 1982), which provides researchers with a way to study multiple competing or collaborative frames. Finally, topic modeling facilitates the creation of new theory because it produces theoretical artifacts that can be paired with other forms of analysis such as sentiment analysis or network analysis.

Understanding Cultural Dynamics

Management scholars have sought to leverage psychological and sociological research on culture—"the interaction of shared cognitive structures and supraindividual cultural phenomena (e.g., material culture, media messages, or conversation) that activate those structures" (DiMaggio, 1997: 264)—to explain diverse phenomena. For example, in research on institutional logics (Thornton et al., 2012), strategic action fields (Fligstein & McAdam, 2011), and professions (Abbott, 1988), scholars have theorized the evolution and impact of cultural meanings at the level of an institutional field. In research on organizational culture (Hatch, 1993) and organizational identity (Gioia & Thomas, 1996), scholars have theorized the evolution and impact of cultural meanings at the level of the organization. In research on cultural entrepreneurship (Lounsbury & Glynn, 2001, 2019; Martens et al., 2007) and institutional work (Lawrence, Suddaby, & Leca, 2009), scholars have attempted to understand how individuals leverage cultural material to achieve strategic objectives. In all of these areas, researchers have attempted to theorize both the dynamics of cultural influences and the evolution of cultural concepts.

Overall, this research on culture has faced significant challenges. One such challenge relates to the measurement of cultural constructs. For example, scholars have defined institutional logics as "the socially constructed, historical pattern of material practices, assumptions, values, beliefs, and rules by which individuals produce and reproduce their material subsistence, organize time and space, and provide meaning to their social reality" (Thornton & Ocasio, 1999: 804). But in empirical studies, it has been harder to specify them. A second challenge is to understand the temporal dynamics associated with culture. For example, in cultural entrepreneurship research, scholars attempt to understand how entrepreneurial organizations are able to legitimate a new market category over an extended period of time (Navis & Glynn, 2010). Researchers also attempt to connect cultural meanings with events and actions, for example, by connecting the content of organizational discourse with changes in organizational networks and broader social discourse (Bail, 2012).

Scholars have used topic modeling methods to push the boundaries of our understanding of such cultural dynamics. Our analysis reveals five themes in this research: understanding the professionalization of a field (#2), using topic modeling to analyze big data to understand cultural trends (#5), understanding dynamics associated with literary meanings (#9), understanding how cultural meanings change over time (#19), and understanding the evolution of cultural trends (#28). Topic modeling has enabled scholars to generate novel theory by providing an operational means to identify cultural concepts and then trace the evolution of those concepts over time and across different locations of social space.

The first topic in this area revolves around developing new theory about the professionalization of fields. Specifically, Croidieu and Kim (2018) theorized the rise of alternate fields and quasi-professions by studying the emergence of U.S. wireless radio broadcasting field and the "lay professional legitimation" of amateur radio operators from 1899 to 1927. To understand the legitimation process for amateur operators, the authors had to gather a wide, diverse constellation of documents from various archival sources: U.S. government regulations, radio operators from the era, radio corporations, and the *New York Times*. They analyzed the distribution of topics over time and by audience to determine the meanings of those patterns using historical (or case) records. This process enabled the authors to identify first- and second-order mechanisms by period. They paired topic modeling of diverse archival materials with standard historical reading and complementary content analysis to create and defend a theoretical account of professionalization based on historical data.

A second topic focuses on how big data can be used to understand cultural trends. These articles describe and illustrate nuances of the processes scholars use to extract meanings from large corpora. For example, Wagner-Pacifici, Mohr, and Breiger (2015) summarized a special issue in *Big Data & Society* on assumptions of sociality that synthesized the results of several other subjects. First, they highlighted the importance of recognizing that big data methods, unreflexively applied, can lead to biased results. Second, they discussed the importance of the interpretive role of analysts who use big data and related methods to generate theory. Third, they emphasized how big data methods require a move away from traditional deductive science, highlighting their inherently inductive and abductive nature. Finally, they showed how analyzing big data requires scholars to ask fundamental questions such as "What is a thing? What is an agent? What is time? What is context? What is cause?" (Wagner-Pacifici et al., 2015: 5). Thus, scholars must reflexively consider the cultural implications of studying big data.

Interestingly, in sociological research that has provided analogical inspiration for management scholars, Mohr and Bogdanov (2013) used topic modeling to analyze literary meanings. In the humanities, Tangherlini and Leonard (2013) introduced a technique called subcorpus topic modeling to compare canonical texts with broader literature and societal discourse. Specifically, they used the technique to "develop a well-curated topic model of a subcorpus" and then used "the ensuing model to discover passages from the large, unlabeled corpus" (Tangherlini & Leonard, 2013: 728). To illustrate the utility of their method, they showed how topics associated with Charles Darwin's intellectual ideas penetrated "into the broader literary world" (Tangherlini & Leonard, 2013: 735). They thus used topic modeling to understand topics associated with well-known texts and then applied the outputs to analyze other, less well-known cultural meanings.

Another evident topic focuses on how cultural meanings evolve over time. An example of this can be seen in the work of DiMaggio et al. (2013), who identified the frames invoked and crafted by news outlets in their coverage of the public controversy surrounding the U.S. government's support of artists and art organizations. The authors rendered corpora using data from five mainstream media outlets; after applying unsupervised LDA to isolate and link topics, they inductively identified different frames. Their results reveal not only the differences across frames by time period but also how a single text produced by these media outlets might use multiple frames. Applying a fractional multinomial logit analysis, they calculated the expected relative prominence of topics based on their LDA analysis. By further aggregating those topics into particular topic groupings, then classifying them as conflict or comparison frames, they were able to reveal the likely link between the relative increase in conflict topics that accompanied the growing sentiment against public funding for U.S. arts organizations starting in the 1980s. The authors thus used topic modeling to identify different frames of cultural meaning in the public sphere and then showed how these meanings changed over time.

A final topic looks at the impact of cultural meanings on societal actions. For instance, Marshall (2013) sought to understand the evolution of cultural trends by contrasting how different academic theories of demography unfolded over a 60-year period in Great Britain and France. Specifically, she used correlated topic modeling (to account for the assumption that topics in her corpus might be correlated across documents) to understand how concepts associated with fertility were understood (and unfolded) differently in different cultural contexts. She used topic modeling to identify topics, measure the prevalence of those topics in the corpus, and then connect those topics to the dominant theories of demography in effect during that time. The topic modeling analysis enabled her to identify differences between the responses of French and British academics to changing demographics during the study period. Topic modeling thus enables scholars to trace the evolution of cultural trends by connecting the prevalence of themes in discourse to historical events.

Overall, topic modeling has provided management scholars with a new methodology for generating novel insights about cultural dynamics. First, topic modeling provides a means to develop an unbiased understanding of the prevalence of distinct cultural concepts over an extended period of time, thereby enabling scholars to measure cultural concepts more precisely. Second, topic modeling enables scholars to compare a well-known subset of knowledge to broader corpora that might reflect that knowledge structure more generally, thereby enabling scholars to develop new theories and link constructs that previously had been difficult to connect, both empirically and theoretically. Similarly, topic modeling enables researchers to see how different meanings within the discourse surrounding a particular topic exist and shift over time. Finally, topic modeling can connect shifts in discourse to broader cultural trends.

NEW TRENDS RELATED TO TOPIC MODELING AND RENDERING

Many new trends in management and computer science research are relevant to management scholars' use of topic modeling to render corpora, topics, and theoretical artifacts (see Figure 2). Each trend within a rendering process has a unique trajectory that is important to discuss and respect. For instance, some trends broaden specific rendering processes (e.g., creating corpora), whereas others deepen them (e.g., fitting topic models). Trends also involve some of the aforementioned management subject areas. In this section, we discuss not only trends but also their implications for rendering and building management knowledge.

Trends in Rendering Corpora

As management researchers embrace approaches that move beyond dictionary-centric content analysis, corpus selection becomes an even more critical step in topic modeling research. Recent articles on text analysis reveal a broad effort to engage more closely both with computational linguistics and NLP (Kobayashi et al., 2018; Schmeidel et al., 2018). These efforts were precipitated by an important shift toward conceptualizing corporal dimensions to enable comparison.

Corpus linguistics. Within management, this trend of engaging with computational linguistics is most evident in a recent special issue of Organizational Research Methods (Tonidandel, King, & Cortina, 2018) on big data and modern data analytics. This special issue demonstrates the arc of preprocessing corpora as a precursor to higher order text analyses with big data (Kobayashi et al., 2018; Schmiedel et al., 2018). However, many of these preprocessing techniques were highlighted several years earlier by Pollach (2012), who pointed management researchers to a branch of linguistics known as "corpus linguistics" to show how word patterns can lead to meaningful insights by virtue of the corpora in which they appear. Techniques for analyzing corpora themselves—both qualitatively and quantitatively—include word frequency lists, keyword-in-context searches, comparison of corpora, word collocations, and statistical methods for assessing word-frequency patterns.

Pollach (2012) originally positioned corpus linguistics techniques as methodological innovations for content analysis. In very recent work, Kobayashi et al. (2018: 1) took a broader approach, suggesting that such preprocessing considerations represent a "fundamental logic" of mining "text data." As part of that mining, articles in this vein have stressed the imperative of preprocessing as "wrangling" text data into a corpus (Braun, Kuljanin & DeShon, 2018). Schmiedel et al. (2018) have laid out some steps that recognize the fundamental importance of data collection and cleaning in topic modeling analysis. Theoretically speaking, these articles draw on core ideas from linguistics, such as the famous distributional hypothesis (Firth, 1957)—that is, "words that occur in the same contexts tend to have similar meanings" (Turney & Pantel, 2010: 142). Inferring meanings, in other words, depends on the context created by the corpus. As a result, these recent articles are raising the bar in terms of the level of sophistication and reporting standards required for scholars who use topic modeling and other text analysis methods.

In fact, we built our rendering process on the insight that corpora curation has implications for theoretical work because meaning is inferred from context. A source corpus begins as natural language, which can be messy and thus requires selecting and trimming. These two steps standardize documents, which then enable topics in the corpus to be rendered at a higher level of abstraction. Moretti (2013) called this "distant reading," where a corpus can be fully and adequately represented in terms of topics. Sharpening this reading requires iteration; for this reason, our rendering process has an arrow pointing back from rendering topics to rendering corpora. The trends we identified in preprocessing point to the adaption of techniques from corpus linguistics for the purposes of corpus curation, thereby expanding the toolkit for rendering.

NLP. Innovations in NLP are advancing how scholars prepare and preprocess the words in corpora. NLP research highlights two key concerns: first, as the base unit of meaning, a token (a word, parts of words, or phrase combining words) is a function of grammar; and, second, structures of grammar are embedded in sentences, which have codependencies across words and paragraphs within a document. Uttered meanings correspond to parts of speech. For example, the meaning of the token Google changes based on whether it is a noun (i.e., referring to the company or software), or a verb (i.e., referring to use of the search engine), and can be referred to in a similar manner through a pronoun in a subsequent sentence. Thus, a token as a unit of meaning may be a word or multiple words (i.e., a phrase) (Chomsky, 1956).

NLP research suggests that latent meaning in texts can be captured by bigrams, or two-word units rather than individual words, as in the standard "bag of words" approach (Manning, Raghavan, & Schütze, 2010). Some management researchers have, therefore, shifted the unit of analysis to a "bag of sentences" (Bao & Datta, 2014; Büschken & Allenby, 2016). Determining the boundary of analysis is technically tricky. For example, because a sentence break is not just a function of searching for the full stop character (i.e., "."), researchers have developed NLP methods to determine sentence boundaries in a common task called sentence segmentation (Kiss & Strunk, 2006). Moreover, advanced deep learning algorithms (e.g., neural networks) are being introduced that go beyond "bag of words" approaches altogether to consider syntactic position and context when identifying linguistic structures such as constituency and dependency parsing representations (Manning et al., 2014). Deep learning is an unsupervised algorithm that can be trained on large text corpora to "learn" latent structures, including semantic compositionality (Socher et al., 2013) within texts (or other kinds of data) that can then be used for explanatory or predictive purposes.

Additional advances have improved the precision of identifying tokens. For example, mentions of individual actors may be standardized by using NLP technologies such as Named Entity Recognition (Mohr et al., 2013) and coreference resolution (Manning et al., 2014). The former is an NLP method that can automatically identify entities based on their appearance in texts and can annotate analytical codes as actors, organizations, and countries. The latter is an NLP tool that can extend named entity recognition to pronouns and other references to entities across sentences. Standardizing entities to resolve ambiguities inherent in manifested natural language facilitates machine-based reading.

Approaches to making such transformations are particularly salient in topic modeling because this trimming determines the token unit upon which topics are established (Schmiedel et al., 2018). These decisions regarding rendering corpora have theoretical implications. The NLP methods discussed here are largely inductive tools, with machine learning algorithms annotating texts. Although inductive methods have become more widely accepted in management journals, there is still considerable risk of over-fitting findings to the data if scholars generalize too quickly (i.e., engage in "theoretical over-fitting") (Tchalian, 2019). Thus, researchers must continue to check the validity of such annotating.

Non-Western languages. Another new corpus-rendering trend that touches on these developments in corpus linguistics is the treatment of languages that are structurally dissimilar to most Western languages—in particular, languages without spaces between words (or, scriptio continua), including many Southeast Asian writing systems (e.g., Thai, Burmese, and Lao) and those that use Chinese characters (i.e., Chinese and Japanese). Treatment of these languages is not straightforward. For example, each Chinese clause can be recognized as a group of characters. Each Chinese character corresponds to a syllable; although some characters represent individual (i.e., one-syllable) words, many words consist of more than one character. These linguistic features make preprocessing necessary to ensure effective topic modeling and theorizing, thereby enabling the algorithm to identify the tokens that comprise the texts.

The traditional content-analytical method of using preset dictionaries to match characters with possible words in the corpus confronts computational problems, and the permutations and ambiguities of language often lead to poor results. Customized dictionaries improve fit, but still yield substantial inaccuracies (Allen et al., 2017; Slingerland, Nichols, Neilbo, & Logan, 2017). Today, statistical and machine learning models are complementing, if not replacing, preset dictionaries. These models build internal lists of words by training algorithms through iterative learning. This training can be performed using extant language libraries (e.g., the People's Daily Language Library) to segment unknown texts.

The introduction and development of these methods has opened the door to using topic models to investigate a wide range of novel data sources and cultures. For example, Huang, Li, Zhang, Liu, Chiu, and Zhu (2015) used topic modeling to analyze one of China's biggest online social network platforms, Weibo, to track the real-time ideation process of suicide, which is traditionally assessed by surveys and interviews and thus suffers self-reporting and retrospective biases. Their approach has shed new light on future studies of various ideation processes such as entrepreneurial ideation.

Such word segmentation processes also make comparative analysis and theorization of multiple-language corpora feasible. In particular, with appropriate preprocessing, topic models can be used to analyze the diffusion and translation of new ideas, frames, and categories crossing national borders. For example, the cross-national diffusion of corporate social responsibility (CSR) has attracted scholarly attention (Kim & Bae, 2016; Lim & Tsutsui, 2012). But identifying the extent to which CSR has been locally translated and innovated would require fine-grained analysis of multiple-language corpora, which topic modeling can facilitate. Because the topic outputs from non-English corpora must be translated into their English equivalents to be used in comparison and theorization, and because the cultural context still matters for those identified topics, such comparative projects are best developed by teams with at least one researcher who knows the language and culture and can apply that knowledge to help validate the rendering of the corpora.

Summary. New trends in rendering corpora hold great promise for addressing the technical and theoretical limitations of current topic modeling approaches. They show that corpus selection as well as lemmatizing and other forms of corpus preparation have theoretical implications and, therefore, must be explicitly discussed in the methods sections of articles, likely under the aegis of "data preprocessing." The use of foreign languages only magnifies these challenges, just as they do in any form of archivalism applied to other cultures.

Trends in Rendering Topics

Researchers are continuing to refine how topics are rendered in an effort to manage the degree of supervision required and how fit can be defined. In Figure 2, we show how the rendering of topics revolves around the criteria for identifying robust, applicable topics (i.e., around supervision and fit criteria). Supervision and fitting, in turn, depend on the form of theorizing taken—inductive, abductive, or deductive—with induction aligning with less supervision and fitting than deduction.

Integrating topic rendering with other approaches. Many scholars today are finding that topic modeling works best when integrated with other methods of analysis, which has implications for the rendering of topics. One recent style of work covered by labels such as "big qual" (Davidson, Edwards, Jamieson, & Weller, 2019) and "RiCH (Reader in Control of Hermeneutics)" (Breiger et al., 2018) gives the interpretive human reader primacy but leans on the affordances of computational tools for forming rich representations of topics. Other styles in recent work integrate topic modeling with more traditional deductive methods (Haans, 2019; Hankammer, Antons, Kleer, & Piller, 2016; Roberts et al., 2014), where topics are rendered according to a logic of variable coherence. Topic modeling in these correlational analyses seems to rely on a parsimony principle, where topics are presented in articles as tables with applied labels and fewer than 10 highly associated words per topic (i.e., Schmiedel et al., 2018). Our reading of this trend reveals that the dominant method in the research design affects how topics are rendered.

Recent trends in topic modeling within management research have also shifted attention toward alternate ways of capturing latent patterns to reveal new (sometimes provisional) meaning structures that change over time. The LDA-based analyses we reviewed in this article mostly followed a pattern of rendering one set of topics in a corpus. Through iterative steps in the rendering process, Hannigan et al. (2019) found that a key topic in a scandal's media coverage was changing because of the disclosure of a social control agent's judgments of wrongdoing. To overcome this challenge, they split their corpus in two, rendering topics across each subcorpus. They used the word–topic matrices from both models to find comparable topics, which they subsequently used as independent variables representing media effects of a scandal in event history models at different time periods. Similar efforts to periodize data can be seen in work by Croidieu and Kim (2018). We see such efforts as contextualizing topics in ongoing theoretical concerns.

As another example, Cho et al. (2017) embedded topic modeling with other commonly used methods of conducting a literature review. The concept of topic was used to approximate an "author community" of researchers exhibiting certain topics prominently in their work. This framing affected the logic of how they rendered topics. They rendered latent author communities using topic modeling against those derived using bibliometric network analysis to show similarities and differences in approaches, but this comparison governed the validity of topics rendered. Alternative analytical approaches that help generate theory (Bail, 2012; Kennedy, 2008), especially emergence processes, also promise the ability to better articulate latent patterns to reveal hierarchical linguistic structure (Mohr et al., 2013). Therefore, the rendering of topics is part of the overall theory generation process itself.

STM. Just as LDA disrupted latent semantic indexing (LSI) (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990), scholars are attempting to modify LDA by improving fit algorithms and making it more structured and systematic. One major development is STM (Bail et al., 2017; Roberts et al., 2014; Schmiedel et al., 2018), which extends LDA by incorporating metadata about documents, such as who wrote each text and when or where they were written. This information can be re-applied to the topic estimation procedure and help improve model fit. In so doing, STM enables researchers to identify relationships not just between topics and documents but also between the producers of documents and the texts and topics. It can be used in a linear regression framework to analyze specific metadata (as covariates) to identify statistically significant relationships to each topic. It can also be used in mixed methods approaches such as with critical discourse analysis to tie textual data analyzed using topic models with richer qualitative analysis (Vaara, Aranda, Etchanchu, Guyt, & Sele, 2019).

In recent working articles appearing in Academy of Management Annual Meeting Proceedings, researchers have adopted mixed STM approaches. For instance, Aggarwal, Lee, and Hwang (2017) used topic modeling to operationalize review diversity in Yelp reviews to show that status gains are correlated with higher quality reviews and nonelite conformity to those same reviews. Likewise, Karanovic, Berends, and Engel (2018) used topic modeling to study actors' perceptions of "platform capitalism" (Davis, 2016) in a popular online forum for Uber drivers. Their analysis reveals consistent patterns in a large corpus representing more than 120,000 forum posts and shows that drivers' reactions can both contribute to and critically evaluate the legitimacy of a new organizational form, despite being imposed from the aforementioned.

Hierarchical LDA. Another promising extension to LDA topic modeling is hierarchical LDA (hLDA) (Blei, Griffiths, & Jordan, 2010). Whereas LDA traditionally requires that a researcher set the number of topics (the k parameter), hLDA can generate the optimal number of topics based on other researcher-defined parameters, such as the number of hierarchical levels and number of terms per topic. Although different software implementations of hLDA use different algorithms to generate the hierarchical models, generally speaking, the hLDA algorithm generates a set of subtopics after identifying an aggregate topic. The algorithm then "reshuffles the deck" by reclassifying documents or document segments into synthetic document groupings and rerunning the algorithm for each grouping to generate additional subtopics. The result is a hierarchy representing the topics and subtopics, or subdimensions, of the texts being analyzed.

The ability to generate a hierarchical representation of the internal structure of a discourse can provide substantial theoretical insights. Tchalian, Glaser, Hannigan, and Lounsbury (2019) are using hLDA to identify the competing and complementary messaging efforts of stakeholders in the emergent electric vehicle (EV) industry: automobile manufacturers, newspaper reporters, automotive experts, and government officials. The hierarchical structure of the hLDA output is enabling Tchalian et al. (2019) to trace both the longitudinal appearance of different topics involved with the construction of the emergent EV category and their prominence within the discourse. This approach allows them to define the theoretical concept of "institutional attention"—the field-level convergence that both isolates and aggregates the various interests involved in the social construction of the EV as a market category. The hierarchical arrangement of topics in their article and others (Smith, Hawes, & Myers, 2014) reveals not only the primacy of ideas over time but also the sociocognitive meaning structures emphasized in cultural sociology (Mohr, 1998) and content analysis (Duriau et al., 2007), thus highlighting the great potential of topic modeling approaches for generating novel theoretical insights.

Summary. Advances in rendering topics have broadened topic modeling's use by pairing it with other techniques and deepened its use by creating variants that structure topics (e.g., hLDA). Rendering topics, at least for the near future, appears sufficiently robust to work with developments in near variants such as NLP and specific machine processing algorithms (i.e., "trained" algorithms in specific domains). These trends have the potential to extend the theoretical deltas we identified in our analysis of management subject areas. However, applying new algorithms for topic modeling and determining proper logics of fit and validity also raises important questions about research design. For example, use of STM reinforces critical decisions about appropriate measurement and variation in econometric-based approaches, and hLDA simply shifts a researcher's interpretive choices from determining the number of topics to deciding the number of levels and words per topic. These advances demonstrate that the most powerful path of development in topic modeling is not to displace but rather complement traditional research designs by enabling the use of different approaches to abstract and measure phenomena using text.

Trends in Rendering Theoretical Artifacts

Trends in rendering theoretical artifacts may offer the richest, most open-ended area of development in the field. Three trends are of particular interest: delineating latent structures, mapping new meaning, and blending artificial intelligence (AI) with human supervision to generate new artifacts. Each trend has been pursued using a range of theorizing approaches from inductive to deductive, and each has the ability to both extend and build theory, as indicated by the iterative arrows in Figure 2.

Latent structures and the "new structuralism." Increasingly, scholars are using topic modeling to assess structural relations in fields (Bail, 2014; Jha & Beckman, 2017; McFarland, Ramage, Chuang, Heer, Manning, & Jurafsky, 2013). Structural artifacts formed through rendering may enable theorists to identify new mechanisms for uncovering organizational or institutional structures, including those flexible enough to allow for a variety of instantiations in studies of fields (Lounsbury & Ventresca, 2003). The central thread relates to the use of topic modeling to map cultural dynamics around social structures. A macro approach involves mapping the meaning structures that comprise business environments (Pröllochs & Feuerriegel, 2018), knowledge profiles of firms (Suominen et al., 2017), emerging fields (Hannigan & Casasnovas, 2019), and political issues (Kim, Ahn, & Jung, 2018). Researchers have modeled the topics and rhetorical attributes of scientific articles, in turn finding links between the hidden topic structure of scientific communities as "thought collectives" and impacts on knowledge consumption patterns (Antons et al., 2018). Others have identified the "backstage" influences of stakeholder groups in the sustainability movement in higher education and have used measures of discursive distance to identify field-level coherence (Augustine & King, 2017).

More micro approaches involve modeling the formation of social network ties using topic-based proximity measures (Lee, Qui, & Whinston, 2016), or tracking the signatures of content authorship using author-topic models (Rosen-Zvi, Griffiths, Steyvers, & Smyth, 2004). Scholars are using these micro approaches to revisit a classic question in social science: How are social structures and meanings coconstituted? Lee et al. (2016) considered the mechanism of homophily in network formation by topic modeling texts of user-generated biographies and their associated tweets. In turn, they found that people with similar topic vectors were more likely to check-in to the same locations and form similar online social network ties. Rosen-Zvi et al. (2004) used an extension to LDA to model the contents of documents and authors' interests. They created the "author-topic model" artifact, which can be used to compare documents for similarity and applied to automatically match article authors to reviewers. In each of these articles, researchers used topic modeling to render and theorize structural dimensions as artifacts.

Scholars are extending the new structuralist approach by using topic modeling to analyze dynamics of culture and meaning (Lounsbury & Glynn, 2019; Mohr & Bogdanov, 2013). The simultaneous rendering of topics and contents of identified topic clusters reveals how social structure and meanings can be co-constituted at the field level. An example of a classic approach in this style of work is an exploration of "grass-fed beef" (Weber, Patel, & Heinze, 2013) as a construct that conveys particular meanings and describes the evolving structure of a market. Topic modeling enables social structures and meanings to be studied in new ways. Hannigan and Casasnovas (2019) used topic modeling and named entity recognition to map the co-occurrence of actors and topics appearing in media coverage to identify the spatial and temporal arrangements of an emerging field. Following classic works in the new structuralist tradition (i.e., Mohr & Duquenne, 1997), Hannigan and Casasnovas created incidence matrices of topic and actor co-occurrence and used them to generate maps of hierarchical Galois lattice structures. These lattice artifacts are visual maps that demonstrate co-constitution by showing the nesting of substructures formed through two modes of analysis. Mohr and Duquenne (1997) used lattices to show how practices and meanings co-constituted institutional logics, whereas Hannigan and Casasnovas (2019) used lattices to reveal the types of actors and topics co-constituting spatial and temporal arrangements in field formation. Advances in relational topic modeling (Chang & Blei, 2009; Gerlach, Peixoto, & Altmann, 2018) that identify document networks are also being used to render more document-based theoretical artifacts, perhaps representing different audience perspectives. These audience perspectives, including those captured using STM, enable latent structures among knowledge creators to be identified.

Bringing back meaning. Although topic modeling provides tools for extracting and presenting constellations of words and phrases that appear in patterns across documents in corpora, the question of whether such topics represent meaning structures is an important one (Mohr, 1998). During the initial analytical stage, analysts interpret topics based on logics of fit and interpretability. However, presenting topics without careful concern for theoretical artifacts risks presenting disembodied arguments about meaning. Thus, a naive machine learning analysis may omit important distinctions if applied crudely. An important topic modeling trend thus centers on how to capture meaning and meaning structures.

Organizational scholars have long been interested in studying meanings, particularly in light of recent concerns about measuring the construction and deployment of culture (i.e., Gehman & Soublière, 2017; Lounsbury & Glynn, 2019; Weber & Dacin, 2011). Although topic modeling–based research promises the potential to study cultural dynamics with increased scale and precision, scholars acknowledge that the technique must be paired with a respect for symbolic and social boundaries (Lounsbury & Glynn, 2019; Mohr et al., 2013). For example, Mohr et al. (2013) pointed to Burke's (1945) classic analytical structure of the pentad to study scenes of action. They used topic modeling and NLP to study the pentad in a corpus of U.S. national security documents. Analytically, they used named entity recognition to map actors, topic modeling to identify scenes, and NLP-based semantic grammar parsers to identify acts. Other scholars have described the utility of applying related computational methods such as semantic network analysis to contextualize topic modeling through theoretical artifacts (Carley & Kaufer, 1993; Diesner & Carley, 2005). Combined with a concern for theoretical artifacts, topic modeling thus opens the door to rendering modes of meaning, such as observing connotations and denotations of an institutional field.

Blending topic modeling and AI. A third fertile area of enhancing the theoretical artifacts built with topic modeling lies at the intersection of artifacts derived from AI and those derived from topic model rendering. AI and the deep learning models on which it is built can be blended with topic models in at least two ways. First, in the class of AI models known as "deep neural networks," two relevant methods enable blending with topic modeling: convolutional neural network (CNN) methods and recurrent neural network (RNN) methods. Unlike machine learning models such as LDA that use minimal inferences about context, these models retain more contextual information and thus are becoming increasingly relevant for social science researchers. They are more appropriate for dealing with streaming data such as Facebook updates and Amazon reviews, in which local contexts (e.g., prior words in a word sequence) affect the position of each topic term (Jin, Luo, Zhu, & Zhuo, 2018). Combining these methods with topic models may enable a more complex and dynamic rendering of theoretical artifacts such as frames, logics, and the latent value orientations discussed previously. When applied to large text corpora, both CNN and RNN are particularly effective in managing the tradeoff of specificity, enabling the analysis and modeling of latent structures that better balance under- and over-fitting. Moreover, they may help generate entirely new theoretical artifacts to help identify and explain social and role structure, partisanship, ideological contestation, discursive fields, and other sociocultural structures and institutional regimes more dynamically.

Second, deep learning can be integrated with topic models to analyze images—alone or along with verbal text—which opens a new path to rendering theoretical artifacts. Whereas verbal text is descriptive, linear, additive, and temporal, images and visual features are embodied, spatial, holistic, and simultaneous, which defies conventional analytical techniques. The integration of deep learning into topic models creates potential for future theoretical development that considers both visual features and verbal text (Krizhevsky, Sutskever, & Hinton, 2012). In particular, scholars have argued that the role of visual features in the process of institutionalization is significant, but largely underexamined (Meyer, Jancsary, Höllerer, & Boxenbaum, 2017).

In other words, deep learning helps manage tradeoffs around specificity and configuration and represents an effective solution to the ever-present issue of theoretical parsimony, but it also comes with a caution. Because deep learning is a computationally inductive modeling tool, many of its operationalizations are "black boxed," making its feature permutations challenging to reconstruct mathematically. It ironically highlights the tradeoff of human supervision and reinforces the need to apply it along with other analytical techniques within a mixed-methods approach to generating theoretical artifacts.

Summary. All three new trends in topic modeling—eliciting latent structures, capturing meaning, and using AI to help generate theoretical artifacts—open up new avenues for theory building. They complement the agnostic assumptions about meaning that are embedded in the LDA algorithm and, in this way, echo how trends related to corpora selection and trimming and to supervising and fitting topics are helping scholars overcome some of topic modeling's foibles while preserving its power. In particular, by revealing latent patterns and meaning structures, topic modeling is increasingly able to generate social, cultural, and political constructs that define evolving cultural meanings, discursive fields, and political action.

FROM THE BALCONY

Topic modeling, a method adapted from computer science, "represents a novel tool for analyzing large collections of qualitative data in a scalable and reproducible way" [Schmiedel et al., 2018: 3; see also Kobayashi et al. (2018)]. Our review reveals that topic modeling has been used in surprisingly diverse ways by management scholars, demonstrating that it is a malleable methodological and theoretical tool for tackling a variety of research questions. Although many articles we examined described the technical underpinnings of the LDA algorithm, we found that topic modeling practices are part of an often-implicit process of rendering corpora, topic models, and theoretical artifacts from raw data. We applied topic model rendering in this review to curate and make sense of the topic modeling corpus in the management literature. Our analysis reveals that topic modeling is gaining steam in management research (see Figure 1), particularly in five areas: detecting novelty and emergence, developing inductive classification systems, understanding online audiences and products, analyzing frames and social movements, and understanding cultural dynamics. Topic modeling has both strengthened knowledge in each area and enabled scholars to explore subjects in new ways. The current trends in rendering with topic modeling have only increased the value added by the technique. We now wish to briefly consider the topic modeling field in management research from a broader perspective, touching on important challenges and debates that will shape the direction of research and the evolution of the domain.

Challenges and Debates

Perhaps, the biggest challenge in the near future stems from how topic modeling has helped open the door to a plethora of work based on the quantitative structural study of meaning (Mohr, 1998; Ventresca & Mohr, 2002). Emergent classification systems based on meaning structures, such as those we have examined in topic modeling research, provide a reflexive contrast to others recognized and used to parse meaning in materialized structures, such as patent classification, risk typification, and industrial categorization. In this sense, we see management moving in a direction that reflects current trends in cultural sociology, political science, and linguistics;

a machine learning approach such as topic modeling can reveal shared cultural meanings that in turn can be integrated into the analytical process alongside traditional sociocultural variables and constructs. Our identified trends in topic modeling reveal that this integration is indeed occurring. Thus, topic modeling is not necessarily disrupting or displacing existing methods, so much as augmenting and extending them.

By highlighting the different modes of studying meaning (Mohr et al., 2013), we also acknowledge to the views of semiotics and qualitatively oriented scholars who have long recognized that meanings are grounded in practice and take on different levels of ambiguity. In the debates around semiotics and modeling, it is important to recognize that topic modeling combines the poetic (or connotative) with the semantic (or denotative) meanings of words in topics and subjects; although the words in "bags" are independent, they are combined in proximity and recognized in context. Integrating machine reading within studies of meaning necessitates a discussion around the tradeoffs of standardizing content and linking to theoretical artifacts. This also highlights that topic modeling practice in management is a deeply theoretical endeavor. Now that topic modeling algorithms are becoming more readily available through toolkits in R, Python, and other open source software, we worry that topic modeling risks being pigeon-holed as an LDA algorithm and "black boxed" as just another textual analysis technique. By attending to the rendering process, we hope we have helped scholars understand the choices inherent in the creation and preprocessing of corpora, the parameters used in the topic models themselves, and in the creation of theoretical artifacts from the analysis. Indeed, by articulating the rendering process, we have highlighted how topic modeling using machine learning algorithms actually foregrounds analysts' interpretive decisions and theory work.

Ultimately, theory is paramount for grounding claims around meaning. Our review has emphasized that incorporating topic modeling in a theoretical manner entails careful engagement with the cultural ecology of a social space. Our definition of the rendering process was created along these lines; particularly when using topic modeling to study the meanings of a social space, one cannot neglect its structural foundations. The ecology imagery evokes connotations of a structured space, contoured by theoretical concerns of social structure, such as boundaries, stratification, and reputations of actors. This also invokes the imagery by philosophers of science in assemblage theory, where a sociocultural ecology is constituted by relationships formed through processes of encoding meanings, such as stratification and territory (DeLanda, 2006).

The assemblage theory approach to conceptualizing knowledge-based fields is relevant to our consideration of the researcher generating knowledge alongside algorithms with machine learning. Such work is not performed by the human or the machine alone; rather, it is a combined effort. We reflect on how assemblage theory has illustrated the institution of science operating against the backdrop of two ideal styles of action—"nomadic" versus "state"—where the former is paradigm breaking and smooth, concerned with variation and problematization, and the latter is striated and contoured, concerned with precision and advances in structured fields of knowledge (Deleuze & Guattari, 1987; Jensen & Rödje, 2010). Machine learning approaches that are not configured with contextual structural knowledge may be nomadic—that is, overly fluid and rendering meaning structures across fields, only looking for what is statistically significant, but not necessarily socially or culturally significant. Understanding these ideal "nomadic" and "state" approaches to scientific endeavors can help us understand the ideal types of machine learning reading (nomadic: naive, fast, fluid, and distant) and human-only reading (state: careful, slow, narrowly focused, and deep). Our hope is that by delineating the rendering process, we are striking a middle ground between the two; in reflexively using machine learning tools in this manner, the analyst can see possibilities (latent meaning structures) against materialized social structures (formal classification systems).

To render meaning in this manner is to engender engagement with data, where the researcher zooms in and zooms out based on distant reading (Moretti, 2013) and representations of meaning structures. By conceptualizing topic modeling as part of a rendering process, we hope that we have also avoided the fear that social science researchers are just "squeezing [their] unstructured texts, sounds, or images into some special-purpose data model" (Underwood, 2015: 1). Instead, researchers use rendering processes for topic modeling as a "discovery strategy" to infer meaning. This blending of formal analytical methodologies with an interpretive focus helps reveal meanings and is echoed in an emerging stream of work in organizational theory that Ventresca and Mohr (2002) labeled "new archivalism."

Nevertheless, one challenge remains: as topic modeling has diffused into management research, the practices for applying it have not remained static. Indeed, by adapting this method, management scholars have contributed the rendering process itself. We see this contribution as being aligned with movements that draw on formal methods to generate representations of meanings, which can then be analyzed in a plethora of ways (Brieger et al., 2018; Davidson et al., 2019; Ventresca & Mohr, 2002). We found that many authors did indeed use computational modeling tools in a manner similar what Ventresca and Mohr described in 2002; however, we also found that the process of rendering goes further, particularly as it relates to rendering meanings. In our opinion, topic modeling tends to naturally ally more with mixed approaches to studying text (Brieger et al., 2018; Davidson et al., 2019; Ventresca & Mohr, 2002). Moreover, because meaning schema (i.e., dictionaries, and coding categories) is rejected *a priori*, the technique often seems to be more inductive in nature.

Of course, this is by no means the only mode of theorizing enabled through topic modeling. Other work has been more abductive in nature. For example, the frame analysis by Fligstein et al. (2017) helps explain how the Federal Open Market Committee underestimated the risks to the economy leading up to the 2008 financial crisis; their research design enabled them to use topic modeling to connect hypotheses to texts via a combination of qualitative and quantitative techniques. Indeed, topic modeling has also been used with partially deductive forms of theorizing [see Haans (2019) and Kaplan and Vakili (2015)].

As a final, cumulative point, we think that the flexibility of topic modeling—its utility in creating corpora, its ability to be paired with different quantitative and qualitative methods, and its applicability in variety of theoretical approaches—underpins its power and promise for management research. By surfacing topic modeling's flexibility, we hope our detailed exploration of the rendering process has persuaded the reader, at least to some extent, to consider engaging with topic modeling to build new management theory.

REFERENCES

Abbott, A. 1988. ***The system of professions: An essay on the division of expert labor*** (1st ed.). Chicago: University of Chicago Press.

Abrahamson, E., & Fairchild, G. 1999. Management fashion: Lifecycles, triggers, and collective learning processes. ***Administrative Science Quarterly***, 44(4): 708–740.

Aggarwal, V., Lee, M. K., & Hwang, E. 2017. Status gains and subsequent effects on evaluations. ***Academy of Management Proceedings***, 2017(1): 16355.

Ahonen, P. 2015. Institutionalizing big data methods in social and political research. ***Big Data & Society***, 2(2): 1–12.

Al Sumait, L., Barbará, D., Gentle, J., & Domeniconi, C. 2009. Topic significance ranking of LDA generative models. In En W. Buntine, M. Grobelnik, D. Mladenić, & J. Shawe-Taylor (Eds.), ***Machine learning and knowledge discovery in databases***, vol. 5781: 67–82. Berlin: Springer.

Allen, C., Luo, H., Murdock, J., Pu, J., Wang, X., Zhai, Y., & Zhao, K. 2017. ***Topic modeling the Han dian ancient classics***. ArXiv preprint arXiv:1702.00860. Retrieved from https://arxiv.org/abs/1702.00860. Accessed on October 1, 2018.

Almquist, Z. W., & Bagozzi, B. E. 2017. Using radical environmentalist texts to uncover network structure and network features. ***Sociological Methods & Research***: 1–56, https://doi.org/10.1177/0049124117729696.

Alvesson, M., & Kärreman, D. 2000a. Taking the linguistic turn in organizational research: Challenges, responses, consequences. ***Journal of Applied Behavioral Science***, 36(2): 136–158.

Alvesson, M., & Kärreman, D. 2000b. Varieties of discourse: On the study of organizations through discourse analysis. ***Human Relations***, 53(9): 1125–1149.

Antons, D., Joshi, A. M., & Salge, T. O. 2018. Content, contribution, and knowledge consumption: Uncovering hidden topic structure and rhetorical signals in scientific texts. ***Journal of Management***, https://doi.org/10.1177/0149206318774619.

Antons, D., Kleer, R., & Salge, T. O. 2016. Mapping the topic landscape of JPIM, 1984–2013: In search of hidden structures and development trajectories. ***Journal of Product Innovation Management***, 33(6): 726–749.

Argote, L., & Greve, H. R. 2007. "A behavioral theory of the firm": 40 years and counting: Introduction and impact. ***Organization Science***, 18(3): 337–349.

Arora, A. 1995. Licensing tacit knowledge: Intellectual property rights and the market for know-how. ***Economics of Innovation and New Technology***, 4(1): 41–60.

Arora, A., Gittelman, M., Kaplan, S., Lynch, J., Mitchell, W., & Siggelkow, N. 2016. Question-based innovations in strategy research methods. ***Strategic Management Journal***, 37(1): 3–9.

Augustine, G., & King, B. G. 2017. Behind the scenes: A backstage look at field formation within sustainability in higher education. ***Academy of Management Proceedings***, 2017(1): 15788.

Azzopardi, L., Girolami, M., & van Risjbergen, K. 2003. ***Investigating the relationship between language model perplexity and IR precision-recall measures***. Presented at the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Toronto: Association for Computing Machinery.

Bail, C. A. 2012. The fringe effect: Civil society organizations and the evolution of media discourse about Islam since the September 11th attacks. ***American Sociological Review***, 77(6): 855–879.

Bail, C. A. 2014. The cultural environment: Measuring culture with big data. ***Theory and Society***, 43(3–4): 465–482.

Bail, C. A., Brown, T. W., & Mann, M. 2017. Channeling hearts and minds: Advocacy organizations, cognitive-emotional currents, and public conversation. ***American Sociological Review***, 82(6): 1188–1213.

Bakhtin, M. M. 1982. ***The dialogic imagination: Four essays***. Austin, TX: University of Texas Press.

Ballinger, G. A., & Rockmann, K. W. 2010. Chutes versus ladders: Anchoring events and a punctuated-equilibrium perspective on social exchange relationships. ***Academy of Management Review***, 35(3): 373–391.

Bansal, P., & Corley, K. 2011. The coming of age for qualitative research: Embracing the diversity of qualitative methods. ***Academy of Management Journal***, 54(2): 233–237.

Bao, Y., & Datta, A. 2014. Simultaneously discovering and quantifying risk types from textual risk disclosures. ***Management Science***, 60(6): 1371–1391.

Baumer, E. P. S., Mimno, D., Guha, S., Quan, E., & Gay, G. K. 2017. Comparing grounded theory and topic modeling: Extreme divergence or unlikely convergence? ***Journal of the Association for Information Science and Technology***, 68(6): 1397–1410.

Benford, R. D., & Snow, D. A. 2000. Framing processes and social movements: An overview and assessment. ***Annual Review of Sociology***, 26: 611–639.

Berelson, B. 1952. ***Content analysis in communication research***. New York: Free Press.

Blanchard, S. J., Aloise, D., & Desarbo, W. S. 2017. Extracting summary piles from sorting task data. ***Journal of Marketing Research***, 54(3): 398–414.

Blei, D. M. 2012. Probabilistic topic models. ***Communications of the ACM***, 55(4): 77–84.

Blei, D. M., Griffiths, T. L., & Jordan, M. I. 2010. The nested Chinese restaurant process and Bayesian nonparametric inference of topic hierarchies. ***Journal of the ACM***, 57(2): 7:1–7:30.

Blei, D. M., Ng, A. Y., & Jordan, M. I. 2003. Latent Dirichlet allocation. ***Journal of Machine Learning Research***, 3: 993–1022.

Borgman, C. L. 2015. ***Big data, little data, no data: Scholarship in the networked world***. Cambridge, MA: MIT Press.

Braun, M. T., Kuljanin, G., & DeShon, R. P. 2018. Special considerations for the acquisition and wrangling of Big Data. ***Organizational Research Methods***, 21(3), 633–659, https://doi.org/10.1177/1094428117690235.

Breiger, R. L., Wagner-Pacifici, R., & Mohr, J. W. 2018. Capturing distinctions while mining text data: Toward low-tech formalization for text analysis. ***Poetics***, 68: 104–119.

Burke, K., 1945. ***A grammar of motives***. Berkeley, CA: University of California Press.

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

Buurma, R. S. 2015. The fictionality of topic modeling: Machine reading Anthony Trollope's Barsetshire series. ***Big Data & Society***, 2(2): 1–6, https://doi.org/10.1177/2053951715610591.

Campbell, J. L., Chen, H., Dhaliwal, D. S., Lu, H., & Steele, L. B. 2014. The information content of mandatory risk factor disclosures in corporate filings. ***Review of Accounting Studies***, 19(1): 396–455.

Cao, L., & Fei-Fei, L. 2007. ***Spatially coherent latent topic model for concurrent segmentation and classification of objects and scenes***. 2007 IEEE 11th International Conference on Computer Vision, ICCV: 1–8. Piscataway, NJ: IEEE.

Carley, K. M., & Kaufer, D. 1993. Semantic connectivity: An approach for analyzing semantic networks. ***Communication Theory***, 3(3): 183–213.

Chang, J., & Blei, D. 2009. Relational topic models for document networks. In ***Artificial intelligence and statistics***: 81–88. Retrieved from http://proceedings.mlr.press/v5/chang09a.html.

Chang, J., Boyd-Graber, J., Gerrish, S., Wang, C., & Blei, D. M. 2009. Reading tea leaves: How humans interpret topic models. In Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, & A. Culotta (Eds.), ***Advances in neutral information processing systems***, vol. 22. Retrieved from https://papers.nips.cc/paper/3700-reading-tea-leaves-how-humans-interpret-topic-models.pdf.

Charmaz, K. 2014. ***Constructing grounded theory*** (2nd ed.). Thousand Oaks, CA: Sage.

Cho, Y.-J., Fu, P.-W., & Wu, C.-C. 2017. Popular research topics in marketing journals, 1995–2014. ***Journal of Interactive Marketing***, 40: 52–72.

Chomsky, N. 1956. Three models for the description of language. ***IRE Transactions on Information Theory***, 2(3): 113–124.

Cornelissen, J. P., Durand, R., Fiss, P. C., Lammers, J. C., & Vaara, E. 2015. Putting communication front and center in institutional theory and analysis. ***Academy of Management Review***, 40(1): 10–27.

Croidieu, G., & Kim, P. H. 2018. Labor of love: Amateurs and lay-expertise legitimation in the early U.S. radio field. ***Administrative Science Quarterly***, 63(1): 1–42.

Crossley, S. A., Dascalu, M., & McNamarac, D. S. 2017. ***How important is size? An investigation of corpus size and meaning in both latent semantic analysis and latent Dirichlet allocation***. 30th International Florida Artificial Intelligence Research Society Conference, FLAIRS 2017. Menlo Park, CA: AAAI Press.

Dalpiaz, E., Rindova, V., & Ravasi, D. 2016. Combining logics to transform organizational agency: Blending industry and art at Alessi. ***Administrative Science Quarterly***, 61(3): 347–392.

Davidson, E., Edwards, R., Jamieson, L., & Weller, S. 2019. Big data, qualitative style: A breadth-and-depth method for working with large amounts of secondary qualitative data. ***Quality & Quantity***, 53(1): 363–376.

Davis, G. F. 2016. Can an economy survive without corporations? Technology and robust organizational alternatives. ***Academy of Management Perspectives***, 30(2): 129–140.

de Saussure, F. 1959. ***Course in general linguistics***. New York: McGraw-Hill.

Deephouse, D. L. 1999. To be different, or to be the same? It's a question (and theory) of strategic balance. ***Strategic Management Journal***, 20(2): 147–166.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. 1990. Indexing by latent semantic analysis. ***Journal of the American Society for Information Science***, 41(6): 391.

DeLanda, M. 2006. ***A new philosophy of society: Assemblage theory and social complexity***. London: A&C Black.

Deleuze, G., & Guattari, F. 1987. ***A thousand plateaus***. New York: Continuum.

Denzin, N. K., & Lincoln, Y. S. 2011. ***The SAGE handbook of qualitative research***. Thousand Oaks, CA: Sage.

Dey, I. 1995. Reducing fragmentation in qualitative research. In U. Kelle (Ed.), ***Computer-aided qualitative data analysis: Theory, methods, and practice***: 69–79. London: Sage.

Diesner, J., & Carley, K. M. 2005. Revealing social structure from texts: Meta-matrix text analysis as a novel method for network text analysis. In V. K. Narayanan & D. J. Armstrong (Eds.), ***Causal mapping for research in information technology***: 81–108. Hershey, PA: IGI Global.

DiMaggio, P. J. 1997. Culture and cognition. ***Annual Review of Sociology***, 23: 263–287.

DiMaggio, P. J. 2015. Adapting computational text analysis to social science (and vice versa). ***Big Data & Society***, 2(2): 2053951715602908.

DiMaggio, P. J., Nag, M., & Blei, D. 2013. Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. ***Poetics***, 41(6): 570–606.

Durand, R., & Khaire, M. 2017. Where do market categories come from and how? Distinguishing category creation from category emergence. ***Journal of Management***, 43(1): 87–110.

Durand, R., & Paolella, L. 2013. Category stretching: Reorienting research on categories in strategy, entrepreneurship, and organization theory. ***Journal of Management Studies***, 50(6): 1100–1123.

Duriau, V. J., Reger, R. K., & Pfarrer, M. D. 2007. A content analysis of the content analysis literature in organization studies: Research themes, data sources, and methodological refinements. ***Organizational Research Methods***, 10(1): 5–34.

Eisenhardt, K. M. 1989. Building theories from case study research. ***Academy of Management Review***, 14(4): 532–550.

Evans, J. A., & Aceves, P. 2016. Machine translation: Mining text for social theory. ***Annual Review of Sociology***, 42(1): 21–50.

Fama, E. F., & French, K. R. 1993. Common risk factors in the returns on stocks and bonds. ***Journal of Financial Economics***, 33(1): 3–56.

Firth, J. R. 1957. ***Papers in linguistics***. London: Oxford University Press.

Fiss, P. C. 2007. A set-theoretic approach to organizational configurations. ***Academy of Management Review***, 32(4): 1180–1198.

Fiss, P. C., & Hirsch, P. M. 2005. The discourse of globalization: Framing and sensemaking of an emerging concept. ***American Sociological Review***, 70(1): 29–52.

Fleming, L. 2001. Recombinant uncertainty in technological search. ***Management Science***, 47(1): 117–132.

Fligstein, N., & McAdam, D. 2011. Toward a general theory of strategic action fields. ***Sociological Theory***, 29(1): 1–26.

Fligstein, N., Stuart Brundage, J., & Schultz, M. 2017. Seeing like the Fed: Culture, cognition, and framing in the failure to anticipate the financial crisis of 2008. ***American Sociological Review***, 82(5): 879–909.

Forbes, D. P., & Kirsch, D. A. 2011. The study of emerging industries: Recognizing and responding to some central problems. ***Journal of Business Venturing***, 26: 589–602.

Gehman, J., Glaser, V. L., Eisenhardt, K. M., Gioia, D., Langley, A., & Corley, K. G. 2018. Finding theory-method fit: A comparison of three qualitative approaches to theory building. ***Journal of Management Inquiry***, 27(3): 284–300.

Gehman, J., & Soublière, J.-F. 2017. Cultural entrepreneurship: From making culture to cultural making. ***Innovation***, 19(1): 61–73.

Gerlach, M., Peixoto, T. P., & Altmann, E. G. 2018. A network approach to topic models. ***Science Advances***, 4(7): eaaq1360.

Gioia, D. A., Corley, K. G., & Hamilton, A. L. 2013. Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. ***Organizational Research Methods***, 16(1): 15–31.

Gioia, D. A., & Thomas, J. B. 1996. Identity, image, and issue interpretation: Sensemaking during strategic change in academia. ***Administrative Science Quarterly***, 41(3): 370–403.

Giorgi, S., & Weber, K. 2015. Marks of distinction: Framing and audience appreciation in the context of investment advice. ***Administrative Science Quarterly***, 60(2): 333–367.

Goffman, E. 1974. ***Frame analysis: An essay on the organization of experience***. Cambridge, MA: Harvard University Press.

Greve, H. R. 2003. ***Organizational learning from performance feedback: A behavioral perspective on innovation and change***. Cambridge, UK: Cambridge University Press.

Griffiths, T. L., & Steyvers, M. 2004. Finding scientific topics. ***Proceedings of the National Academy of Sciences***, 101(suppl 1): 5228–5235, https://doi.org/10.1073/pnas.0307752101.

Grimmer, J., & Stewart, B. M. 2013. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. ***Political Analysis***, 21(3): 267–297.

Guerreiro, J., Rita, P., & Trigueiros, D. 2016. A text mining-based review of cause-related marketing literature. ***Journal of Business Ethics***, 139(1): 111–128.

Guo, L., Sharma, R., Yin, L., Lu, R., & Rong, K. 2017. Automated competitor analysis using big data analytics: Evidence from the fitness mobile app business. ***Business Process Management Journal***, 23(3): 735–762.

Haans, R. F. J. 2019. What's the value of being different when everyone is? The effects of distinctiveness on performance in homogeneous versus heterogeneous categories. ***Strategic Management Journal***, 40(1): 3–27.

Hankammer, S., Antons, D., Kleer, R., & Piller, F. 2016. Researching mass customization: Mapping hidden structures and development trajectories. ***Academy of Management Proceedings***, 2016(1): 10900.

Hannan, M. T., & Carroll, G. R. 1992. ***Dynamics of organizational populations: Density, legitimation, and competition***. Oxford: Oxford University Press.

Hannan, M. T., Pólos, L., & Carroll, G. R. 2007. ***Logics of organization theory: Audiences, codes, and ecologies***. Princeton, NJ: Princeton University Press.

Hannigan, T. R., & Casasnovas, G. 2019. New structuralism and field emergence: The co-constitution of meanings and actors in the early moments of impact investing. In ***Research in the sociology of organizations (forthcoming)***.

Hannigan, T. R., Porac, J. F., Bundy, J., Wade, J. B., & Graffin, S. D. 2019. ***Crossing the line or creating the line: Media effects in the 2009 British MP expense scandal***. Working paper.

Hatch, M. J. 1993. The dynamics of organizational culture. ***Academy of Management Review***, 18(4): 657–693.

Hatch, M. J., & Schultz, M. 2017. Toward a theory of using history authentically: Historicizing in the Carlsberg Group. ***Administrative Science Quarterly***, 62(4): 657–697.

Heugens, P. P., & Lander, M. W. 2009. Structure! Agency! (and other quarrels): A meta-analysis of institutional theories of organization. ***Academy of Management Journal***, 52(1): 61–85.

Houghton, J. P., Siegel, M., Madnick, S., Tounaka, N., Nakamura, K., Sugiyama, T., Nakagawa, D., & Shirnen, B. 2017. Beyond keywords: Tracking the evolution of conversational clusters in social media. ***Sociological Methods & Research***, https://doi.org/10.1177/0049124117729705.

Hu, D. J., & Saul, L. K. 2009. A probabilistic topic model for music analysis. ***Proceedings of NIPS***, 9: 1–4.

Huang, A. H., Lehavy, R., Zang, A. Y., & Zheng, R. 2017. Analyst information discovery and interpretation roles: A topic modeling approach. ***Management Science***, 64(6): 2833–2855.

Huang, X., Li, X., Zhang, L., Liu, T., Chiu, D., & Zhu, T. 2015. ***Topic model for identifying suicidal ideation in Chinese microblog***. 29th Pacific Asia Conference on Language, Information and Computation: 553–562.

Humphreys, A., & Wang, R. J.-H. 2018. Automated text analysis for consumer research. ***Journal of Consumer Research***, 44(6): 1274–1306.

Jacobs, B. J. D., Donkers, B., & Fok, D. 2016. Model-based purchase predictions for large assortments. ***Marketing Science***, 35(3): 389–404.

Jensen, C. B., & Rödje, K. 2010. ***Deleuzian intersections: Science, technology, anthropology***. New York: Berghahn Books.

Jha, H. K., & Beckman, C. M. 2017. A patchwork of identities: Emergence of charter schools as a new organizational form. ***Research in the Sociology of Organizations***, 50: 69–107.

Jin, M., Luo, X., Zhu, H., & Zhuo, H. H. 2018. ***Combining deep learning and topic modeling for review understanding in context-aware recommendation***. Proceedings of NAACL-HLT 2018: 1605–1614.

Jockers, M. L., & Mimno, D. 2013. Significant themes in 19th-century literature. ***Poetics***, 41(6): 750–769.

Kaminski, J., Jiang, Y., Piller, F., & Hopp, C. 2017. ***Do user entrepreneurs speak different? Applying natural language processing to crowdfunding videos***. Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems: 2683–2689. New York: ACM.

Kaplan, S. 2008a. Cognition, capabilities, and incentives: Assessing firm response to the fiber-optic revolution. ***Academy of Management Journal***, 51(4): 672–695.

Kaplan, S. 2008b. Framing contests: Strategy making under uncertainty. ***Organization Science***, 19(5): 729–752.

Kaplan, S., & Vakili, K. 2015. The double-edged word of recombination in breakthrough innovation. ***Strategic Management Journal***, 36(10): 1435–1457.

Karanovic, J., Berends, H., & Engel, Y. 2018. Is platform capitalism legit? Ask the workers. ***Academy of Management Proceedings***, 2018(1): 17038.

Kennedy, M. T. 2005. Behind the one-way mirror: Refraction in the construction of product market categories. ***Poetics***, 33(3–4): 201–226.

Kennedy, M. T. 2008. Getting counted: Markets, media, and reality. ***American Sociological Review***, 73(2): 270–295.

Kennedy, M. T., & Fiss, P. C. 2013. An ontological turn in categories research: From standards of legitimacy to evidence of actuality. ***Journal of Management Studies***, 50(6): 1138–1154.

Kim, H., Ahn, S.-J., & Jung, W.-S. 2018. Horizon scanning in policy research database with a probabilistic topic model. ***Technological Forecasting and Social Change***, https://doi.org/10.1016/j.techfore.2018.02.007.

Kim, S., & Bae, J. 2016. Cross-cultural differences in concrete and abstract corporate social responsibility (CSR) campaigns: Perceived message clarity and perceived CSR as mediators. ***International Journal of Corporate Social Responsibility***, 1: 1–14.

Kiss, T., & Strunk, J. 2006. Unsupervised multilingual sentence boundary detection. ***Computational Linguistics***, 32(4): 485–525.

Kitchin, R., & McArdle, G. 2016. What makes big data, big data? Exploring the ontological characteristics of 26 datasets. ***Big Data & Society***, 3(1): 1–10.

Kline, S. J., & Rosenberg, N. 1986. An overview of innovation. In R. Landau & N. Rosenberg (Eds.), ***The positive sum strategy: Harnessing technology for economic growth***: 275–306. Washington, DC: National Academy of Sciences.

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B. E., Bussonnier, M., Frederic, J., & Corlay, S. 2016. Jupyter Notebooks—A publishing format for reproducible computational workflows. In F. Loizides & B. Schmidt (Eds.), ***Positioning and power in academic publishing: Players, agents and agendas***: 87–90. Clifton, VA: IOS Press.

Kobayashi, V. B., Mol, S. T., Berkers, H. A., Kismihók, G., & Den Hartog, D. N. 2018. Text classification for organizational researchers: A tutorial. ***Organizational Research Methods***, 21(3): 766–799.

Krippendorff, K. 1980. ***Content analysis*** (1st ed.). Beverly Hills, CA: Sage.

Krippendorff, K. 2004. ***Content analysis: An introduction to its methodology*** (2nd ed.). Thousand Oaks, CA: Sage.

Krippendorff, K. 2012. ***Content analysis: An introduction to its methodology*** (3rd ed.). Thousand Oaks, CA: Sage.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. 2012. ***ImageNet classification with deep convolutional neural networks***. Proceedings of the 25th International Conference on Neural Information Processing Systems, vol. 1: 1097–1105. Red Hook, NY: Curran Associates Inc.

Kuhn, T. S. 1996. ***The structure of scientific revolutions*** (3rd ed.). Chicago: University of Chicago Press.

Langley, A. 1999. Strategies for theorizing from process data. ***Academy of Management Review***, 24(4): 691–710.

Langley, A., & Abdallah, C. 2011. Templates and turns in qualitative studies of strategy and management. In D. D. Bergh & D. J. Ketchen (Eds.), ***Building methodological bridges***: 201–235. Bingley, UK: Emerald Group Publishing.

Lasswell, H. D. 1948. ***Power and personality***. New York: W.W. Norton.

Lawrence, T. B., Suddaby, R., & Leca, B. 2009. ***Institutional work*** (1st ed.). Cambridge, UK: Cambridge University Press.

Lee, G. M., Qiu, L., & Whinston, A. B. 2016. A friend like me: Modeling network formation in a location-based social network. ***Journal of Management Information Systems***, 33(4): 1008–1033.

Lee, H., Kwak, J., Song, M., & Kim, C. O. 2015. Coherence analysis of research and education using topic modeling. ***Scientometrics***, 102(2): 1119–1137.

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

Levy, K. E. C., & Franklin, M. 2014. Driving regulation: Using topic models to examine political contention in the U.S. trucking industry. ***Social Science Computer Review***, 32(2): 182–194.

Lim, A., & Tsutsui, K. 2012. Globalization and commitment in corporate social responsibility: Cross-national analyses of institutional and political economy effects. ***American Sociological Review***, 77(1): 69–98.

Liu, Y., Mai, F., & MacDonald, C. 2018. A big-data approach to understanding the thematic landscape of the field of business ethics, 1982–2016. ***Journal of Business Ethics***, https://doi.org/10.1007/s10551-018-3806-5.

Locke, K. D. 2001. ***Grounded theory in management research***. London: Sage.

Loewenstein, J., Ocasio, W., & Jones, C. 2012. Vocabularies and vocabulary structure: A new approach linking categories, practices, and institutions. ***Academy of Management Annals***, 6(1): 41–86.

Lounsbury, M., & Glynn, M. A. 2001. Cultural entrepreneurship: Stories, legitimacy, and the acquisition of resources. ***Strategic Management Journal***, 22(6/7): 545–564.

Lounsbury, M., & Glynn, M. A. 2019. ***Cultural entrepreneurship: A new agenda for the study of entrepreneurial processes and possibilities***. Cambridge, UK: Cambridge University Press.

Lounsbury, M., & Ventresca, M. 2003. The new structuralism in organizational theory. ***Organization***, 10(3): 457–480.

Manning, C., Raghavan, P., & Schütze, H. 2010. Introduction to information retrieval. ***Natural Language Engineering***, 16, 100–103.

Manning, C. D., & Schütze, H. 1999. ***Foundations of statistical natural language processing*** (1st ed.). Cambridge, MA: MIT Press.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., & McClosky, D. 2014. ***The Stanford CoreNLP Natural Language Processing Toolkit***. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations: 55–60. Available at https://www.aclweb.org/anthology/P14-5010. Accessed July 15, 2018.

Marciniak, D. 2016. Computational text analysis: Thoughts on the contingencies of an evolving method. ***Big Data & Society***, 3(2): 1–5.

Marquis, C., Glynn, M. A., & Davis, G. F. 2007. Community isomorphism and corporate social action. ***Academy of Management Review***, 32(3): 925–945.

Marshall, E. A. 2013. Defining population problems: Using topic models for cross-national comparison of disciplinary development. ***Poetics***, 41(6): 701–724.

Martens, M. L., Jennings, J. E., & Jennings, P. D. 2007. Do the stories they tell get them the money they need? The role of entrepreneurial narratives in resource acquisition. ***Academy of Management Journal***, 50(5): 1107–1132.

Mattmann, C. A. 2013. Computing: A vision for data science. ***Nature***, 493(7433): 473.

McCallum, A. K. 2002. ***MALLET: A machine learning for language4 toolkit***. Retrieved from http://mallet.cs.umass.edu. Accessed July 15, 2018.

McFarland, D. A., Ramage, D., Chuang, J., Heer, J., Manning, C. D., & Jurafsky, D. 2013. Differentiating language usage through topic models. ***Poetics***, 41(6): 607–625.

Meyer, R. E., Jancsary, D., Höllerer, M. A., & Boxenbaum, E. 2017. The role of verbal and visual text in the process of institutionalization. ***Academy of Management Review***, 43(3): 392–418.

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. 1990. Introduction to WordNet: An on-line lexical database. ***International Journal of Lexicography***, 3(4): 235–234.

Miller, I. M. 2013. Rebellion, crime and violence in Qing China, 1722–1911: A topic modeling approach. ***Poetics***, 41(6): 626–649.

Mimno, D. 2012. Computational historiography: Data mining in a century of classics journals. ***Journal on Computing and Cultural Heritage***, 5(1): 3:1–3:19.

Mimno, D., Wallach, H. M., Talley, E., Leenders, M., & McCallum, A. 2011. ***Optimizing semantic coherence in topic models***. Proceedings of the Conference on Empirical Methods in Natural Language Processing: 262–272. Stroudsburg, PA: Association for Computational Linguistics.

Mohr, J. W. 1998. Measuring meaning structures. ***Annual Review of Sociology***, 24: 345–370.

Mohr, J. W., & Bogdanov, P. 2013. Introduction—Topic models: What they are and why they matter. ***Poetics***, 41(6): 545–569.

Mohr, J. W., & Duquenne, V. 1997. The duality of culture and practice: Poverty relief in New York city, 1888–1917. ***Theory and Society***, 26(2/3): 305–356.

Mohr, J. W., Wagner-Pacifici, R., Breiger, R. L., & Bogdanov, P. 2013. Graphing the grammar of motives in national security strategies: Cultural interpretation, automated text analysis and the drama of global politics. ***Poetics***, 41(6): 670–700.

Mollick, E. 2014. The dynamics of crowdfunding: An exploratory study. ***Journal of Business Venturing***, 29(1): 1–16.

Momeni, A., & Rost, K. 2016. Identification and monitoring of possible disruptive technologies by patent-development paths and topic modeling. ***Technological Forecasting and Social Change***, 104: 16–29.

Moretti, F. 2013. ***Distant reading***. New York: Verso.

Morgeson, F. P., Mitchell, T. R., & Liu, D. 2015. Event system theory: An event-oriented approach to the organizational sciences. ***Academy of Management Review***, 40(4): 515–537.

Mützel, S. 2015. Facing big data: Making sociology relevant. ***Big Data & Society***, 2(2): 2053951715599179.

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

Navis, C., & Glynn, M. A. 2010. How new market categories emerge: Temporal dynamics of legitimacy, identity, and entrepreneurship in satellite radio, 1990–2005. ***Administrative Science Quarterly***, 55(3): 439–471.

Navis, C., & Glynn, M. A. 2011. Legitimate distinctiveness and the entrepreneurial identity: Influence on investor judgments of new venture plausibility. ***Academy of Management Review***, 36(3): 477–499.

Nelsen, B. J., & Barley, S. R. 1997. For love or money? Commodification and the construction of an occupational mandate. ***Administrative Science Quarterly***, 42(4): 619–653.

Nelson, L. K. 2017. Computational grounded theory: A methodological framework. ***Sociological Methods & Research***: 0049124117729703.

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.

Ocasio, W. 1997. Towards an attention-based view of the firm. ***Strategic Management Journal***, 18: 187–206.

Oh, J., Stewart, A. E., & Phelps, R. E. 2017. Topics in the *Journal of Counseling Psychology*, 1963–2015. ***Journal of Counseling Psychology***, 64(6): 604–615.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. 2015. ***The development and psychometric properties of LIWC2015***. Austin, TX: University of Texas at Austin.

Pentland, B. T., & Feldman, M. S. 2005. Organizational routines as a unit of analysis. ***Industrial and Corporate Change***, 14(5): 793–815.

Pfarrer, M., Pollock, T., & Rindova, V. 2010. A tale of two assets: The effects of firm reputation and celebrity on earnings surprises and investors' reactions. ***Academy of Management Journal***, 53(5): 1131–1152.

Podolny, J. M. 1993. A status-based model of market competition. ***American Journal of Sociology***, 98(4): 829–872.

Pollach, I. 2012. Taming textual data: The contribution of corpus linguistics to computer-aided text analysis. ***Organizational Research Methods***, 15(2): 263–287.

Porac, J. F., Wade, J. B., & Pollock, T. G. 1999. Industry categories and the politics of the comparable firm in CEO compensation. ***Administrative Science Quarterly***, 44(1): 112–144.

Pratt, M. G. 2009. From the editors: For the lack of a boilerplate: Tips on writing up (and reviewing) qualitative research. ***Academy of Management Journal***, 52(5): 856–862.

Prein, G., & Kelle, U. 1995. Using linkages and networks for theory building. In U. Kelle (Ed.), ***Computer qualitative data analysis: Theory, methods, and practice***: 62–68. London: Sage Publications.

Pröllochs, N., & Feuerriegel, S. 2018. Business analytics for strategic management: Identifying and assessing corporate challenges via topic modeling. ***Information & Management***, https://doi.org/10.1016/j.im.2018.05.003.

Puranam, D., Narayan, V., & Kadiyali, V. 2017. The effect of calorie posting regulation on consumer opinion: A flexible latent Dirichlet allocation model with informative priors. ***Marketing Science***, 36(5): 726–746.

Raffaelli, R. 2018. Technology reemergence: Creating new value for old technologies in Swiss mechanical watchmaking, 1970–2008. ***Administrative Science Quarterly***: 0001839218778505, https://doi.org/10.1177/0001839218778505.

Ragin, C. C. 2008. ***Redesigning social inquiry: Fuzzy sets and beyond***. Chicago: University of Chicago Press.

Rao, H., Monin, P., & Durand, R. 2003. Institutional change in Toque Ville: Nouvelle cuisine as an identity movement in French gastronomy. ***American Journal of Sociology***, 108(4): 795–843.

Rhee, E. Y., & Fiss, P. C. 2014. Framing controversial actions: Regulatory focus, source credibility, and stock market reaction to poison pill adoption. ***Academy of Management Journal***, 57(6): 1734–1758.

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. 2014. Structural topic models for open-ended survey responses. ***American Journal of Political Science***, 58(4): 1064–1082.

Rosen-Zvi, M., Griffiths, T., Steyvers, M., & Smyth, P. 2004. ***The author-topic model for authors and documents***. Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence: 487–494. Arlington, VA: AUAI Press.

Rothenberg, A. 2014. ***Flight from wonder—An investigation of scientific creativity***. New York: Oxford University Press.

Ruckman, K., & McCarthy, I. 2017. Why do some patents get licensed while others do not? ***Industrial and Corporate Change***, 26(4): 667–688.

Schmiedel, T., Müller, O., & vom Brocke, J. 2018. Topic modeling as a strategy of inquiry in organizational research. ***Organizational Research Methods***, 3(1), https://doi.org/10.1177/1094428118773858.

Shi, Z., Lee, G. M., & Whinston, A. B. 2016. Toward a better measure of business proximity: Topic modeling for industry intelligence. ***MIS Quarterly***, 40(4): 1035–1056.

Sievert, C., & Shirley, K. 2014. ***LDAvis: A method for visualizing and interpreting topics***. Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces: 63–70. Stroudsburg, PA: Association for Computational Linguistics.

Slingerland, E., Nichols, R., Neilbo, K., & Logan, C. 2017. The distant reading of religious texts: A "big data" approach to mind-body concepts in early China. ***Journal of the American Academy of Religion***, 85(4): 985–1016.

Smith, A., Hawes, T., & Myers, M. 2014. ***Hiéarchie: Visualization for hierarchical topic models***. Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces: 71–78. Baltimore, MD: Association for Computational Linguistics.

Snow, D. A., & Benford, R. D. 1988. Ideology, frame resonance, and participant mobilization. In B. Klandermans, H. Kriesi, & S. G. Tarrow (Eds.), ***International social movement research***, vol. 1: 197–218. Greenwich, CT: JAI Press.

Snow, D. A., Rochford, E. B., Worden, S. K., & Benford, R. D. 1986. Frame alignment processes, micromobilization, and movement participation. ***American Sociological Review***, 51(4): 464–481.

Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A., & Potts, C. 2013. ***Recursive deep models for semantic compositionality over a sentiment treebank***. Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing: 1631–1642. Retrieved from https://nlp.stanford.edu/~socherr/EMNLP2013_RNTN.pdf. Accessed July 15, 2018.

Song, M., Heo, G. E., & Lee, D. 2015. Identifying the landscape of Alzheimer's disease research with network and content analysis. ***Scientometrics***, 102(1): 905–927.

Song, M., & Kim, S. Y. 2013. Detecting the knowledge structure of bioinformatics by mining full-text collections. ***Scientometrics***, 96(1): 183–201.

Sørensen, J. B., & Stuart, T. E. 2000. Aging, obsolescence, and organizational innovation. ***Administrative Science Quarterly***, 45(1): 81–112.

Stevens, K., Kegelmeyer, P., Andrzejewski, D., & Buttler, D. 2012. ***Exploring topic coherence over many models and many topics***. Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning: 952–961.

Strauss, A. C., & Corbin, J. M. 1998. ***Basics of qualitative research: Techniques and procedures for developing grounded theory*** (2nd ed.). Los Angeles, CA: Sage.

Strothotte, T., & Schlechtweg, S. 2002. ***Non-photorealistic computer graphics: Modeling, rendering, and animation***. San Francisco, CA: Morgan Kaufmann.

Suominen, A., Toivanen, H., & Seppänen, M. 2017. Firms' knowledge profiles: Mapping patent data with unsupervised learning. ***Technological Forecasting and Social Change***, 115: 131–142.

Tangherlini, T. R., & Leonard, P. 2013. Trawling in the sea of the great unread: Sub-corpus topic modeling and humanities research. ***Poetics***, 41(6): 725–749.

Tchalian, H. 2019. Microfoundations and recursive analysis: A mixed-methods framework for language-based research, computational methods, and theory development. In ***Research in the Sociology of Organizations***. Forthcoming.

Tchalian, H., Glaser, V. L., Hannigan, T. R., & Lounsbury, M. 2019. ***Institutional attention: Cultural entrepreneurship and the dynamics of category construction***. Working paper.

Thornton, P. H., & Ocasio, W. 1999. Institutional logics and the historical contingency of power in organizations: Executive succession in the higher education publishing industry, 1958–1990. ***American Journal of Sociology***, 105(3): 801–843.

Thornton, P. H., Ocasio, W., & Lounsbury, M. 2012. ***The institutional logics perspective: A new approach to culture, structure and process***. New York: Oxford University Press.

Timmermans, S., & Tavory, I. 2012. Theory construction in qualitative research: From grounded theory to abductive analysis. ***Sociological Theory***, 30(3): 167–186.

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.

Tolbert, P. S., & Zucker, L. G. 1996. The institutionalization of institutional theory. In S. Clegg, C. Hardy, & W. Nord (Eds.), ***Handbook of organization studies***: 175–190. London: Sage.

Tonidandel, S., King, E. B., & Cortina, J. M. 2018. Big data methods: Leveraging modern data analytic techniques to build organizational science. ***Organizational Research Methods***, 21(3): 525–547.

Toubia, O., & Netzer, O. 2016. Idea generation, creativity, and prototypicality. ***Marketing Science***, 36(1): 1–20.

Trajtenberg, M. 1990. A penny for your quotes: Patent citations and the value of innovations. ***RAND Journal of Economics***, 21(1): 172–187.

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.

Turney, P. D., & Pantel, P. 2010. From frequency to meaning: Vector space models of semantics. ***Journal of Artificial Intelligence Research***, 37: 141–188.

Underwood, T. 2015. The literary uses of high-dimensional space. ***Big Data & Society***, 2(2): 1–6.

Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. 2013. Atypical combinations and scientific impact. ***Science***, 342(6157): 468–472.

Vaara, E. 2010. Taking the linguistic turn seriously: Strategy as a multifaceted and interdiscursive phenomenon. In J. Baum & J. B. Lampel (Eds.), ***The globalization of strategy research***: 29–50. Bingley, UK: Emerald Group Publishing.

Vaara, E., Aranda, A., Etchanchu, H., Guyt, J., & Sele, K. 2019. ***How to make use of structural topic modeling in critical discourse analysis?*** Working paper.

Ventresca, M. J., & Mohr, J. W. 2002. Archival research methods. In J. A. C. Baum (Ed.), ***Blackwell companion to organizations***: 805–828. Oxford: Blackwell.

Venugopalan, S., & Rai, V. 2015. Topic based classification and pattern identification in patents. ***Technological Forecasting and Social Change***, 94: 236–250.

Vergne, J.-P., & Wry, T. 2014. Categorizing categorization research: Review, integration, and future directions. ***Journal of Management Studies***, 51(1): 56–94.

Wagner-Pacifici, R., Mohr, J. W., & Breiger, R. L. 2015. Ontologies, methodologies, and new uses of big data in the social and cultural sciences. ***Big Data & Society***, 2(2): 2053951715613810.

Wang, X., Bendle, N. T., Mai, F., & Cotte, J. 2015. The journal of consumer research at 40: A historical analysis. ***Journal of Consumer Research***, 42(1): 5–18.

Wang, X., McCallum, A., & Wei, X. 2007. ***Topical n-grams: Phrase and topic discovery, with an application to information retrieval***. Proceedings of the Seventh IEEE International Conference on Data Mining (ICDM 2007): 697–702. Piscataway, NJ: IEEE.

Wang, Y., & Chaudhry, A. 2018. When and how managers' responses to online reviews affect subsequent reviews. ***Journal of Marketing Research***, 55(2): 163–177.

Weber, K., & Dacin, M. T. 2011. The cultural construction of organizational life: Introduction to the special issue. ***Organization Science***, 22(2): 287–298.

Weber, K., Patel, H., & Heinze, K. L. 2013. From cultural repertoires to institutional logics: A content-analytic method. In M. Lounsbury & E. Boxenbaum (Eds.), ***Institutional logics in action***, vol. 39B: 351–382. Bingley, UK: Emerald Group Publishing.

Weber, R. P. 1990. ***Basic content analysis*** (2nd ed.). London: Sage.

Whetten, D. A. 1989. What constitutes a theoretical contribution? ***Academy of Management Review***, 14(4): 490–495.

Whorf, B. L. 1956. ***Language, thought, and reality: Selected writings of Benjamin Lee Whorf***. Cambridge, MA: Technology Press of Massachusetts Institute of Technology.

Wilson, A. J., & Joseph, J. 2015. ***Organizational attention and technological search in the multibusiness firm: Motorola from 1974 to 1997***. Bingley, UK: Emerald Group Publishing.

Yau, C.-K., Porter, A., Newman, N., & Suominen, A. 2014. Clustering scientific documents with topic modeling. ***Scientometrics***, 100(3): 767–786.

Zajac, E. J., & Fiss, P. C. 2006. The symbolic management of strategic change: Sensegiving via framing and decoupling. ***Academy of Management Journal***, 49(6): 1173–1193.

Zajac, E. J., & Westphal, J. D. 1994. The costs and benefits of managerial incentives and monitoring in large U.S. corporations: When is more not better? ***Strategic Management Journal***, 15(S1): 121–142.

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.

Zhao, E. Y., Fisher, G., Lounsbury, M., & Miller, D. 2017. Optimal distinctiveness: Broadening the interface between institutional theory and strategic management. ***Strategic Management Journal***, 38(1): 93–113.

Zhao, E. Y., Ishihara, M., Jennings, P. D., & Lounsbury, M. 2018. Optimal distinctiveness in the console video game industry: An exemplar-based model of proto-category evolution. ***Organization Science***, 29(4): 588–611.

Zott, C., & Huy, Q. N. 2007. How entrepreneurs use symbolic management to acquire resources. ***Administrative Science Quarterly***, 52(1): 70–105.

Zuckerman, E. W. 1999. The categorical imperative: Securities analysts and the illegitimacy discount. ***American Journal of Sociology***, 104(5): 1398–1438.

Timothy R. Hannigan (tim.hannigan@ualberta.ca) is an assistant professor of Organization Theory and Entrepreneurship at the University of Alberta. He completed doctoral work at the Said Business School, University of Oxford. His research interests cover dynamics in organizational

discontinuities—including the contexts of innovation, entrepreneurship, and political scandals.

Richard F. J. Haans (haans@rsm.nl) is an Assistant Professor of Strategic Management & Entrepreneurship at the Rotterdam School of Management, Erasmus University Rotterdam. He studies optimal distinctiveness—the question of how different firms should strive to be from their competitors—and creativity more generally.

Keyvan Vakili (kvakili@london.edu) is an assistant professor of strategy and entrepreneurship at the London Business School. He received his PhD from the Rotman School of Management at University of Toronto. His research interests include creativity, knowledge recombination, innovation and technology strategy, and the relationship between innovation and collaboration.

Hovig Tchalian (Hovig.Tchalian@cgu.edu) studies the impact of language and language-based processes on social and institutional innovation. He uses a mixed-methods approach combining qualitative and computational methods to study the re-emergence of the modern Electric Vehicle market, the values that underlie corporate governance, and the upscaling of the Canadian whisky category.

Vern L. Glaser (vglaser@ulberta.ca) is an assistant professor at the Alberta School of Business, University of Alberta. He received his Ph.D. from the University of Southern California. His research investigates how organizations strategically change practices and culture.

Milo Shaoqing Wang (swang7@ualberta.ca) is a doctoral candidate at the Alberta School of Business, University of Alberta. His research examines the construction of various social evaluations, and how it socially and culturally affects organizational strategy and entrepreneurial activity.

Sarah Kaplan (sarah.kaplan@rotman.utoronto.ca) is Distinguished Professor and Director of the Institute for Gender and the Economy at the University of Toronto's Rotman School of Management. Formerly a professor at the Wharton School, she has her PhD from MIT's Sloan School of Management. She is a Senior Editor at Organization Science.

P. Devereaux (Dev) Jennings is the T.A. Graham Professor of Business and Coordinator of the Canadian Centre of CSR (CCCSR) at the Alberta School of Business. He received his M.A. and Ph.D. from Stanford University. Dev has many research interests, including the effects of category evolution in video games and nanotechnology, and the financing of high tech and cleantech innovation.

APPENDIX

TOPIC MODELING RESEARCH IN MANAGEMENT

Following recent efforts by scholars using topic modeling to map literatures (Antons et al., 2016, 2018; Cho et al., 2017; Guerreiro, Rita, & Trigueiros, 2016; Liu, Mai, & MacDonald, 2018; Oh et al., 2017), we used the method to inductively analyze our topic modeling corpus. In this appendix, we provide additional details about our rendering process (see Figure 2 in the main text) that we did not have the space to discuss in the body of the article. To do so sensibly, we need to provide those details within the context of the rendering steps that we discussed in the body. As a result, this appendix represents a standalone description of our topic modeling effort.

Rendering a Corpus

As highlighted in the main text, to identify management subjects on which topic modeling has been making an impact, we first curated relevant journal articles that leveraged topic modeling methods—not a simple task, for it required rounds of selection and trimming. Specifically, we created a corpus by conducting a computerized text search in Scopus and the Web of Science for article abstracts with keywords signaling topic modeling: "topic model*", "LDA", "Latent Dirichlet Allocation". After pruning articles containing false positives for the LDA acronym (such as "linear discriminant analysis" or "loss distribution approach"), and duplicates, this yielded a vast set of articles (N = 1,466 in 639 publications). Many articles were from computer or information science, so we narrowed out the corpus by curating only include articles from publications that were identified by Scopus and Web of Science as "business" (N = 566 articles in 219 publications). We analyzed this preliminary corpus using topic modeling techniques; we found that there were still many topics that were about algorithms, big textual data, computer science, logistics, and MIS—or just not very interpretable. We continued to narrow our analysis by selecting a subset of articles published in mainstream management journals (e.g., *ASQ* and *SMJ*) and journals from related disciplines that management scholars using topic modeling methods read and cited. For example, we found that many management scholars were influenced by and referenced articles from the special issue in *Poetics* (Mohr & Bogdanov, 2013). Using this approach, we ultimately trimmed the corpus to 66 articles that were directly relevant to management theory.

More specifically, to effectively manage our rendering process in one place, we used Jupyter Notebooks with Python (Kluyver et al., 2016) alongside the libraries Gensim, Pandas, and the Natural Language Toolkit (NLTK). We also used Python to interface (using shell commands) with the Java software packages Mallet and Stanford CoreNLP. In our initial analysis, we relied on abstracts and titles for topic modeling. However, following on Mohr and Bogdanov (2013)—particularly in light of caution by Crossley et al. (2017) to use more than 1,000 documents and 20,000 words for good convergence—we downloaded the full content of articles as PDFs, then used Python to break them down into paragraphs and clean the text. Our paragraph tokenization process was custom-written in Python and based on regular expressions corresponding to common patterns manually found in improper paragraph breaks. This analysis was applied across all 66 articles and resulted in 5,362 paragraphs, the latter serving as the "documents" for LDA.

Before doing detailed cleaning of the text, we first attempted to identify common phrases. Following the procedure from Antons et al. (2016) to identify and replace n-grams in each paragraph, we used an algorithm from NLTK that analyzed common bigrams and trigrams appearing in each paragraph. We then manually coded each phrase as interpretable, given our domain expertise. For all phrases coded as interpretable, we collapsed them into a single token by substituting a "-" character for space characters (i.e., "big data" became "big-data"). The insight here was to collapse common phrases such as "social media" that have interpretable meaning, which would be lost when LDA scrambles word order in the bag of words projection (Wang, McCallum, & Wei, 2007). We also examined high and low relevance and common phrases to be sure that we had stable and unique keywords for our topics, thus removing phrases such as "latent Dirichlet allocation."

After processing phrases, we cleaned each paragraph using the NLP parsing approach with the Stanford CoreNLP software. This computational linguistics/NLP tool broke down each paragraph into constituent sentences, removed punctuation, then analyzed each word according to their part of speech to determine an adequate lemma. For the collapsed phrases, this analysis just reported the full phrase (i.e., "big-data"). Each paragraph was thus converted into a single unordered list of lemmatized words and n-gram phrases. We then assessed that corpus using LDA (applying the Gibbs algorithm for its convergence method) with the number of topics based on the coherence measure data and interpretability. This final corpus used for the LDA contained 5,362 documents with 351,786 distinct words. Table A1 summarizes the end result of our rendered corpus by detailing our final list of 66 articles.

Rendering Topics

To render topics from this corpus, we used the LDA algorithm in two major steps: first, we derived an LDA model from the paragraph dataset, and second, we applied that model to the corpus of 66 articles to derive a topic document matrix. This two-step approach was used by Mohr and Bogdanov (2013) to analyze the paragraph as a unit of analysis in deriving the model, where the corpus needs to be sufficiently large to confidently project a specification for the LDA algorithm that converges. Statistical significance and convergence are functions of the model specification, but this model can then be applied to individual documents to derive a topic probability distribution. The major analytical move here is in using individual paragraphs from all articles ($N = 5362$) generate the model, but then applying it back on the full articles ($N = 66$) to determine the topic document matrix.

The LDA procedure was executed by the software tool Mallet (McCallum, 2002) (see Table A3 for a listing of software packages for topic modeling). A key concern in conducting this procedure is determining the proper number of topics, that is, fitting the topic model. In this process, we initially built on quantitative evidence, using the popular "UMass" measure of topic coherence (Mimno et al., 2011). Topic coherence is a metric done at the level of a topic, developed to match human evaluations of topic quality [see Chang et al. (2009) for a discussion on intrinsic measures of topics not correlating with human judgments]. The UMass metric of coherence considers high scoring words in a topic, tracking the semantic similarity of documents in which they co-occur [see Mimno et al. (2011) for full description]. Stevens, Kegelmeyer, Andrzejewski, and Buttler (2012) extended this coherence score as a measure of overall topic model quality. They generated different topic models based on specifications varying the number of topics (i.e., across a reasonable range generating models in steps of 5 or 10). They then graphed the average topic coherence in each model and looked for evidence of a plateau. We conducted a similar analysis, generating nine different models in Mallet ranging from 10 topics to 50, in steps of 5 (see Figure A1). We followed the procedures from Mallet documentation, setting the hyper-parameters at recommended values and computing diagnostic files for each model. Each diagnostic file was processed in Python to compute average coherence scores. In summary, we projected different LDA models for a range of topics *k*, graphing the coherence measure for each value of *k* between 5 and 50 topics (in increments of 5, so 5, 10, and 15, topics and so on). The coherence graph indicated that 35 topics were ideal as a plateau. For models two steps away on each side of 35 (i.e., 20, 25, 40, and 45 topics), we manually inspected the top topic words for interpretability and confirmed that 35 was adequate.

Rendering Theoretical Artifacts

To render theoretical artifacts from the topic output, inspired by articles such as Croidieu and Kim (2017), Antons et al. (2016), and Mohr et al. (2013), we sought to approach this visually using tools such as LDAvis (Sievert & Shirley, 2014). From this, we developed a four-step process. First, for each topic, we analyzed the MDS plot, reordering the top words according to the relevance metric in Sievert and Shirley (2014), which altered the order between extremes of common words across topics and those uniquely within. We also tracked linkages between topics and documents, using topic weights to form a Topic Significance Ranking (Al Sumait, Barbará, Gentle, & Domeniconi, 2009) to sense the meaning of topics based on domain expertise of articles. Second, we created a "rendering artifact" that synthesized critical information about each topic on one page (see Figure A2). Specifically, we showed the words in the topic (along with the weight of the words), the documents the topic was found in (along with topic weights in documents), and the MDS chart.

Third, three of the coauthors went through each topic and independently assessed the theoretical meaning of these topics and their keywords. Each examined the words and weighted documents (paragraphs in articles) by topic and created first and second-order codes of the topics, which the authors then aggregated into management subject areas. Fourth, the authors compared codes to determine the level of agreement and generated a master spreadsheet of words, topics, articles, key contributions, and subjects (see Table 2). In keeping with theoretical rendering, we paid particular attention to how subject areas were signaled and extended by particular topics, as well as the ways in which topic modeling research introduced new constructs, relationships, and mechanisms into those areas. Both represented the theoretical "delta" of using topic modeling. Such grounded theorizing using axial codes, used by trained experts is relatively standard in management theory today [Bansal & Corley, 2014; Denzin & Lincoln, 2011; Gioia et al., 2013; Pratt, 2009; see also Croidieu and Kim (2018)].

TABLE A1
Rendering the Corpus

Authors	Year	Article Title	Journal
Ahonen, P.	2015	Institutionalizing Big Data methods in social and political research	*Big Data & Society*
Almquist, Z. W., & Bagozzi, B. E.	2017	Using radical environmentalist texts to uncover network structure and network features	*Sociological Methods and Research*
Antons, D., Joshi, A. M., & Salge, T. O.	2018	Content, contribution, and knowledge consumption: Uncovering hidden topic structure and rhetorical signals in scientific texts	*Journal of Management*
Antons, D., Kleer, R., & Salge, T. O.	2016	Mapping the topic landscape of JPIM, 1984–2013: In search of hidden structures and development trajectories	*Journal of Product Innovation Management*
Bail, C. A., Brown, T. W., & Mann, M.	2017	Channeling hearts and minds: Advocacy organizations, cognitive-emotional currents, and public conversation	*American Sociological Review*
Bao, Y., & Datta, A.	2014	Simultaneously discovering and quantifying risk types from textual risk disclosures	*Management Science*
Bendle, N. T., & Wang, X.	2016	Uncovering the message from the mess of Big Data	*Business Horizons*
Blanchard, S. J., Aloise, D., & DeSarbo, W. S.	2017	Extracting summary piles from sorting task data	*Journal of Marketing Research*
Büschken, J., & Allenby, G. M.	2016	Sentence-based text analysis for customer reviews	*Marketing Science*
Buurma, R. S.	2015	The fictionality of topic modeling: Machine reading Anthony Trollope's Barsetshire series	*Big Data & Society*
Cho, Y. J., Fu, P. W., & Wu, C. C.	2017	Popular research topics in marketing journals, 1995–2014	*Journal of Interactive Marketing*
Croidieu, G., & Kim, P. H.	2018	Labor of love: Amateurs and lay-expertise legitimation in the early US radio field	*Administrative Science Quarterly*
DiMaggio, P.	2015	Adapting computational text analysis to social science (and vice versa)	*Big Data & Society*
DiMaggio, P., Nag, M., & Blei, D.	2013	Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of US government arts funding	*Poetics*
Evans, J. A., & Aceves, P.	2016	Machine translation: Mining text for social theory	*Annual Review of Sociology*
Fligstein, N., Stuart Brundage, J., & Schultz, M.	2017	Seeing like the Fed: Culture, cognition, and framing in the failure to anticipate the Financial Crisis of 2008	*American Sociological Review*
Giorgi, S., & Weber, K.	2015	Marks of distinction: Framing and audience appreciation in the context of investment advice	*Administrative Science Quarterly*
Grimmer, J., & Stewart, B. M.	2013	Text as data: The promise and pitfalls of automatic content analysis methods for political texts	*Political Analysis*
Guerreiro, J., Rita, P., & Trigueiros, D.	2016	A text mining-based review of cause-related marketing literature	*Journal of Business Ethics*
Guo, L., Sharma, R., Yin, L., Lu, R., & Rong, K.	2017	Automated competitor analysis using big data analytics	*Business Process Management Journal*
Guo, X. H., Wei, Q., Chen, G. Q., Zhang, J., & Qiao, D. D.	2017	Extracting representative information on intra-organizational blogging	*MIS Quarterly*
Haans, R.	2019	What's the value of being different when everyone is? (Move to in press? No clean text to topic model)	*Strategic Management Journal*

TABLE A1
(Continued)

Authors	Year	Article Title	Journal
Houghton, J. P., Siegel, M., Madnick, S., Tounaka, N., Nakamura, K., Sugiyama, T., Nakagawa, D., & Shirnen, B.	2017	Beyond keywords: Tracking the evolution of conversational clusters in social media	*Sociological Methods and Research*
Huang, A. H., Lehavy, R., Zang, A. Y., & Zheng, R.	2017	Analyst information discovery and interpretation roles: A topic modeling approach	*Management Science*
Humphreys, A., & Wang, R. J. H.	2018	Automated text analysis for consumer research	*Journal of Consumer Research*
Jacobs, B. J. D., Donkers, B., & Fok, D.	2016	Model-based purchase predictions for large assortments	*Marketing Science*
Jha, H. K., & Beckman, C. M.	2017	A patchwork of identities: Emergence of charter schools as a new organizational form	*Research in the Sociology of Organizations*
Jockers, M. L., & Mimno, D.	2013	Significant themes in 19th-century literature	*Poetics*
Kaplan, S., & Vakili, K.	2015	The double-edged sword of recombination in breakthrough innovation	*Strategic Management Journal*
Kinney, A. B., Davis, A. P., & Zhang, Y.	2018	Theming for terror: Organizational adornment in terrorist propaganda	*Poetics*
Kobayashi, V. B., Mol, S. T., Berkers, H. A., Kismihók, G., & Den Hartog, D. N.	2018	Text mining in organizational research	*Organizational Research Methods*
Lee, H., Kwak, J., Song, M., & Kim, C.	2015	Coherence analysis of research and education using topic modeling	*Scientometrics*
Lee, T., & Bradlow, E.	2011	Automated marketing research using online customer reviews	*Journal of Marketing Research*
Levy, K. E. C., & Franklin, M.	2014	Driving regulation: Using topic models to examine political contention in the U.S. trucking industry	*Social Science Computer Review*
Liu, Y., Mai, F., & MacDonald, C.	2018	A Big-Data approach to understanding the thematic landscape of the field of business ethics, 1982–2016	*Journal of Business Ethics*
Luo, J. H., Pan, X. W., & Zhu, X. Y.	2015	Identifying digital traces for business marketing through topic probabilistic model	*Technology Analysis & Strategic Management*
Marciniak, D.	2016	Computational text analysis: Thoughts on the contingencies of an evolving method	*Big Data & Society*
Marshall, E. A.	2013	Defining population problems: Using topic models for cross-national comparison of disciplinary development	*Poetics*
McFarland, D. A., Ramage, D, Chuang, J., Heer, J., Manning, C. D., & Jurafsky, D.	2013	Differentiating language usage through topic models	*Poetics*
Miller, I. M.	2013	Rebellion, crime and violence in Qing China, 1722–1911: A topic modeling approach	*Poetics*
Moe, W. W., & Schweidel, D. A.	2017	Opportunities for innovation in social media analytics	*Journal of Product Innovation Management*
Mohr, J. W., & Bogdanov, P.	2013	Introduction-Topic models: What they are and why they matter	*Poetics*
Mohr, J. W., Wagner-Pacifici, R., Breiger, R. L., & Bogdanov, P.	2013	Graphing the grammar of motives in National Security Strategies: Cultural interpretation, automated text analysis and the drama of global politics	*Poetics*
Momeni, A., & Rost, K.	2016	Identification and monitoring of possible disruptive technologies by patent-development paths and topic modeling	*Technological Forecasting and Social Change*
Mützel, S.	2015	Facing Big Data: Making sociology relevant	*Big Data & Society*
Nam, H., Joshi, Y. V., & Kannan, P. K.	2017	Harvesting brand information from social tags	*Journal of Marketing*

TABLE A1
(Continued)

Authors	Year	Article Title	Journal
Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M.	2012	Mine your own business	*Marketing Science*
Oh, J., Stewart, A., & Phelps, R.	2017	Topic modeling journal topics	*Journal of Counseling Psychology*
Puranam, D., Narayan, V., & Kadiyali, V.	2017	The effect of calorie posting regulation on consumer opinion: A flexible latent Dirichlet allocation model with informative priors	*Marketing Science*
Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S., Albertson, B., & Rand, D.	2014	Structural topic models for open-ended survey responses	*American Journal of Political Science*
Ruckman, K., & McCarthy, I.	2017	Why do some patents get licensed while others do not?	*Industrial and Corporate Change*
Schmiedel, T., Müller, O., & vom Brocke, J.	2018	Topic modeling as a strategy of inquiry in organizational research: A tutorial with an application example on organizational culture	*Organizational Research Methods*
Shi, Z., Lee, G. M., & Whinston, A. B.	2016	Toward a better measure of business proximity: Topic modeling for industry intelligence	*MIS Quarterly*
Suominen, A., Toivanen, H., & Seppänen, M.	2017	Firm's knowledge profiles	*Technological Forecasting and Social Change*
Tangherlini, T. R., & Leonard, P.	2013	Trawling in the sea of the great unread: Sub-corpus topic modeling and humanities research	*Poetics*
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*
Toubia, O., & Netzer, O.	2017	Idea generation, creativity, and prototypicality	*Marketing Science*
Trusov, M., Ma, L. Y., & Jamal, Z.	2016	Crumbs of the cookie: User profiling in customer-base analysis and behavioral targeting	*Marketing Science*
Underwood, T.	2015	The literary uses of high-dimensional space	*Big Data & Society*
Venugopalan, S., & Rai, V.	2015	Topic based classification and pattern identification in patents	*Technological Forecasting and Social Change*
Wagner-Pacifici, R., Mohr, J. W., & Breiger, R. L.	2015	Ontologies, methodologies, and new uses of Big Data in the social and cultural sciences	*Big Data & Society*
Wang, X., Bendle, N. T., Mai, F., & Cotte, J.	2015	*The Journal of Consumer Research* at 40: A historical analysis	*Journal of Consumer Research*
Wang, Y., & Chaudhry, A.	2018	When and how managers' responses to online reviews affect subsequent reviews	*Journal of Marketing Research*
Wilson, A. J., & Joseph, J.	2015	Organizational attention and technological search in the multibusiness firm: Motorola from 1974 to 1997	*Advances in Strategic Management*
Yau, C., Porter, A., Newman, N., & Suominen, A.	2014	Clustering scientific documents with topic modeling	*Scientometrics*
Zhang, Y. C., Moe, W. W., & Schweidel, D. A.	2017	Modeling the role of message content and influencers in social media rebroadcasting	*International Journal of Research Marketing*

TABLE A2
Rendering Topics

Topic #	Topic Weight (Rank)	Raw Topics
1	14	problem, search, structure, attention, concept, process, exist, unit, create, general
2	32	amateur, field, professional, public, space, radio, actor, theme, expertise, expert
3	20	sample, company, set, select, point, follow, test, dataset, describe, section
4	33	make, pile, task, datum, set, summary, consumer, sort, propose, item
5	1	social, conversation, big-data, language, theory, cognitive, public, shift, meaning, emotional
6	27	frame, context, audience, important, framing, make, process, give, individual, part
7	9	researcher, identify, discuss, insight, decision, subject, culture, specific, approach, organizational
8	12	show, figure, table, top, average, represent, high, present, compare, higher
9	8	work, author, write, literary, passage, read, corpus, series, gender, stm
10	24	form, identity, community, logic, organizational, actor, institutional, application, distinct, school
11	30	group, network, identify, radical, movement, pair, environmental, action, strategy, finding
12	3	user, content, message, social-media, consumer, influence, individual, role, activity, platform
13	10	product, dimension, customer, consumer, attribute, purchase, market, prediction, review, online
14	29	scientific, impact, focus, app, knowledge, article, content, find, rhetorical, attribute
15	5	document, corpus, label, identify, blei, process, algorithm, collection, text, latent
16	4	model, distribution, probability, parameter, observe, estimate, give, latent, assume, fit
17	26	organization, theme, individual, effort, people, comment, strategy, day, term, field
18	23	firm, category, industry, performance, position, distinctiveness, competitor, show, level, competitive
19	35	art, support, term, percent, view, recombination, newspaper, assign, agency, grant
20	11	text, category, approach, human, researcher, code, text-analysis, classification, construct, automate
21	13	effect, variable, significant, increase, estimate, coefficient, test, positive, regression, control
23	21	brand, approach, car, text-mining, map, keyword, association, mention, tag, consumer
24	28	idea, weight, distribution, edge, measure, base, node, combination, average, semantic
25	22	patent, technology, knowledge, technological, citation, identify, path, base, cite, highly
26	6	word, term, sentence, frequency, assign, matrix, common, represent, meaning, count
27	34	financial, fomc, economy, price, market, hypothesis, macroeconomic, primary, discussion, real
28	15	time, period, trend, change, fertility, population, country, context, british, demographic
29	19	review, response, rating, health, restaurant, post, hotel, regulation, find, treatment
30	18	relationship, licensor, characteristic, increase, similar, find, size, licensing, licensee, choice
31	31	information, analyst, report, investor, risk, discovery, interpretation, manager, role, find
32	25	major, rebellion, job, event, state, report, case, crime, level, related
33	2	datum, text, information, analyze, application, collect, tool, amount, online, extract
34	7	article, journal, field, publish, year, citation, scholar, papers, author, paper
35	16	model, text, unsupervised, assumption, political, apply, make, scale, grimmer, learn

TABLE A3
Software for Rendering in Topic Modeling

Software	Environment	Relevant Rendering Steps	URL
Gensim	Python	Corpora, Topics	https://radimrehurek.com/gensim/
Natural Language Tookit (NLTK)	Python	Corpora	http://www.nltk.org
Stanford CoreNLP	Java (with Python wrapper)	Corpora	https://stanfordnlp.github.io/CoreNLP/
Jupyter notebook	Python, R	(All)	https://jupyter.org
Anaconda	Python, R	(All)	https://www.anaconda.com
Matplotlib	Python	Theoretical Artifacts	https://matplotlib.org
Pandas	Python	(All)	https://pandas.pydata.org
MALLET	Java (with Python wrapper)	Topics	http://mallet.cs.umass.edu
RStudio	R	(All)	https://www.rstudio.com
tm (R package)	R	Corpora	https://cran.r-project.org/web/packages/tm/index.html
tidytext (R package)	R	Corpora	https://cran.r-project.org/web/packages/tidytext/index.html
snowballC (R package)	R	Corpora	https://cran.r-project.org/web/packages/SnowballC/index.html
topic models (R package)	R	Topics	https://cran.r-project.org/web/packages/topicmodels/index.html
stm (R package)	R	Topics, Theoretical Artifacts	https://cran.r-project.org/web/packages/stm/index.html
lda (R package)	R	Topics	https://cran.r-project.org/web/packages/lda/index.html
David Blei research group code	Python/R/C/C++	Topics	http://www.cs.columbia.edu/~blei/topicmodeling_software.html
David Mimno Topic Modeling Bibliography of papers and software	Python/R/C/C++/Java	Topics	https://mimno.infosci.cornell.edu/topics.html
LDAvis	R	Theoretical Artifacts	https://cran.r-project.org/package=LDAvis
pyLDAvis	Python	Theoretical Artifacts	https://pyldavis.readthedocs.io/
igraph	Python/R	Theoretical Artifacts	https://igraph.org

FIGURE A1
Rendering Topics with Coherence Scores

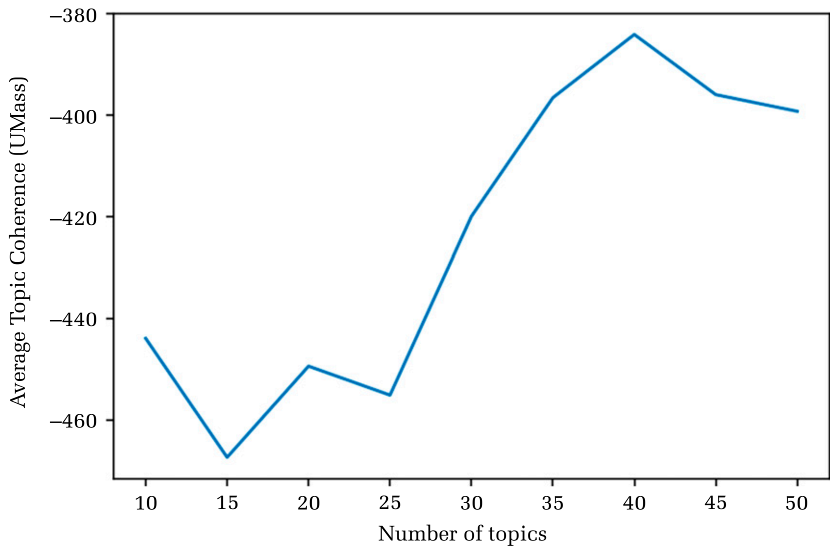


FIGURE A2
Rendering Theoretical Artifact Based on Topic Output