Università Commerciale "Luigi Bocconi"

PhD School

PhD program in: BUSINESS ADMINISTRATION AND MANAGEMENT

Cycle: XXXIII

Disciplinary Field (code): SECS - P/08

NLP methods to inform Marketing Strategy

Advisor: Dirk Hovy

Co-Advisor: Gaia Rubera

PhD Thesis by

Serena Pugliese

ID number: 1442748

Year: 2022

Abstract

This thesis explores the intersection between the marketing literature and Natural Language Processing. From a methodological perspective, NLP has provided marketing scholars with new metrics and tools, like sentiment analysis and topic modelling, that enabled them to analyze textual data more in-depth and on a larger scale. One of the research areas most affected by these methodological advancements is research on Online Word Of Mouth and social media interactions in general. Within this field, we generally observe two types of research that use NLP methods: studies on the effect of social media interaction on external phenomena, like company performance, and studies about the interaction among the online actors. The three essays of this thesis follow this three folded classification (methodology, performance and online dynamics).

In the first one, "The Telephone Game: the effect of online communication similarity on market performance", we study the effect of semantic similarity on market performance. Semantic similarity has been so far neglected in marketing to our knowledge. However, we argue that it is an important dimension of online communication dynamics since it can measure how much of the original brand message gets retained in consumers' communication. This information helps consumers evaluate their fit with the brand, and hence it contributes to the effect of online communication on market performance. We find that semantic similarity positively affects market performance.

The second essay, "Culture of Innovation: A Comprehensive Literature Review Using Natural Language Processing", is intended to provide a methodological contribution. We

argue that, although there is no ready-to-use algorithm for literature reviews, different NLP methods can be used to assist researchers during the literature review process. Hence we try to apply them to review the literature about the culture of innovation.

Finally, the third essay, "Disentangling the "echoverse" for brand communication", is a research proposal about social media dynamics between brands and consumers. The tendency of consumers to diverge from brand content creates a constant tension for brands between the need to keep up with its consumers to keep them engaged and the need to keep control of its own narrative. To assess how the conversational content between brands and consumers changes over time and who drives the change, we will observe the shifts in topics discussed by brands and consumers across ten years of Twitter data.

Contents

1 The Telephone Game: the effect of online communication similarity on market performance 4

1.1 Abstract . . . 4
1.2 Introduction . . . 5
1.3 Theoretical Background . . . 7
1.3.1 Existing metrics to measure social content . . . 7
1.3.2 Semantic Similarity . . . 8
1.3.3 Differences with other metrics . . . 10
1.3.4 Online communication and market performance . . . 12
1.4 Data description . . . 15
1.5 Tweets preprocessing . . . 16
1.6 Measures . . . 17
1.6.1 Similarity . . . 17
1.6.2 Sentiment . . . 19
1.6.3 Volume . . . 19
1.6.4 Sales . . . 19
1.7 Model . . . 20
1.8 Results . . . 21
1.9 Robustness Checks . . . 23
1.9.1 Doc2vec . . . 23

1.9.2 Relationship between similarity and number of users 24
1.10 Discussion 25
1.11 Limitations and Further research 27
Appendices 29
A Regressions with different seeds 29

2 Culture of Innovation: A Comprehensive Literature Review Using Natural Language Processing 54

2.1 Abstract 54
2.2 Introduction 55
2.3 Research Scope 58
2.3.1 Innovation 58
2.3.2 Corporate Culture and National Culture 59
2.3.3 Corporate Culture and National Culture 59
2.4 Methodology 60
2.4.1 Paper Retrieval and Selection: Manual 60
2.4.2 Alternative Approach to Paper Selection: Automated 62
2.4.3 Paper Classification: Latent Dirichlet Allocation (LDA) 67
2.4.4 Results 71
2.5 Discussion 75
2.5.1 Topic 1: Market Orientation and Innovation Performance 75
2.5.2 Sub-topic 1.1: Customer and Competitor Orientation 77
2.5.3 Sub-topic 1.2: Inter-functional Coordination and Cultural Cohesiveness 78
2.5.4 Topic 2: Cultural Advantages and Innovation Performance 79
2.5.5 Sub-topic 2.1: Corporate Culture Advantages 80
2.5.6 Sub-topic 2.2: National Culture Advantages 81

2.6 Towards an Agenda on Culture of Innovation 82
2.7 Contributions 85
2.8 Limitations 87
Appendices 88
A Alternative Automated Approaches to Paper Selection 88
B Papers classified with LDA 89

3 Disentangling the "echoverse" for brand communication 141
3.1 Abstract 141
3.2 Introduction and Theoretical Background 142
3.3 Data Description 145
3.3.1 Brands 145
3.3.2 Consumers 145
3.4 Tweets cleaning 147
3.5 Methodology 148
3.6 Expected results 149

Chapter 1

The Telephone Game: the effect of online communication similarity on market performance

Serena Pugliese, Gaia Rubera, Dirk Hovy

1.1 Abstract

Marketing literature has widely explored the relationship between online communication and market performance, looking at the volume and valence of online communication between consumers and brands. There is a large consensus that volume has a positive effect on sales because it helps spread brand awareness (Babić Rosario et al., 2016), whereas valence can have a positive or negative effect depending on the sentiment expressed (Sonnier et al., 2011; Tang et al., 2014; Hennig-Thurau et al., 2015). However, these approaches do not account for the fact that consumers tend to reframe the brand message when communicating (Kozinets et al., 2010) and that the consumers' evaluation about whether the product is right for them is affected by the combination of information they received

from the company and other consumers (Kuksov et al., 2013). This combination depends on whether one of the two sources provides new information or echoes the information already available (i.e. informativeness, Kuksov et al. 2013). The first aim of this study is to measure how much of the original brand message gets retained in consumers' communication, using semantic similarity, a construct largely neglected in marketing literature. Since this information helps consumers evaluate their fit with the brand, we expect it to contribute to the effect of online communication to market performance. We study this effect using Twitter data and brand sales for brands in the soft drinks industry across ten years. We find that semantic similarity positively affects market performance.

1.2 Introduction

Imagine there are N people in a room playing the telephone game. The first player sets the message. The second player can hear it directly from the first one. From the third position on, the players can receive the message from another player or might overhear something when other players are talking, but in the end, did they get the message intended by the first player?

This situation is very similar to how consumers today are exposed to brand communication. Before social media, brand communication used to be top-down, meaning that the brand (1st player) defined the content of the message it wanted to deliver to consumers (2nd player) and the means to convey the message. The communication received from the company was almost exclusively the only piece of information regarding the brand that could influence the consumer besides word of mouth from people very close to them (i.e., family and friends). With social media, consumers share information about the brand with many other consumers (player 2 to N). Hence, consumers receive two types of communication: (1) from the brand (always top-down) and (2) from other consumers (both as direct conversation or as being exposed to what other consumers are saying to somebody else). Given that brand communication is meant to present the brand in the most interesting way to its consumers, it becomes important for the brand to achieve consistency between its official communication and other consumers' communication. In other words, it becomes important that these two types of communication provide similar information.

Although there is extensive literature on online communication, its effect on market performance have been studied almost exclusively in terms of volume and valence (Kumar et al., 2013; Hennig-Thurau et al., 2015). This approach implicitly assumes that whenever someone mentions a brand, they are repeating the brand message, i.e. "amplification" assumption, (Kozinets et al., 2010), but often consumers do not merely report the brand message. They generally rephrase it using their own words or emphasize aspects of their experience that they find more interesting, whether or not the brand initially mentioned it (Kozinets et al., 2010). So consumers' final evaluation about whether the brand is a good fit will ultimately depend on the information provided by the brand and the information provided by consumers about the brand (Kozinets et al., 2010). However, only the information provided by the brand is accurately selected to serve its strategic goals, so consumers contribution might interfere with its effectiveness.

Following that, we aim to answer two main questions in this study. First, "Can we measure how much of the original message gets retained in consumers' online communication about brands?" Drawing from the latest advancements in the computational socio-linguistics literature, we bring to marketing a novel method to measure the similarity between brand communication and consumers perceptions: semantic similarity, an information-theoretic construct to measure the similarity of meaning (Resnik, 1995; Mikolov et al., 2013). This construct captures how similar the official brand communication is to consumers' perceptions, as evident in consumers' social media activity. To our knowledge, this construct has been largely neglected so far in marketing literature. Second, what is the relationship between this similarity and market performance? Since high similarity might signal consistency between how the brand wants to position itself and consumers' perceptions and should provide more credibility to the brand (Ludwig et al., 2013), it might affect market performance positively. On the other hand, consumers reframe the brand's messages to make them more appealing to their own audience (Berger, 2015), suggesting that preserving the original message might reduce the effect of online communication.

We explore this effect within the online activity of consumers and brands in the soft drinks industry. We analyse a brand's meaning as extracted from its own tweets and other users' tweets that mention a brand to understand how close they are. Then we test whether this similarity affects market performance in terms of sales.

1.3 Theoretical Background

First, we review existing metrics for social media content. Then, we define semantic similarity and highlight how it differs from previously studied marketing constructs. Finally, we review previous work about consumers and brands online communication to explain why similarity might affect this relationship.

1.3.1 Existing metrics to measure social content

The marketing literature has proposed several metrics to capture consumer perceptions, but they suffer from key limitations that semantic similarity might overcome. The oldest and most common metric is volume Asur and Huberman (2010), i.e., the amount of content shared about a specific topic: e.g., a brand, a product. It is generally operationalized as the number of tweets, reviews or posts depending on the platform studied. While volume can be very helpful to measure reach or popularity, it does not provide any information about the content of the message. To understand how volume alone might be misleading, consider what happened in 1985 with the new coke: Coca-Cola launched a new formula and immediately got many reactions from consumers. Consumers felt that it didn't respect the traditional values associated with the brand (Oliver, 2013), and even though the new formula was performing better in blind tastings, the product was a huge failure. If we mapped the consumers' and brand's communication about new coke, we would have observed that Coca-Cola was related to the concept of "new" in its communication, but at the same time, what was relevant about Coca Cola in consumers communication was its "tradition". These are two opposite concepts in terms of meaning, and this misalignment determined the product's failure.

To overcome these issues, the marketing literature has then moved to analyze valence (Sonnier et al., 2011) or attribute specific valence (Liu et al., 2019; Rust et al., 2021), i.e. the distinction between positive, negative, and neutral sentiment. While helpful, valence presents limitations, mainly due to the fact that the sentiment toward a product or brand may remain stable, but the meaning associated with it might change. Take, for instance, the case of gluten-free products. Initially, they were associated with food intolerances. Then, consumers started to perceive them as healthy food regardless of doctors' suggestions. While the valence remained stable, the meaning changed. Thus, we propose a third relevant metric to analyze social content: i.e., semantic similarity, namely the strength of the semantic interactions between two elements only considering taxonomic relationships (Harispe et al., 2015). As we are going to show, semantic similarity is able to capture the change in meaning: from medical necessity to diet preference.

1.3.2 Semantic Similarity

Semantic similarity is an information-theoretic construct rooted in artificial intelligence's goal to replicate the human ability to learn. To achieve this goal, numerous researchers have designed and studied semantic measures. Semantic measures are "mathematical tools used to estimate the strength of the semantic relationship between units of language, concepts or instances, through a (numerical) description obtained according to the comparison of information supporting their meaning" (Harispe et al., 2015). Within the category of semantic measures, it is possible to distinguish semantic relatedness, which captures "the strength of the semantic interactions between two elements with no restrictions on the types of the semantic links considered" (Harispe et al., 2015), and semantic similarity, which is a subset of semantic relatedness that focuses only on taxonomic relationships (Harispe et al., 2015).

To clarify the difference between these two constructs, consider the following example with three very simple phrases:

1. I rented a flat in Milan

2. I rented an apartment in Milan

3. I booked a hotel in Rome

From these phrases, we might infer that "flat" is semantically related to "Milan". Still, semantically similar words are only those with a similar meaning by dint of occurring in similar contexts. So, among all the words used in the example phrases, only "apartment" could be semantically similar to "flat" because they are both things that we learn can be rented and are located in Milan. Meanwhile, "hotel" is less similar to "flat" because it is something that we book and it is in Rome (at least according to the available data). Now, suppose we have access to a much larger corpus of examples. In that case, we could learn more subtle semantic distinctions between concepts (including that hotels and apartments are similar in that we live in them, etc.). As we will explain more extensively in the measures section, similarity can be expressed as a numerical value, measuring the distance between concepts based on their co-occurrence in the corpus. Following this approach, in this research, we represent what a brand "means" based on its own online communication and what it "means" based on its consumers' online communication and map how close these two perceptions are.

1.3.3 Differences with other metrics

We can use semantic similarity to measure brand/product positioning, as in the case of gluten-free products described above. Following on the same example, when gluten-free products are perceived (and hence described) as a medical need they will be more semantically similar to other products consumed for medical reasons. Whereas now, that they are consumed also for other reasons their meaning is partially changing and semantic similarity can capture the fact that the way the product is described is becoming more similar to healthy food or low calories food that it was in the past. This example suggests that similarity can capture brand/product positioning. Past studies have used network measures to capture relative positioning towards competitors (Netzer et al., 2012) or specific attributes the company is interested in (Culotta and Cutler, 2016). Though very informative and easy to implement, these measures can only capture positioning towards pre-identified concepts. Differently, semantic similarity can capture any association expressed by consumers, whether or not the brand initially seeded it.

Two studies are closest to ours. First, Nam et al. (2017), which uses user-defined keywords associated with online content (i.e., social tags) to understand consumers' heterogeneous interpretations of content about brands and products. Our study differs in the following key ways. First, tags are mainly names, so while they are informative of concepts the brand is associated with, they might be less informative about attributes (i.e., adjectives) or relationships among concepts (i.e., verbs). Instead, by using the entire text that consumers and brands posts, we can get a complete picture of brand perceptions. Second, tags sometimes might not reflect the post content, as in the case of users including trending hashtags in their posts only to make them more visible. Finally, semantic similarity captures not only consumers' association but also how far they are from the perception the brand wants to achieve.

The second study closest to ours is Zhong and Schweidel (2020), which analyzes the entire content and develop a model to capture topic shifts in social media content. Their findings suggest how the message's content can shift even without changes in traditional metrics as volume or valence. However, their main focus is to extend LDA to capture latent change-points in the conversation about brands or products that might be more informative than static topic modelling for a company. Topic changes might, in fact, signal brand crisis or attention to new products. Hence the purpose of the method they are proposing is limited to monitoring online activity, while we are also interested in understanding its impact on sales. Answering their calls for future research using "models in which the words are presented in vectors (e.g., Mikolov et al. 2013)", we use Doc2vec(Le and Mikolov, 2014) to measure semantic similarity. Though they both provide information about the content of a text, LDA models and models for measuring semantic similarity, such as Doc2Vec, differ in terms of purpose. LDA (Blei et al., 2003) and its extensions (Zhong and Schweidel, 2020; Büschken and Allenby, 2020) try to identify the topic (or topics) of input text provided. On the other hand, models like Doc2vec (Le and Mikolov, 2014) that allow computing semantic similarity, do not try to identify specific topics discussed. As we will explain more in detail in the measures section, semantic similarity focuses on the differences among different texts, independently from the topics discussed.

This method has already proved to be informative for the marketing strategy, but to our knowledge, their application within the marketing literature has been limited to (1) understanding customer needs and (2) monitor audience communication around political debates. In the first case, Timoshenko and Hauser (2019) extracted customer needs from UGC, but differently from our study, they did not explore the possible effects of semantic similarity. In the second case, Berman et al. (2019) used semantic similarity to capture if and how much the audience tends to detach from the original message around specific points in time of presidential debates. This study also confirms that similarity can vary independently from sentiment, suggesting the need to better understand its theoretical role in brand-consumers online dynamics. However, it focuses on very short time windows(4 mins) and immediate audience reactions. In this study, we focus on a larger time frame to capture a more general brand perception.

1.3.4 Online communication and market performance

There is solid and consistent empirical evidence in the marketing literature about the reliability of online communication as a market performance predictor. This relationship depends both on consumers communication and company communication. For what concerns consumer communication, this literature has focused almost exclusively on only two dimensions, volume and valence (You et al., 2015; Babić Rosario et al., 2016). Though not formally tested in these studies, many of their findings suggest that part of the effect of online communication might be related to the type of information provided.

For example, Asur and Huberman (2010) found a moderate positive correlation between retweets and box office performance. Suggesting that consumers repeating brand messages affects performance positively, presumably because of the increased visibility of promotional messages. This finding is also consistent with the results observed accounting for both consumers and brands communication: brand and consumer communication on the same aspects strengthen each other (Gopinath et al., 2014).

Moreover, the literature about the effect of consumers communication suggests that reviews whose content matches the average linguistic style of a specific type of content tend to have a stronger effect on sales (Ludwig et al., 2013). Linguistic style matches focus on a similar usage of function words, i.e. pronouns, prepositions, articles, conjunctions, auxiliary verbs. However, the effect of linguistic style matches in online communities (Ludwig et al., 2013) is related to a broader phenomenon: communication accommodation theory (CAT; Giles 1979). According to CAT, subjects that communicate with each other tend to conform on every level, even non-verbal communication, increasing approval and trust (Pickering and Garrod, 2004). So, if linguistic similarities as the one studied by Ludwig et al. (2013) increase conversion rates because it conveys part of the effects of CAT theory, it might be possible that also other similarities within the text shared have the same effect. More specifically, in this study, we focus on words that would not be classified as functional according to Ludwig et al. (2013) definition. Differently from Ludwig et al. (2013), we removed pronouns, prepositions, articles, conjunctions, and auxiliary verbs from the text as they are generally considered *stop words* in textual analysis. Stop words generally include the same categories of words that Ludwig et al. (2013) listed when defining functional words (i.e., pronouns, prepositions, articles, conjunctions, and auxiliary verbs). In textual analysis, these words are commonly removed during text preprocessing because they are extremely frequent but do not convey any specific meaning. Hence, they are generally considered less likely to convey specific information about the text in which they are contained.

The ability of other similarities to convey part of the CAT benefits might be especially important in the case of brands' communication. As Colicev et al. (2018) pointed out, brands communication tends to be less effective than consumers communication because it tends to be perceived as less honest. However, if communicating similarly to the rest of the community (i.e. the consumers) makes the content more credible and effective (Ludwig et al., 2013), also brand communication might result more credible and effective than usual when it is similar to consumers communication. So our first hypothesis is:

H1a: Semantic similarity between brand and consumers online communication is positively related with brand market performance .

On the other hand, Gong et al. (2017) find that company retweets are effective only for informative tweets, i.e. tweets that contain objective information about the product like where and when it will be available. This finding suggests that only objective information preserve their value when repeated, while other messages shared by the company are not effective when retained in consumers' communication. Consumers, in fact, tend to reframe the message they receive to include their ideas (Asur and Huberman, 2010; Kozinets et al., 2010; Tan et al., 2016; Melumad et al., 2021). The additional information that they provide about their own impression of the brand has two effects. On one side, this additional information affects the expected utility of the product since it contributes to the brand image and evaluations about whether the product is a good fit for the consumer (Kuksov et al., 2013). On the other, the message is modified to make it more interesting for their audience (Kozinets et al., 2010; Berger, 2015), suggesting that maintaining the message similar might reduce the effect of consumers communication, making it less "tailored" for their audience.

The idea that additional information compared to the one already provided makes consumer communication more effective is also consistent with the research findings about sentiment. Tang et al. (2014) suggest that mixed neutral statements (i.e., tweets containing both positive and negative claims) increase the effect of other information (like positive or negative tweets) because it provides more novel information about the product that might increase curiosity and attention towards it. If that is true, we would expect that when consumers diverge from official brand communication, consumers communication will become more effective since it would provide more novel information. Also, Hennig-Thurau et al. (2015) observed that negative comments are more effective in driving sales also because they are perceived as more honest than company official communication. In contrast, positive comments tend to confirm the information already provided by the company, making them less effective in influencing other consumers. This finding suggests that the effect of consumers' communication on sales is related to the fact that consumers provide information that diverges from information provided by the company.

Therefore, given that:

1. highly similar messages like retweets are not very effective in driving sales (Gong et al., 2017);

2. part of the effect of consumers communication is posited to be related to the fact

that it provides novel information compared to the one already available from other sources (Tang et al., 2014; Hennig-Thurau et al., 2015);

3. consumers reframe messages in a way that would make them more attractive to their audience (Kozinets et al., 2010; Berger, 2015);

we expect:

H1b Semantic similarity between brand and consumers online communication is negatively related with brand market performance.

1.4 Data description

We collected sales data from Euromonitor's Passport database (2020) about the soft drink industry by country from 2010 to 2019. We also collected the new products within the soft drinks industry, launched from 2009 to 2018 from Mintel Global New Products Database (2019), which we use as a control variable. Matching the brands from Passport sales and Mintel new products, we obtained a list of brands for which we checked whether they have an official Twitter account in English. We obtained a list of 282 brands for which we identified at least one English account. However, a brand can have multiple accounts because of national accounts (for example, @CocaCola and @CocaColaUK). The maximum number of English accounts per brand we identified is 8. This process resulted in a list of 493 accounts. For these accounts, we scraped from Twitter the user timeline of the brands (i.e., all the tweets posted by the brand) and all the tweets posted by other users mentioning the brand (i.e., @brand) from 2009 to 2018. In total, we were able to collect over 25.5 million tweets.

Not all of these tweets were suitable for our analysis. First, we removed duplicates or non-informative tweets, like tweets containing only links, mentions or hashtags. Of the accounts collected, 18 are not suitable to compute similarity because they were private accounts (hence their tweets could not be scraped), because they never tweeted, or because they lacked any tweets mentioning the brand.

Moreover, many accounts refer to a specific country, so the effect of those tweets will likely reflect only on that specific country. So, we checked whether sales for the country the account was referring to were available on Passport. To identify which country the account was referring to, we first looked at the account description. If a location was not available in the description, we manually checked their tweets and which offline events, national holidays or other accounts they were referring to. We then assigned the location based on the country where these elements were located. We excluded accounts for which sales were not available in the same country the brand account referred to. This matching significantly reduced our sample's number of accounts and brands: we went from 493 accounts and 282 brands to 268 accounts and 156 brands. Although this process was not based on the dimensions of the accounts in terms of followers or tweets, it ended up excluding very small accounts, as evident from the fact that though we had to exclude almost half of the accounts collected, the remaining ones left us with a sample of over 19 million tweets.

Though we searched for tweets containing the mention of the brands we are interested in, the scraper captured also tweets that did not contain the mention in the text but were posted in reply to tweets mentioning the brand. Given that these tweets are outside our research criteria, our final sample included 14.3 million tweets.

1.5 Tweets preprocessing

Before computing similarity, we cleaned the tweets by removing anything else from the text (i.e., hashtags, links, mentions, emojis, numbers). Given the frequent misspellings in social media to emphasize words (for example, "cooooooool" instead of "cool"), we automatically replaced each word with three or more equal letters consecutively, with its correct form. To remove other misspellings that a computer would classify as entirely different words making it harder to understand the meaning of the text, we removed words that occurred less than five times. Finally, we reduced each word to its lemma, identified collocations (i.e., pair of words that tend to occur together), and joined their tokens (i.e., replace "nice day" with "nice_day"). With these "clean tweets," we compared the meaning of a brand, as extracted from the brand tweets, with the brand's meaning as extracted from other consumers' tweets.

1.6 Measures

1.6.1 Similarity

We computed semantic similarity using Doc2Vec (Le and Mikolov, 2014) which represents the "meaning" of a brand as a vector, built to represent the content of the text. Before building our vectors, we aggregated tweets from brands per year to form a single body of text and compared it with the text created with consumers' tweets, about the same brand, from the same year. While it might be counterintuitive, aggregating tweets allows us to have a more precise measure of meaning. We could also treat each tweet individually, build a vector for each of them, and then compute semantic similarity between these two sets. However, this option would compute the similarity between the average vectors of each set. These average vectors are built in a purely mathematical way, without necessarily accounting for the meaning of the original vectors. While aggregating them implies building a vector to represent everything that a brand has said in that group of tweets and compare it directly with the representation of everything its' consumers said. We aggregated them yearly to provide consistency with our dependent variable: unfortunately, sales at the brand level are available only at the year level. So, we computed the similarity between the brand's tweets at time t and consumers' tweets at time t, where t is a year.

After aggregating the tweets to have each group as a single document, we passed the documents through Doc2Vec (Le and Mikolov, 2014). Doc2Vec is an extension of Word2Vec (Mikolov et al., 2013) that learns continuous distributed vector representations for documents. Doc2vec builds two matrixes, a document matrix D where every column represents a document and a word matrix W where every column represents a word. The document and word vectors are concatenated to predict the next word in a context.

Moreover, documents can be tagged with multiple labels. The labels help the computer understand that some documents are related, despite being separate. For example, documents with the same label might be about the same topic, referring to the same entity or any other meaningful information that they have in common. In our case, we labelled each document with a unique id and additional labels for the brand that document was referred to, whether that document contained tweets from the brand or consumers, the year and the category of soft drinks the brand commercializes. The soft drinks category labels were obtained from Passport. To set model parameters, we followed Lau and Baldwin (2016). The vector distance between two "meaning vectors" (i.e., the one extracted from the brand's text and the one extracted from consumers' text) captures how semantically similar their content is. We computed vector distance as cosine similarity. Given two vectors A and B, cosine similarity can be computed as:

$$\cos(\theta) = \frac{\mathbf{AB}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} \mathbf{A}_i \mathbf{B}_i}{\sqrt{\sum_{i=1}^{n} (\mathbf{A}_i)^2} \sqrt{\sum_{i=1}^{n} (\mathbf{B}_i)^2}} \tag{1.1}$$

We preferred cosine similarity to other vector distance measures, as Euclidean distance, because cosine similarity is constrained between -1 (diametrical opposite) and 1 (the same). This range is helpful for interpretation as negative values can be interpreted as more or less dissimilar documents and positive values as more or less similar documents.

1.6.2 Sentiment

For the sentiment, we computed the sentiment of each tweet using Vader (Hutto and Gilbert, 2014), which computes a probability that the tweet is positive, negative or neutral. We then assigned each tweet to the most probable category. Many studies use the number of positive, negative and neutral tweets as sentiment measure. Babić Rosario et al. (2016) argue that the number of tweets per sentiment is a composite volume/valence measure. We agree on this perspective, so we compute sentiment as the percentage of positive, negative and neutral tweets for the tweets mentioning the brand in each year. We computed only sentiment for tweets mentioning the brand and not for tweets from the brand. Tweets posted by the brand will be generally positive or neutral since they are carefully crafted to provide promotional messages which are unlikely to suggest negative emotions given their purpose. Furthermore, sentiment is generally used to capture what people think about the brand, so the sentiment expressed by the brand is uninformative from a theoretical standpoint.

1.6.3 Volume

Consistently with the literature, we measure volume as the number of tweets shared in the period we are interested in. Given that our dependent variable is at the year level, we distinguish two types of volume:
(1) the number of tweets per year shared by the brand;
(2) the number of tweets per year shared by other users mentioning the brand.

1.6.4 Sales

As mentioned, we collected sales data from Passport. Sales are available per brand, per year, by country. We downloaded sales in million litres per year. We used the data only for those countries that matched the brand Twitter account location, identified as previously explained. For brands with multiple accounts and countries available, we summed the sales by brand year.

1.7 Model

To exclude possible reverse causality, we regress sales at time t on Twitter variables (similarity, volume and sentiment) at t-1. Before estimating the model, we checked the distribution of our dependent variable. As expected, sales distribution was heavily skewed, given that many brands can sell very small quantities, while very few brands get a very high volume of sales (skewness=4.34, kurtosis= 25.03). Hence, we log-transformed our variable to normalize it. After transformation our dependent variable appeared to be normally distributed (skewness= -0.46, kurtosis= 2.97). Then we tested for non-stationarity and serial correlation of errors. The unit root test was significant (χ^2= 1743.97, p < 0.001), suggesting that non-stationarity is not a problem in our data. Wooldridge (2015) test for autocorrelation showed that there is first-order serial correlation (F(1, 136) = 29,001, p < 0.001) in our sample, so we will use first difference estimator. Equation 1 represents our model:

$$\begin{aligned}\Delta(y_{bt} - y_{bt-1}) = \alpha_1 &+ \beta_1\Delta(Sim@_{bt-1} - Sim@_{bt-2}) \\ &+ \beta_2\Delta(Volumefrom_{bt-1} - Volumefrom_{bt-2}) \\ &+ \beta_3\Delta(Volume@_{bt-1} - Volume@_{bt-2}) \\ &+ \beta_4\Delta(\%positive_{bt-1} - \%positive_{bt-2}) \\ &+ \beta_5\Delta(\%negative_{bt-1} - \%negative_{bt-2}) \\ &+ \beta_6\Delta(Launches_{bt-1} - Launches_{bt-2}) \\ &+ \beta_7 Controls_{bt} + \Delta u_{bt}\end{aligned} \tag{1.2}$$

Where:

- $\Delta(y_{bt} - y_{bt-1})$ is the difference between the natural logarithm of million litres sold by brand b in year t and the quantity sold at t-1.

- $\Delta(Sim@_{bt-1}$ - $Sim@_{bt-2})$ is the difference between the cosine similarity between the vector representing what brand b posted and what consumers posted mentioning brand b at t-1 and similarity at t-2.

- $\Delta(Volumefrom_{bt-1}$ - $Volumefrom_{bt-2})$ is the difference between the number of tweets posted by brand b at time t -1 and t-2

- $\Delta(Volume@_{bt-1}$ - $Volume@_{bt-2})$ is the difference between the number of tweets posted by consumers mentioning the brand b at time t -1 and t-2

- $\Delta(\%positive_{bt-1}$ - $\%positive_{bt-2})$ and $\Delta(\%negative_{bt-1}$ - $\%negative_{bt-2})$ are the differences in the percentages of positive and negative tweets shared by consumers mentioning brand b between year t-1 and t-2

- $\Delta(Launches_{bt-1}$ - $Launches_{bt-2})$ is the difference in the number of new products launched by brand b between year t-1 and year t-2 reported on Mintel. We controlled for new products since Zhong and Schweidel (2020) observed a shift in topics discussed about brands around a new product launch and Berman et al. (2019) use topics discussed to estimate similarity. Hence new products might be an antecedent of similarity.

- Controls: we control for years, and product category fixed effects. Product category follows the classification reported on Passport, distinguishing among Bottled Water, Carbonates, Concentrates, Energy Drinks, Juice, RTD Coffee, RTD Tea, Sports Drinks. These categories are a standard industry classification consistent also with Mintel's product categories for soft drinks.

1.8 Results

Table 1.1 shows the result for the estimation of the model in Eq 1. After model estimation, we checked variance inflation factors (VIFs) to exclude multicollinearity within our

VARIABLES	Δ Sales
Δ Similarity	0.169***
	(0.057)
Δ Volume @	-0.000
	(0.000)
Δ Volume from	0.000
	(0.000)
Δ %positive	0.034
	(0.06)
Δ %negative	-0.303
	(0.265)
Δ Launches	-0.000
	(0.000)
Constant	0.107***
	(0.041)
Year FE	✓
Soft drink category FE	✓
Observations	795
R-squared	0.057

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 1.1

variables. The VIFs ranged between 1 and 2.9 (only Year dummies VIFs > 2), suggesting no significant multicollinearity in our model. Results show a significant positive effect of similarity between brand and consumers at t-1 on sales at time t (β=0,169, $p < 0.01$), supporting H1a instead of H1b. We do not find a significant effect of volume. This result might be related to the fact volume effects are generally studied in a shorter period like weeks or months, so the effect of the volume of tweets in our sample might have faded away before the subsequent year. The sentiment did not prove to be significant. This result might seem in contrast with its wide usage in the literature. However, the findings about sentiment are less consistent. Some studies, in fact, found no significant effect of sentiment.

1.9 Robustness Checks

1.9.1 Doc2vec

Doc2vec is not deterministic, so we get a different value every time we run it, even with the same data and parameters. As we mentioned earlier, Doc2Vec builds two matrixes, a document matrix and a word matrix. The different values come from the word vectors initialization that are generally seeded with a random number. The random initialization is not problematic for the validity of the performance itself. The task that Doc2Vec is trying to perform is to sort documents so that their relative distance represents the difference in the content. The random initialization essentially changes the "observation point". Imagine sorting N objects so that more similar objects are closer to each other and less similar objects are more distant. To do that, one will need to list first the visible details about each object (i.e. build their vector representation). If the same person tries to repeat the task with the same objects, the list of visible details will slightly change if they observe them any time from a different random point. All the final representations will be different and yet all true. So to have replicable results, we fixed the "starting point" of the word vectors, we manually set an initial seed for Doc2Vec. The seed will assure that any time we run the model with the same data and same parameters, we will get the same similarity values. However, any integer can be used as a seed. To prove that our results are not conditional on the seed value, we computed similarity values with 100 different seeds (from 1 to 100) and then tested our model in Eq 1 on each seed. Table 1.1 shows results obtained using similarity with seed 5, and Appendix A reports the regression results for all the seeds. To compare the results of the different models, we followed the logic of p-curve analysis (Simonsohn et al., 2014). Suppose there is no effect of similarity on sales. In that case, we are equally likely to observe any possible p-value, so the distribution of p-values observed in our different models should be uniformly distributed. As evident in Figure 1.1, the distribution of the p-values of similarity is heavily right-skewed, suggesting that highly significant values are more likely and that the significance of similarity is not conditional on the specific seed used.

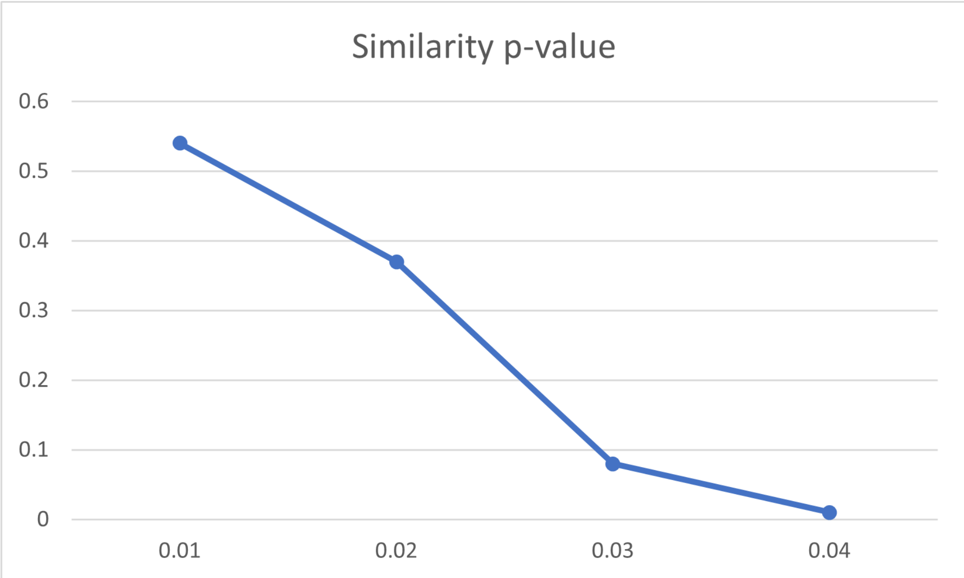


Figure 1.1

1.9.2 Relationship between similarity and number of users

A potential concern about our model might come from the relationship between similarity and the number of people involved in the conversation about the brand. It could be argued that the number of users involved affects both sales and similarity. The more users are involved in the conversations, the higher sales are. At the same time, the more users are involved in the conversation; the more different opinions will be expressed in the tweets, hence reducing similarity. To further exclude any confounding effect of the number of users involved, we re-run our analysis replacing the number of tweets with the number of users involved. Each tweet's author was identified with a unique anonymous number in our data set. We could not include both the number of users and the number of tweets they shared because these variables were highly correlated (0.99). The results of the analysis based on the number of users (Table 1.2 and Figure 1.2) are consistent with the

VARIABLES	**Δ Sales**
Δ Similarity	0.169***
	(0.057)
Δ Users @	1.76e-07
	(4.26e-07)
Δ Volume from	3.53e-07
	(3.70e-07)
Δ %positive	0.035
	(0.06)
Δ %negative	-0.304
	(0.265)
Δ Launches	-0.000
	(0.000)
Constant	0.106***
	(0.041)
Year FE	✓
Soft drink category FE	✓
Observations	795
R-squared	0.057

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 1.2

results obtained using the number of tweets.

1.10 Discussion

To our knowledge, this research is the first within the marketing field that explores the effect of semantic similarity within the social media interaction of brands and consumers on market performance. We observed sales and tweets for brands and consumers within the soft drinks industry for ten years to investigate this effect. Given that receiving highly similar messages, both from the brand and consumers, reduces their novelty, it might make consumers communication less effective (Tang et al., 2014; Hennig-Thurau et al., 2015), suggesting that similarity might harm sales.

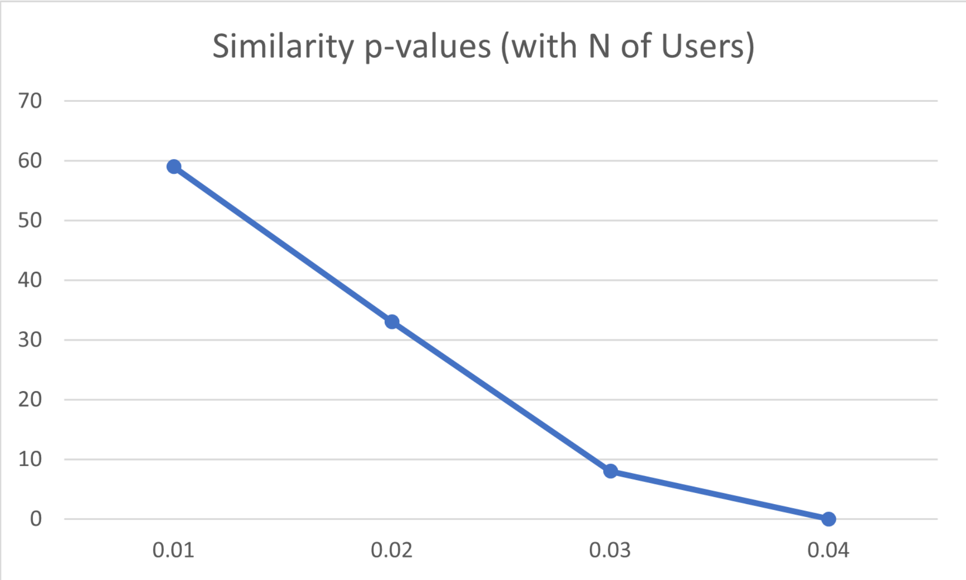


Figure 1.2

Instead, we find that semantic similarity has a positive effect on sales. This effect is in line with the positive effect of retweets observed by Asur and Huberman (2010) and brands and consumers' communication on the same aspects (Gopinath et al., 2014). Similar consumers' communication implies that consumers are repeating the brand message to their audience, hence increasing the visibility of the brand message. Moreover, this positive effect suggests that similarity might help increase consumers trust, as observed for linguistic style match (Ludwig et al., 2013), since it helps the brand to sound more similar to a more reliable source of information (Colicev et al., 2018).

Our results suggest that social media managers should monitor social media activity not only in terms of volume and valence but also in terms of content. According to our findings, it seems important to understand what the brand represents in consumers' minds. Social media managers should frame brand content in a way that is as consistent as possible with consumers' communication. Finally, accounting for the effect of semantic similarity might also help brands identify key opinion leaders among consumers. When identifying key opinion leaders among their consumers, social media managers could evaluate their impact on sales not only in terms of size of their audience and role in their network but also weighting it considering how much they were consistent with the brand official communication.

1.11 Limitations and Further research

Unfortunately, it was not possible to fully exploit the level of details offered by Twitter data. As mentioned earlier, Twitter data proved to be a good predictor of sales even daily. However, despite our extensive search in industry and company databases, sales at the brand level are available only yearly. Some databases, like Orbis, offer quarterly breakdowns, but this information is available only at the company level, which is generally a different entity than the one discussed online. Online communication generally comes from and is referred to individual brands. In contrast, companies can own several brands in multiple industries simultaneously, so their quarterly results will not be based only on the brand observed.

However, we think that this limitation might provide interesting avenues for future research. As explained, some studies posit that the positive effect of online consumers' communication is related to its ability to provide novel information. Increasing semantic similarity is likely to reduce novelty. This reduction of novelty might reduce motivation to engage (Tang et al., 2014), which might in turn decrease volume and interest in the brand in general. So, it is possible that similarity would have a negative effect when evaluated on a shorter term of observation like daily or weekly. At the same time, when observed at a yearly level, it has a positive effect because it reflects slower dynamics, like building trust from consumers. It is important to remember that aggregating text on a yearly level is not equivalent to merely summing the daily effects, similar to what we would observe among two people talking. If we focus only on one conversation, we would likely observe that every time one of the two parties speaks, they are likely to add something new to keep the conversation going. So semantic similarity within that conversation would be low. Whereas if we observe the same two people for one year and simultaneously evaluate the similarity of everything they said to each other in that year, we might observe a higher similarity than the one detected daily. For example, one of the two parties can introduce a new topic in a conversation (which would lower the similarity observed within it). Then the other actor gets interested and starts talking about that topic in the following conversations within the same year, resulting in a higher similarity.

Moreover, it might be interesting to assess whether already explored moderators of online word of mouth effects also affect the relationship between similarity and sales. For example, online word of mouth is more effective when consumers can assess their own similarity (in terms of identity) with the source of the information (Babić Rosario et al., 2016). Suppose we focus only on consumers that receive the brand and consumers messages jointly. In that case, it might also be possible that semantically similar messages are even more effective in contexts in which consumers are aware of how similar they are to the consumers involved. In other words, it might be important to account for the source of similarity and whether it comes from an identity-similar consumer or not.

Finally, the literature in marketing is moving towards more fine-grained measures of sentiment that relate to specific attributes mentioned in the content (Rust et al., 2021; Chakraborty et al., 2021). It might be possible to observe the different effects of similarity based on the attribute on which consumers and brands are converging or diverging. For example, a standard classification for social media content is the distinction between informative content that refers to objective attributes of the brand or product (like promotions, characteristics of the product or service offered) and emotional content that focuses on symbolic attributes of the brand. Intuitively, we expect to observe that similarity on objective attributes is more important than similarity on symbolic attributes.

A Regressions with different seeds

The following table reports the coefficients of the variables included in Eq. 1, when fitting the regression with the similarity values computed with different seeds.

Similarity seed VARIABLES	(1) Δ Sales	(2) Δ Sales	(3) Δ Sales	(4) Δ Sales	(5) Δ Sales
Δ Similarity	0.146**	0.151**	0.151***	0.131**	0.169***
	(0.0581)	(0.0599)	(0.0563)	(0.0601)	(0.0574)
Δ Volume @	-1.43e-08	-6.80e-09	-1.57e-08	-1.71e-08	-1.37e-08
	(3.37e-07)	(3.33e-07)	(3.37e-07)	(3.35e-07)	(3.35e-07)
Δ Volume from	3.85e-07	3.69e-07	3.85e-07	3.72e-07	3.65e-07
	(3.82e-07)	(3.80e-07)	(3.85e-07)	(3.70e-07)	(3.74e-07)
Δ %positive	0.0332	0.0355	0.0346	0.0357	0.0341
	(0.0601)	(0.0602)	(0.0602)	(0.0599)	(0.0600)
Δ %negative	-0.311	-0.286	-0.294	-0.289	-0.303
	(0.267)	(0.265)	(0.265)	(0.266)	(0.265)
Δ Launches	-0.000161	-0.000148	-0.000153	-0.000142	-0.000154
	(0.000324)	(0.000324)	(0.000325)	(0.000324)	(0.000323)
Year FE	✓	✓	✓	✓	✓
Soft Drinks FE	✓	✓	✓	✓	✓
Constant	0.107***	0.106***	0.106**	0.105**	0.107***
	(0.0411)	(0.0411)	(0.0412)	(0.0413)	(0.0410)
Observations	795	795	795	795	795
R-squared	0.056	0.056	0.056	0.054	0.057
F(21, 773)	2.10	2.08	2.13	2.11	2.14
Prob > F	0.003	0.003	0.002	0.003	0.002

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Similarity seed	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
VARIABLES	**Δ Sales**	**Δ Sales**	**Δ Sales**	**Δ Sales**	**Δ Sales**
Δ Similarity	0.151***	0.155***	0.128**	0.158***	0.151***
	(0.0577)	(0.0591)	(0.0550)	(0.0590)	(0.0582)
Δ Volume @	-1.28e-08	-1.80e-08	-1.82e-08	-1.22e-08	-1.91e-08
	(3.38e-07)	(3.36e-07)	(3.37e-07)	(3.36e-07)	(3.34e-07)
Δ Volume from	3.67e-07	3.57e-07	3.69e-07	3.85e-07	3.40e-07
	(3.81e-07)	(3.74e-07)	(3.76e-07)	(3.81e-07)	(3.80e-07)
Δ %positive	0.0336	0.0336	0.0367	0.0363	0.0332
	(0.0597)	(0.0601)	(0.0598)	(0.0596)	(0.0597)
Δ %negative	-0.291	-0.290	-0.286	-0.292	-0.291
	(0.266)	(0.264)	(0.267)	(0.265)	(0.266)
Δ Launches	-0.000127	-0.000161	-0.000125	-0.000128	-0.000151
	(0.000324)	(0.000323)	(0.000323)	(0.000325)	(0.000325)
Year FE	✓	✓	✓	✓	✓
Soft Drinks FE	✓	✓	✓	✓	✓
Constant	0.107***	0.107***	0.106**	0.106**	0.105**
	(0.0411)	(0.0411)	(0.0412)	(0.0412)	(0.0413)
Observations	795	795	795	795	795
R-squared	0.056	0.056	0.054	0.056	0.056
F(21, 773)	2.15	2.15	2.14	2.09	2.08
Prob > F	0.002	0.002	0.002	0.003	0.003

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Similarity seed VARIABLES	(11) Δ Sales	(12) Δ Sales	(13) Δ Sales	(14) Δ Sales	(15) Δ Sales
Δ Similarity	0.130**	0.140**	0.145**	0.142**	0.140**
	(0.0563)	(0.0587)	(0.0571)	(0.0568)	(0.0566)
Δ Volume @	-2.24e-08	-1.40e-08	-1.84e-08	-2.61e-08	-1.47e-08
	(3.36e-07)	(3.40e-07)	(3.36e-07)	(3.37e-07)	(3.35e-07)
Δ Volume from	4.01e-07	4.00e-07	3.59e-07	3.47e-07	3.92e-07
	(3.87e-07)	(3.76e-07)	(3.74e-07)	(3.74e-07)	(3.81e-07)
Δ %positive	0.0345	0.0346	0.0347	0.0357	0.0345
	(0.0601)	(0.0603)	(0.0597)	(0.0595)	(0.0597)
Δ %negative	-0.303	-0.296	-0.295	-0.296	-0.290
	(0.266)	(0.266)	(0.265)	(0.266)	(0.265)
Δ Launches	-0.000130	-0.000136	-0.000140	-0.000148	-0.000143
	(0.000324)	(0.000325)	(0.000323)	(0.000325)	(0.000323)
Year FE	✓	✓	✓	✓	✓
Soft Drinks FE	✓	✓	✓	✓	✓
Constant	0.106***	0.107***	0.105**	0.107***	0.105**
	(0.0411)	(0.0411)	(0.0413)	(0.0411)	(0.0414)
Observations	795	795	795	795	795
R-squared	0.055	0.055	0.055	0.055	0.055
F(21, 773)	2.09	2.07	2.12	2.13	2.11
Prob > F	0.003	0.003	0.003	0.002	0.003

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Similarity seed	**(16)**	**(17)**	**(18)**	**(19)**	**(20)**
VARIABLES	**Δ Sales**	**Δ Sales**	**Δ Sales**	**Δ Sales**	**Δ Sales**
Δ Similarity	0.158***	0.165***	0.147**	0.147**	0.189***
	(0.0588)	(0.0591)	(0.0571)	(0.0596)	(0.0642)
Δ Volume @	-1.71e-08	-1.38e-08	-1.49e-08	-1.70e-08	-1.57e-08
	(3.33e-07)	(3.34e-07)	(3.34e-07)	(3.35e-07)	(3.35e-07)
Δ Volume from	3.90e-07	4.03e-07	3.75e-07	3.70e-07	3.38e-07
	(3.82e-07)	(3.77e-07)	(3.81e-07)	(3.81e-07)	(3.79e-07)
Δ %positive	0.0345	0.0319	0.0339	0.0333	0.0313
	(0.0602)	(0.0596)	(0.0598)	(0.0601)	(0.0594)
Δ %negative	-0.292	-0.305	-0.295	-0.292	-0.316
	(0.266)	(0.266)	(0.266)	(0.265)	(0.267)
Δ Launches	-0.000163	-0.000149	-0.000162	-0.000145	-0.000178
	(0.000324)	(0.000323)	(0.000324)	(0.000325)	(0.000323)
Year FE	✓	✓	✓	✓	✓
Soft Drinks FE	✓	✓	✓	✓	✓
Constant	0.106**	0.106**	0.106**	0.106***	0.106***
	(0.0412)	(0.0412)	(0.0412)	(0.0411)	(0.0411)
Observations	795	795	795	795	795
R-squared	0.056	0.056	0.055	0.055	0.058
F(21, 773)	2.14	2.12	2.08	2.05	2.11
Prob > F	0.002	0.002	0.003	0.004	0.003

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Similarity seed VARIABLES	(21) Δ Sales	(22) Δ Sales	(23) Δ Sales	(24) Δ Sales	(25) Δ Sales
Δ Similarity	0.138**	0.167***	0.143**	0.145**	0.138**
	(0.0581)	(0.0579)	(0.0588)	(0.0582)	(0.0572)
Δ Volume @	-1.75e-08	-1.47e-08	-1.08e-08	-1.35e-08	-1.75e-08
	(3.36e-07)	(3.35e-07)	(3.35e-07)	(3.39e-07)	(3.35e-07)
Δ Volume from	3.93e-07	3.73e-07	3.64e-07	3.51e-07	3.71e-07
	(3.76e-07)	(3.83e-07)	(3.75e-07)	(3.74e-07)	(3.79e-07)
Δ %positive	0.0341	0.0322	0.0356	0.0351	0.0374
	(0.0599)	(0.0601)	(0.0600)	(0.0603)	(0.0603)
Δ %negative	-0.283	-0.288	-0.297	-0.287	-0.291
	(0.266)	(0.265)	(0.265)	(0.265)	(0.266)
Δ Launches	-0.000151	-0.000141	-0.000130	-0.000156	-0.000152
	(0.000324)	(0.000323)	(0.000324)	(0.000323)	(0.000322)
Year FE	✓	✓	✓	✓	✓
Soft Drinks FE	✓	✓	✓	✓	✓
Constant	0.106**	0.106***	0.106**	0.106***	0.107***
	(0.0412)	(0.0411)	(0.0412)	(0.0411)	(0.0411)
Observations	795	795	795	795	795
R-squared	0.055	0.057	0.055	0.055	0.055
F(21, 773)	2.14	2.13	2.09	2.07	2.11
Prob > F	0.002	0.002	0.003	0.003	0.003

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Similarity seed	**(26)**	**(27)**	**(28)**	**(29)**	**(30)**
VARIABLES	**Δ Sales**	**Δ Sales**	**Δ Sales**	**Δ Sales**	**Δ Sales**
Δ Similarity	0.148**	0.154***	0.155**	0.146***	0.148**
	(0.0590)	(0.0557)	(0.0601)	(0.0547)	(0.0588)
Δ Volume @	-1.12e-08	-1.33e-08	-1.43e-08	-3.58e-09	-1.01e-08
	(3.35e-07)	(3.36e-07)	(3.35e-07)	(3.37e-07)	(3.35e-07)
Δ Volume from	3.41e-07	3.85e-07	3.53e-07	3.75e-07	3.96e-07
	(3.84e-07)	(3.83e-07)	(3.68e-07)	(3.73e-07)	(3.82e-07)
Δ %positive	0.0345	0.0355	0.0344	0.0332	0.0340
	(0.0602)	(0.0599)	(0.0598)	(0.0601)	(0.0597)
Δ %negative	-0.285	-0.299	-0.295	-0.292	-0.294
	(0.266)	(0.265)	(0.265)	(0.265)	(0.265)
Δ Launches	-0.000131	-0.000153	-0.000133	-0.000138	-0.000120
	(0.000324)	(0.000324)	(0.000325)	(0.000323)	(0.000323)
Year FE	✓	✓	✓	✓	✓
Soft Drinks FE	✓	✓	✓	✓	✓
Constant	0.106**	0.106**	0.106**	0.107***	0.107***
	(0.0412)	(0.0411)	(0.0412)	(0.0411)	(0.0410)
Observations	795	795	795	795	795
R-squared	0.055	0.056	0.056	0.056	0.055
F(21, 773)	2.09	2.14	2.13	2.10	2.19
Prob > F	0.003	0.002	0.002	0.003	0.002

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Similarity seed VARIABLES	(31) Δ Sales	(32) Δ Sales	(33) Δ Sales	(34) Δ Sales	(35) Δ Sales
Δ Similarity	0.134**	0.149**	0.144***	0.142**	0.148***
	(0.0542)	(0.0606)	(0.0552)	(0.0553)	(0.0536)
Δ Volume @	-1.40e-08	-2.31e-08	-1.53e-08	-1.44e-08	-1.60e-08
	(3.35e-07)	(3.34e-07)	(3.35e-07)	(3.34e-07)	(3.34e-07)
Δ Volume from	4.08e-07	3.85e-07	3.71e-07	3.86e-07	3.39e-07
	(3.74e-07)	(3.78e-07)	(3.78e-07)	(3.79e-07)	(3.87e-07)
Δ %positive	0.0364	0.0348	0.0329	0.0365	0.0324
	(0.0597)	(0.0600)	(0.0593)	(0.0602)	(0.0597)
Δ %negative	-0.294	-0.285	-0.298	-0.285	-0.290
	(0.265)	(0.265)	(0.265)	(0.265)	(0.265)
Δ Launches	-0.000127	-0.000145	-0.000144	-0.000156	-0.000150
	(0.000324)	(0.000322)	(0.000323)	(0.000325)	(0.000323)
Year FE	✓	✓	✓	✓	✓
Soft Drinks FE	✓	✓	✓	✓	✓
Constant	0.105**	0.107***	0.106**	0.105**	0.106**
	(0.0412)	(0.0411)	(0.0412)	(0.0412)	(0.0411)
Observations	795	795	795	795	795
R-squared	0.055	0.056	0.055	0.055	0.056
F(21, 773)	2.14	2.04	2.12	2.13	2.16
Prob > F	0.002	0.004	0.003	0.002	0.002

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Similarity seed	(36)	(37)	(38)	(39)	(40)
VARIABLES	Δ Sales	Δ Sales	Δ Sales	Δ Sales	Δ Sales
Δ Similarity	0.147**	0.147***	0.151***	0.151**	0.151***
	(0.0602)	(0.0552)	(0.0551)	(0.0587)	(0.0553)
Δ Volume @	-1.27e-08	-1.01e-08	-1.46e-08	-1.51e-08	-2.07e-08
	(3.36e-07)	(3.36e-07)	(3.35e-07)	(3.34e-07)	(3.32e-07)
Δ Volume from	3.59e-07	3.72e-07	3.35e-07	3.63e-07	3.83e-07
	(3.75e-07)	(3.70e-07)	(3.69e-07)	(3.72e-07)	(3.76e-07)
Δ %positive	0.0350	0.0340	0.0341	0.0348	0.0354
	(0.0603)	(0.0599)	(0.0598)	(0.0600)	(0.0603)
Δ %negative	-0.283	-0.291	-0.293	-0.283	-0.298
	(0.265)	(0.265)	(0.265)	(0.266)	(0.266)
Δ Launches	-0.000167	-0.000154	-0.000150	-0.000156	-0.000153
	(0.000323)	(0.000323)	(0.000325)	(0.000325)	(0.000323)
Year FE	✓	✓	✓	✓	✓
Soft Drinks FE	✓	✓	✓	✓	✓
Constant	0.105**	0.106**	0.105**	0.105**	0.106***
	(0.0412)	(0.0411)	(0.0412)	(0.0412)	(0.0411)
Observations	795	795	795	795	795
R-squared	0.055	0.056	0.056	0.056	0.056
F(21, 773)	2.17	2.10	2.14	2.07	2.14
Prob > F	0.002	0.003	0.002	0.003	0.002

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Similarity seed VARIABLES	(41) Δ Sales	(42) Δ Sales	(43) Δ Sales	(44) Δ Sales	(45) Δ Sales
Δ Similarity	0.169***	0.151**	0.173***	0.156***	0.173***
	(0.0623)	(0.0589)	(0.0622)	(0.0573)	(0.0607)
Δ Volume @	-2.99e-09	-1.41e-08	-3.70e-09	-1.36e-08	-1.29e-08
	(3.36e-07)	(3.37e-07)	(3.36e-07)	(3.33e-07)	(3.37e-07)
Δ Volume from	3.65e-07	3.82e-07	3.50e-07	3.49e-07	3.58e-07
	(3.88e-07)	(3.75e-07)	(3.74e-07)	(3.82e-07)	(3.82e-07)
Δ %positive	0.0351	0.0371	0.0327	0.0346	0.0335
	(0.0601)	(0.0602)	(0.0596)	(0.0599)	(0.0598)
Δ %negative	-0.286	-0.291	-0.301	-0.296	-0.299
	(0.265)	(0.265)	(0.265)	(0.265)	(0.264)
Δ Launches	-0.000153	-0.000142	-0.000152	-0.000143	-0.000144
	(0.000323)	(0.000324)	(0.000323)	(0.000323)	(0.000323)
Year FE	✓	✓	✓	✓	✓
Soft Drinks FE	✓	✓	✓	✓	✓
Constant	0.106**	0.106**	0.106***	0.107***	0.107***
	(0.0411)	(0.0412)	(0.0411)	(0.0411)	(0.0410)
Observations	795	795	795	795	795
R-squared	0.057	0.056	0.057	0.056	0.057
F(21, 773)	2.14	2.15	2.09	2.15	2.11
Prob > F	0.002	0.002	0.003	0.002	0.003

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Similarity seed VARIABLES	(46) Δ Sales	(47) Δ Sales	(48) Δ Sales	(49) Δ Sales	(50) Δ Sales
Δ Similarity	0.153***	0.142***	0.184***	0.169***	0.118**
	(0.0573)	(0.0531)	(0.0617)	(0.0567)	(0.0539)
Δ Volume @	-9.90e-09	-1.46e-08	-6.88e-09	-1.52e-08	-2.25e-08
	(3.36e-07)	(3.35e-07)	(3.35e-07)	(3.35e-07)	(3.36e-07)
Δ Volume from	3.92e-07	3.48e-07	3.76e-07	3.57e-07	3.69e-07
	(3.81e-07)	(3.78e-07)	(3.70e-07)	(3.87e-07)	(3.76e-07)
Δ %positive	0.0349	0.0347	0.0323	0.0349	0.0357
	(0.0597)	(0.0597)	(0.0601)	(0.0599)	(0.0600)
Δ %negative	-0.291	-0.293	-0.281	-0.285	-0.291
	(0.266)	(0.265)	(0.263)	(0.266)	(0.266)
Δ Launches	-0.000131	-0.000136	-0.000151	-0.000127	-0.000135
	(0.000324)	(0.000325)	(0.000324)	(0.000324)	(0.000323)
Year FE	✓	✓	✓	✓	✓
Soft Drinks FE	✓	✓	✓	✓	✓
Constant	0.106**	0.106***	0.106**	0.106**	0.106**
	(0.0412)	(0.0411)	(0.0411)	(0.0411)	(0.0412)
Observations	795	795	795	795	795
R-squared	0.056	0.055	0.058	0.057	0.054
F(21, 773)	2.17	2.16	2.18	2.18	2.08
Prob > F	0.002	0.002	0.002	0.002	0.003

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Similarity seed VARIABLES	(51) Δ Sales	(52) Δ Sales	(53) Δ Sales	(54) Δ Sales	(55) Δ Sales
Δ Similarity	0.140**	0.126**	0.145**	0.164***	0.131**
	(0.0547)	(0.0549)	(0.0614)	(0.0553)	(0.0579)
Δ Volume @	-1.49e-08	-1.72e-08	-8.11e-09	-3.90e-09	-2.11e-08
	(3.36e-07)	(3.33e-07)	(3.33e-07)	(3.35e-07)	(3.35e-07)
Δ Volume from	3.83e-07	3.80e-07	3.37e-07	3.42e-07	3.69e-07
	(3.83e-07)	(3.78e-07)	(3.80e-07)	(3.85e-07)	(3.77e-07)
Δ %positive	0.0369	0.0367	0.0356	0.0315	0.0355
	(0.0605)	(0.0600)	(0.0598)	(0.0597)	(0.0600)
Δ %negative	-0.297	-0.296	-0.292	-0.305	-0.282
	(0.266)	(0.266)	(0.265)	(0.266)	(0.267)
Δ Launches	-0.000156	-0.000146	-0.000137	-0.000135	-0.000152
	(0.000324)	(0.000324)	(0.000322)	(0.000324)	(0.000325)
Year FE	✓	✓	✓	✓	✓
Soft Drinks FE	✓	✓	✓	✓	✓
Constant	0.106**	0.106**	0.106**	0.107***	0.106**
	(0.0411)	(0.0412)	(0.0412)	(0.0410)	(0.0412)
Observations	795	795	795	795	795
R-squared	0.055	0.055	0.055	0.057	0.055
F(21, 773)	2.24	2.07	2.10	2.30	2.09
Prob > F	0.001	0.003	0.003	0.001	0.003

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Similarity seed	**(56)**	**(57)**	**(58)**	**(59)**	**(60)**
VARIABLES	**Δ Sales**	**Δ Sales**	**Δ Sales**	**Δ Sales**	**Δ Sales**
Δ Similarity	0.167***	0.171***	0.156***	0.158***	0.144**
	(0.0599)	(0.0606)	(0.0581)	(0.0596)	(0.0582)
Δ Volume @	-1.65e-08	-5.94e-09	-1.65e-08	-1.03e-08	-2.11e-08
	(3.36e-07)	(3.36e-07)	(3.38e-07)	(3.34e-07)	(3.37e-07)
Δ Volume from	3.92e-07	3.27e-07	3.43e-07	3.67e-07	3.45e-07
	(3.78e-07)	(3.80e-07)	(3.79e-07)	(3.79e-07)	(3.83e-07)
Δ %positive	0.0345	0.0314	0.0358	0.0330	0.0364
	(0.0605)	(0.0596)	(0.0594)	(0.0596)	(0.0597)
Δ %negative	-0.291	-0.291	-0.298	-0.297	-0.295
	(0.265)	(0.264)	(0.265)	(0.265)	(0.266)
Δ Launches	-0.000149	-0.000144	-0.000144	-0.000149	-0.000152
	(0.000324)	(0.000325)	(0.000323)	(0.000324)	(0.000323)
Year FE	✓	✓	✓	✓	✓
Soft Drinks FE	✓	✓	✓	✓	✓
Constant	0.106**	0.106**	0.106**	0.106**	0.106**
	(0.0411)	(0.0412)	(0.0411)	(0.0411)	(0.0412)
Observations	795	795	795	795	795
R-squared	0.057	0.057	0.056	0.056	0.055
F(21, 773)	2.19	2.12	2.09	2.09	2.05
Prob > F	0.002	0.003	0.003	0.003	0.004

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Similarity seed VARIABLES	(61) Δ Sales	(62) Δ Sales	(63) Δ Sales	(64) Δ Sales	(65) Δ Sales
Δ Similarity	0.161***	0.122**	0.140**	0.158***	0.125**
	(0.0590)	(0.0544)	(0.0547)	(0.0556)	(0.0563)
Δ Volume @	-6.86e-09	-8.66e-09	-1.36e-08	-1.71e-08	-1.22e-08
	(3.35e-07)	(3.34e-07)	(3.35e-07)	(3.34e-07)	(3.37e-07)
Δ Volume from	3.36e-07	3.42e-07	3.72e-07	3.57e-07	3.82e-07
	(3.74e-07)	(3.75e-07)	(3.73e-07)	(3.80e-07)	(3.80e-07)
Δ %positive	0.0327	0.0339	0.0343	0.0356	0.0378
	(0.0599)	(0.0600)	(0.0603)	(0.0601)	(0.0606)
Δ %negative	-0.289	-0.290	-0.285	-0.294	-0.294
	(0.265)	(0.266)	(0.265)	(0.265)	(0.267)
Δ Launches	-0.000150	-0.000124	-0.000150	-0.000149	-0.000160
	(0.000323)	(0.000323)	(0.000324)	(0.000325)	(0.000323)
Year FE	✓	✓	✓	✓	✓
Soft Drinks FE	✓	✓	✓	✓	✓
Constant	0.106**	0.105**	0.106**	0.106**	0.104**
	(0.0411)	(0.0413)	(0.0412)	(0.0412)	(0.0413)
Observations	795	795	795	795	795
R-squared	0.056	0.054	0.055	0.056	0.054
F(21, 773)	2.16	2.11	2.15	2.10	2.01
Prob > F	0.002	0.003	0.002	0.003	0.005

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Similarity seed	(66)	(67)	(68)	(69)	(70)
VARIABLES	Δ Sales	Δ Sales	Δ Sales	Δ Sales	Δ Sales
Δ Similarity	0.161***	0.150**	0.146**	0.167***	0.141**
	(0.0594)	(0.0596)	(0.0591)	(0.0611)	(0.0577)
Δ Volume @	-3.39e-09	-6.65e-09	-1.55e-08	-8.76e-09	-1.71e-08
	(3.34e-07)	(3.38e-07)	(3.36e-07)	(3.36e-07)	(3.34e-07)
Δ Volume from	3.73e-07	3.66e-07	3.95e-07	3.63e-07	3.47e-07
	(3.78e-07)	(3.83e-07)	(3.77e-07)	(3.80e-07)	(3.68e-07)
Δ %positive	0.0359	0.0377	0.0369	0.0345	0.0330
	(0.0603)	(0.0601)	(0.0597)	(0.0597)	(0.0596)
Δ %negative	-0.294	-0.293	-0.294	-0.285	-0.288
	(0.265)	(0.265)	(0.267)	(0.265)	(0.266)
Δ Launches	-0.000159	-0.000149	-0.000152	-0.000153	-0.000163
	(0.000323)	(0.000324)	(0.000324)	(0.000324)	(0.000324)
Year FE	✓	✓	✓	✓	✓
Soft Drinks FE	✓	✓	✓	✓	✓
Constant	0.106**	0.106***	0.105**	0.106**	0.106**
	(0.0412)	(0.0411)	(0.0413)	(0.0412)	(0.0412)
Observations	795	795	795	795	795
R-squared	0.056	0.056	0.055	0.057	0.055
F(21, 773)	2.14	2.13	2.10	2.15	2.09
Prob > F	0.002	0.002	0.003	0.002	0.003

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Similarity seed VARIABLES	(71) Δ Sales	(72) Δ Sales	(73) Δ Sales	(74) Δ Sales	(75) Δ Sales
Δ Similarity	0.168***	0.154**	0.150***	0.201***	0.143**
	(0.0594)	(0.0609)	(0.0572)	(0.0651)	(0.0586)
Δ Volume @	-8.57e-09	-1.35e-08	-1.00e-08	-2.42e-09	-1.04e-08
	(3.37e-07)	(3.36e-07)	(3.35e-07)	(3.33e-07)	(3.36e-07)
Δ Volume from	3.36e-07	3.81e-07	3.67e-07	3.66e-07	3.85e-07
	(3.79e-07)	(3.79e-07)	(3.79e-07)	(3.74e-07)	(3.73e-07)
Δ %positive	0.0361	0.0337	0.0366	0.0319	0.0335
	(0.0599)	(0.0598)	(0.0600)	(0.0594)	(0.0600)
Δ %negative	-0.303	-0.298	-0.293	-0.298	-0.285
	(0.266)	(0.266)	(0.265)	(0.265)	(0.265)
Δ Launches	-0.000150	-0.000150	-0.000134	-0.000150	-0.000151
	(0.000323)	(0.000324)	(0.000325)	(0.000324)	(0.000325)
Year FE	✓	✓	✓	✓	✓
Soft Drinks FE	✓	✓	✓	✓	✓
Constant	0.107***	0.105**	0.106**	0.106***	0.106**
	(0.0411)	(0.0412)	(0.0412)	(0.0411)	(0.0411)
Observations	795	795	795	795	795
R-squared	0.057	0.056	0.056	0.059	0.055
F(21, 773)	2.17	2.10	2.14	2.14	2.10
Prob > F	0.002	0.003	0.002	0.002	0.003

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Similarity seed	(76)	(77)	(78)	(79)	(80)
VARIABLES	Δ Sales	Δ Sales	Δ Sales	Δ Sales	Δ Sales
Δ Similarity	0.153**	0.160***	0.151***	0.148**	0.165***
	(0.0611)	(0.0572)	(0.0573)	(0.0594)	(0.0611)
Δ Volume @	-1.25e-08	-9.30e-09	-1.86e-08	-2.02e-08	0
	(3.37e-07)	(3.34e-07)	(3.36e-07)	(3.37e-07)	(3.38e-07)
Δ Volume from	3.87e-07	3.82e-07	3.50e-07	3.86e-07	3.52e-07
	(3.88e-07)	(3.78e-07)	(3.78e-07)	(3.74e-07)	(3.83e-07)
Δ %positive	0.0352	0.0341	0.0345	0.0350	0.0326
	(0.0598)	(0.0600)	(0.0596)	(0.0596)	(0.0600)
Δ %negative	-0.302	-0.297	-0.286	-0.289	-0.303
	(0.266)	(0.265)	(0.264)	(0.264)	(0.265)
Δ Launches	-0.000162	-0.000160	-0.000161	-0.000158	-0.000138
	(0.000323)	(0.000323)	(0.000325)	(0.000323)	(0.000324)
Year FE	✓	✓	✓	✓	✓
Soft Drinks FE	✓	✓	✓	✓	✓
Constant	0.106**	0.106**	0.106**	0.106**	0.106***
	(0.0411)	(0.0411)	(0.0411)	(0.0412)	(0.0411)
Observations	795	795	795	795	795
R-squared	0.056	0.056	0.056	0.056	0.056
F(21, 773)	2.04	2.08	2.14	2.05	2.16
Prob > F	0.004	0.003	0.002	0.004	0.002

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Similarity seed VARIABLES	(81) Δ Sales	(82) Δ Sales	(83) Δ Sales	(84) Δ Sales	(85) Δ Sales
Δ Similarity	0.130**	0.151**	0.158***	0.149***	0.154***
	(0.0530)	(0.0604)	(0.0588)	(0.0554)	(0.0580)
Δ Volume @	-2.48e-08	-1.40e-08	-1.03e-08	-1.26e-08	-4.35e-09
	(3.36e-07)	(3.34e-07)	(3.35e-07)	(3.35e-07)	(3.35e-07)
Δ Volume from	3.78e-07	3.91e-07	3.76e-07	3.85e-07	3.64e-07
	(3.76e-07)	(3.82e-07)	(3.82e-07)	(3.81e-07)	(3.71e-07)
Δ %positive	0.0347	0.0367	0.0327	0.0353	0.0335
	(0.0598)	(0.0598)	(0.0603)	(0.0601)	(0.0598)
Δ %negative	-0.294	-0.291	-0.295	-0.305	-0.296
	(0.266)	(0.266)	(0.264)	(0.266)	(0.265)
Δ Launches	-0.000150	-0.000142	-0.000149	-0.000144	-0.000128
	(0.000324)	(0.000324)	(0.000324)	(0.000323)	(0.000323)
Year FE	✓	✓	✓	✓	✓
Soft Drinks FE	✓	✓	✓	✓	✓
Constant	0.105**	0.106**	0.106**	0.106**	0.106***
	(0.0412)	(0.0412)	(0.0411)	(0.0411)	(0.0411)
Observations	795	795	795	795	795
R-squared	0.055	0.056	0.056	0.056	0.056
F(21, 773)	2.10	2.05	2.11	2.14	2.17
Prob > F	0.003	0.004	0.003	0.002	0.002

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Similarity seed	(86)	(87)	(88)	(89)	(90)
VARIABLES	Δ Sales	Δ Sales	Δ Sales	Δ Sales	Δ Sales
Δ Similarity	0.149***	0.159***	0.157***	0.153**	0.174***
	(0.0545)	(0.0599)	(0.0567)	(0.0605)	(0.0615)
Δ Volume @	-1.10e-08	-1.35e-08	-1.52e-08	-1.15e-08	-5.99e-09
	(3.34e-07)	(3.34e-07)	(3.33e-07)	(3.34e-07)	(3.32e-07)
Δ Volume from	3.71e-07	3.28e-07	3.85e-07	3.89e-07	3.58e-07
	(3.86e-07)	(3.80e-07)	(3.80e-07)	(3.70e-07)	(3.86e-07)
Δ %positive	0.0368	0.0333	0.0332	0.0368	0.0373
	(0.0597)	(0.0597)	(0.0594)	(0.0601)	(0.0597)
Δ %negative	-0.293	-0.299	-0.289	-0.291	-0.294
	(0.265)	(0.266)	(0.265)	(0.265)	(0.265)
Δ Launches	-0.000170	-0.000162	-0.000132	-0.000161	-0.000144
	(0.000324)	(0.000324)	(0.000324)	(0.000324)	(0.000323)
Year FE	✓	✓	✓	✓	✓
Soft Drinks FE	✓	✓	✓	✓	✓
Constant	0.106**	0.106**	0.106**	0.105**	0.107***
	(0.0411)	(0.0412)	(0.0410)	(0.0412)	(0.0411)
Observations	795	795	795	795	795
R-squared	0.056	0.056	0.056	0.056	0.057
F(21, 773)	2.16	2.08	2.12	2.14	2.09
Prob > F	0.002	0.003	0.002	0.002	0.003

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Similarity seed VARIABLES	(91) Δ Sales	(92) Δ Sales	(93) Δ Sales	(94) Δ Sales	(95) Δ Sales
Δ Similarity	0.168***	0.144***	0.157***	0.151**	0.158***
	(0.0581)	(0.0552)	(0.0585)	(0.0595)	(0.0579)
Δ Volume @	-5.87e-09	-1.64e-08	-1.16e-08	-3.09e-09	-6.92e-09
	(3.35e-07)	(3.34e-07)	(3.36e-07)	(3.35e-07)	(3.32e-07)
Δ Volume from	3.59e-07	3.77e-07	3.67e-07	3.71e-07	3.57e-07
	(3.80e-07)	(3.81e-07)	(3.84e-07)	(3.80e-07)	(3.81e-07)
Δ %positive	0.0333	0.0351	0.0315	0.0328	0.0339
	(0.0593)	(0.0598)	(0.0595)	(0.0602)	(0.0604)
Δ %negative	-0.284	-0.296	-0.297	-0.300	-0.296
	(0.266)	(0.265)	(0.265)	(0.265)	(0.265)
Δ Launches	-0.000142	-0.000135	-0.000139	-0.000128	-0.000158
	(0.000324)	(0.000323)	(0.000323)	(0.000324)	(0.000324)
Year FE	✓	✓	✓	✓	✓
Soft Drinks FE	✓	✓	✓	✓	✓
Constant	0.105**	0.107***	0.106***	0.104**	0.105**
	(0.0411)	(0.0411)	(0.0411)	(0.0414)	(0.0412)
Observations	795	795	795	795	795
R-squared	0.057	0.055	0.056	0.056	0.056
F(21, 773)	2.16	2.13	2.10	2.06	2.07
Prob > F	0.002	0.002	0.003	0.004	0.003

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Similarity seed	(96)	(97)	(98)	(99)	(100)
VARIABLES	Δ Sales	Δ Sales	Δ Sales	Δ Sales	Δ Sales
Δ Similarity	0.146***	0.143**	0.166***	0.159***	0.129**
	(0.0562)	(0.0563)	(0.0580)	(0.0565)	(0.0579)
Δ Volume @	-1.30e-08	-1.89e-08	-1.11e-08	-1.74e-08	-2.44e-08
	(3.38e-07)	(3.36e-07)	(3.37e-07)	(3.36e-07)	(3.37e-07)
Δ Volume from	3.55e-07	3.64e-07	3.48e-07	3.44e-07	3.66e-07
	(3.76e-07)	(3.73e-07)	(3.85e-07)	(3.66e-07)	(3.70e-07)
Δ %positive	0.0376	0.0333	0.0327	0.0338	0.0362
	(0.0601)	(0.0600)	(0.0599)	(0.0598)	(0.0599)
Δ %negative	-0.289	-0.291	-0.294	-0.297	-0.289
	(0.266)	(0.265)	(0.265)	(0.265)	(0.266)
Δ Launches	-0.000154	-0.000174	-0.000136	-0.000142	-0.000160
	(0.000325)	(0.000323)	(0.000324)	(0.000324)	(0.000324)
Year FE	✓	✓	✓	✓	✓
Soft Drinks FE	✓	✓	✓	✓	✓
Constant	0.106**	0.106***	0.105**	0.106***	0.106**
	(0.0411)	(0.0411)	(0.0412)	(0.0411)	(0.0412)
Observations	795	795	795	795	795
R-squared	0.055	0.055	0.057	0.056	0.055
F(21, 773)	2.16	2.10	2.19	2.10	2.04
Prob > F	0.002	0.003	0.002	0.003	0.004

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

References

Asur, S. and Huberman, B. A. (2010). Predicting the future with social media. In *2010 IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology*, volume 1, pages 492–499. IEEE.

Babić Rosario, A., Sotgiu, F., De Valck, K., and Bijmolt, T. H. (2016). The effect of electronic word of mouth on sales: A meta-analytic review of platform, product, and metric factors. *Journal of Marketing Research*, 53(3):297–318.

Berger, J. (2015). Word of mouth and interpersonal communication.

Berman, R., Melumad, S., Humphrey, C., and Meyer, R. (2019). A tale of two twitter-spheres: Political microblogging during and after the 2016 primary and presidential debates. *Journal of Marketing Research*, 56(6):895–917.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.

Büschken, J. and Allenby, G. M. (2020). Improving text analysis using sentence conjunctions and punctuation. *Marketing Science*, 39(4):727–742.

Chakraborty, I., Kim, M., and Sudhir, K. (2021). Attribute sentiment scoring with online text reviews: Accounting for language structure and missing attributes.

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

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

Giles, H. (1979). Accommodation theory: Optimal levels of convergence. *Language and social psychology*, pages 45–65.

Gong, S., Zhang, J., Zhao, P., and Jiang, X. (2017). Tweeting as a marketing tool: A field experiment in the tv industry. *Journal of Marketing Research*, 54(6):833–850.

Gopinath, S., Thomas, J. S., and Krishnamurthi, L. (2014). Investigating the relationship between the content of online word of mouth, advertising, and brand performance. *Marketing Science*, 33(2):241–258.

Harispe, S., Ranwez, S., Janaqi, S., and Montmain, J. (2015). Semantic similarity from natural language and ontology analysis. *Synthesis Lectures on Human Language Technologies*, 8(1):1–254.

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

Hutto, C. and Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 8.

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

Kuksov, D., Shachar, R., and Wang, K. (2013). Advertising and consumers' communications. *Marketing Science*, 32(2):294–309.

Kumar, V., Bhaskaran, V., Mirchandani, R., and Shah, M. (2013). Practice prize winner—creating a measurable social media marketing strategy: increasing the value and roi of intangibles and tangibles for hokey pokey. *Marketing Science*, 32(2):194–212.

Lau, J. H. and Baldwin, T. (2016). An empirical evaluation of doc2vec with practical insights into document embedding generation. *arXiv preprint arXiv:1607.05368*.

Le, Q. and Mikolov, T. (2014). Distributed representations of sentences and documents. In *International conference on machine learning*, pages 1188–1196. PMLR.

Liu, X., Lee, D., and Srinivasan, K. (2019). Large-scale cross-category analysis of consumer review content on sales conversion leveraging deep learning. *Journal of Marketing Research*, 56(6):918–943.

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

Melumad, S., Meyer, R., and Kim, Y. D. (2021). The dynamics of distortion: How successive summarization alters the retelling of news. *Journal of Marketing Research*, 58(6):1058–1078.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

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

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

Oliver, T. (2013). *The real Coke, the real story*. Random House.

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

Resnik, P. (1995). Using information content to evaluate semantic similarity in a taxonomy. *arXiv preprint cmp-lg/9511007*.

Rust, R. T., Rand, W., Huang, M.-H., Stephen, A. T., Brooks, G., and Chabuk, T. (2021). Real-time brand reputation tracking using social media. *Journal of Marketing*, 85(4):21–43.

Simonsohn, U., Nelson, L. D., and Simmons, J. P. (2014). P-curve: a key to the file-drawer. *Journal of experimental psychology: General*, 143(2):534.

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

Tan, C., Friggeri, A., and Adamic, L. (2016). Lost in propagation? unfolding news cycles from the source. In *Tenth International AAAI Conference on Web and Social Media*.

Tang, T., Fang, E., and Wang, F. (2014). Is neutral really neutral? the effects of neutral user-generated content on product sales. *Journal of Marketing*, 78(4):41–58.

Timoshenko, A. and Hauser, J. R. (2019). Identifying customer needs from user-generated content. *Marketing Science*, 38(1):1–20.

Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Cengage learning.

You, Y., Vadakkepatt, G. G., and Joshi, A. M. (2015). A meta-analysis of electronic word-of-mouth elasticity. *Journal of Marketing*, 79(2):19–39.

Zhong, N. and Schweidel, D. A. (2020). Capturing changes in social media content: a multiple latent changepoint topic model. *Marketing Science*.

Chapter 2

Culture of Innovation: A Comprehensive Literature Review Using Natural Language Processing

Serena Pugliese, Verdiana Giannetti, Sourindra Banerjee

2.1 Abstract

It is now widely accepted that firms need to develop a culture where innovation flourishes in order to survive and thrive in today's hypercompetitive environment. As a result, the concept of culture of innovation, i.e., a culture (corporate or national) that fosters relentless innovation, has attracted widespread attention from scholars, resulting in a burgeoning literature across different fields including management, marketing, and international business. Different from other research streams in innovation, the literature on culture of innovation has not been organized into systematic frameworks yet. We argue that gaining a full understanding of the topics and trends in research on culture of innovation is critical for a better comprehension of this important field. In the present research, we conduct

a systematic review of the literature on culture of innovation using Natural Language Processing (NLP). Specifically, we implement a topic modelling analysis based on Latent Dirichlet Allocation (LDA) on 254 papers published in 1996-2019 across 15 well-renowned journals. The LDA analysis revealed two main topics, i.e., (1) market orientation and innovation performance and (2) cultural advantages and innovation performance. Within the first topic, we further identified two sub-topics, i.e., (1.1) customer and competitor orientation and (1.2) inter-functional coordination and cultural cohesiveness. Within the second topic, we identified two sub-topics, i.e., (2.1) corporate culture advantages and (2.2) national culture advantages. To the best of our knowledge, this is the first literature review in innovation scholarship to use LDA. Besides offering a necessary, objective, and exhaustive review of the broad literature on culture of innovation, this study also provides future researchers with detailed guidance on how to use LDA for their literature reviews. Actionable future research directions to advance the concept of culture of innovation are provided.

2.2 Introduction

The concept of culture of innovation, i.e., a culture that fosters relentless innovation, has long captured the attention of academics and practitioners alike (Tellis et al., 2009). The intrigue associated with culture of innovation has been noted in academic research quotes such as "culture [of innovation] is a uniquely human product that develops slowly within firms, is tacit and not easily defined, and is not easily transported across firms" (Tellis et al. 2009, p. 7). The importance of developing and leveraging a culture of innovation has also been highlighted by practitioners such as Vinton G. Cerf, VP and Chief Internet Evangelist at Google, winner of the Presidential Medal of Freedom in 2005: "At Google, we've spent years thinking about how to maintain and improve a culture that fosters transformation and innovation. This has led to alignment around certain core principles that have informed our approach and supported Google's culture for two decades" (Cerf, 2020).

Copious amounts of academic research have examined culture of innovation through the lens of corporate- (e.g., risk-tolerance, willingness to cannibalize, percentage of R&D employees among all employees, etc.) and/or national-level variables (e.g., national values, R&D spending, patents, etc.) and attempted to establish its importance in the creation of new products/processes and, in turn, in terms of market performance and financial value for firms. This vast pool of academic research cuts across different fields such as management, marketing, and international business. Broad generalizations classify research on culture of innovation into two main streams, i.e., (1) research on the impact of corporate culture and (2) research on the impact of national culture on innovative outcomes (Tellis et al., 2009). We argue that the burgeoning number of studies on culture of innovation makes it imminent that we move away from such broad generalizations to a more comprehensive and systematic review of the extant literature. While research in other areas of innovation has been organized into systematic frameworks (see Rubera and Kirca 2012 for an extensive meta-analytic review of the literature on innovation and its outcomes), research on culture of innovation has not been organized into such systematic frameworks yet[1]. We conduct a systematic review of the literature on culture of innovation using Natural Language Processing (NLP). Specifically, we implement a topic modelling analysis based on Latent Dirichlet Allocation (LDA) on 254 automatically selected papers published in 1996-2019 across 15 well-renowned journals. The LDA analysis revealed two main topics, i.e., (1) market orientation and innovation performance and (2) cultural advantages and innovation performance. Within the first topic we further identified two sub-topics, i.e., (1.1) customer and competitor orientation and (1.2) inter-functional coor-

[1]We acknowledge the existence of two meta-analytic reviews on similar topics (Büschgens, Bausch, and Balkin 2013; Eisend, Evanschitzky, and Gilliland 2015). While we build on these notable precedents, we also distinguish ourselves in several ways, i.e., by adopting (1) substantially broader definitions for both culture and innovation (and, as a result by using a substantially broader set of papers), (2) a structured, objective, replicable approach, and (3) an assumption-free approach allowing the literature to speak for itself.

dination and cultural cohesiveness. Within the second topic we identified two sub-topics, i.e., (2.1) corporate culture advantages and (2.2) national culture advantages.

The present study makes at least four contributions to the innovation literature. First, from a theoretical perspective, we respond to calls to "reflect back, take stock of knowledge and organize various subfields and topics in innovation and product development research" (Noble et al. 2021, p. 2), by conducting a literature review on culture of innovation within the broader fields of management, marketing, and international business. To the best of our knowledge, we are the first to provide almost two and half decades of research on culture of innovation with a much-needed structure. Second, again from a theoretical perspective, in addition to integrating contributions from several domains, we provide an extensive research agenda to advance scholarship on culture of innovation. In so doing, we lay the groundwork for the next generation of research in the field, by identifying areas in which further research needs to be pursued. Third, from a methodological perspective, we respond to calls for "utilizing different approaches to reflect on extant knowledge" (Noble et al. 2021, p. 1) by being the first to bring to innovation scholarship the approach of conducting a literature review using NLP. NLP, "a computer-assisted analytical technique aimed at automatically analyzing and comprehending human language (Manning and Schutze, 1999), allows scholars to easily extract beneficial insights contained in textual datasets while avoiding burdensome computational work (Collobert et al., 2011; Green, 2012)" (Kang et al. 2020, p. 139). In this study, NLP allowed us to identify the key topics and sub-topics in research on culture of innovation in an objective and efficient manner, i.e., eliminating the subjectivity and tedious work associated with manual textual coding of research papers. Finally, again from a methodological perspective, this study offers a tutorial detailing how NLP can be employed in literature reviews, with clear benefits in terms of objectivity, replicability, and efficiency.

The rest of the paper is organized as follows. The next section clarifies the research scope. This is followed by a section which details the methodology used in the study. Next, we articulate the topics and sub-topics unraveled by the LDA analysis. Finally, we provide an extensive future research agenda on culture of innovation.

2.3 Research Scope

In the following paragraphs, we clarify the research scope by offering definitions for the key constructs, i.e., innovation, culture, and culture of innovation.

2.3.1 Innovation

Given our objective to offer a comprehensive review of the literature on culture of innovation, we hereinafter adopt a holistic view of innovation, which includes innovation as well as the cognate constructs of creativity and entrepreneurship. In so doing, we adopt the definitions of innovation, creativity, and entrepreneurship offered by Amabile et al. (1996). Specifically, while creativity is the production of novel and useful ideas in any domain, innovation is the successful implementation of creative ideas within an organization (Amabile et al., 1996), making the two constructs closely related. Further, entrepreneurship is a particular type of innovation, i.e., the successful implementation of creative ideas to produce a new business or a new initiative within an existing business (Amabile et al., 1996). Extant research identifies two primary ways through which innovation occurs in firms, i.e., (1) the development of new product-markets, i.e., product (or service) innovation, and (2) the development of new and more efficient processes, ranging from manufacturing processes to distribution processes, i.e., process (or administrative) innovation (Cillo et al., 2018). We consider both the above-mentioned types of innovation as pertinent ways though which we capture innovation in the current research. This is also consistent with the definition of innovation by Crossan and Apaydin (2010) as the "production or adoption, assimilation, and exploitation of value added novelty in economic social spheres; renewal and enlargement of products, services, and markets; development of new methods of production, and establishment of new management systems" (p. 1155).

2.3.2 Corporate Culture and National Culture

We hereinafter include two levels of culture in our review, i.e., corporate culture and national culture (Tellis et al., 2009). On the one hand, to make sense of corporate culture, we start with the definition by Barney (1986), according to whom corporate culture is the "complex set of values, beliefs, assumptions and symbols that define the way in which a firm conducts its business" (p. 657). Extending Schein (1986) contribution and consistent with Homburg and Pflesser (2000), we adopt a holistic view of corporate culture referring to it as a core set of attitudes and practices that define the way in which a firm conducts its business (Barney, 1986), including shared basic values, behavioral norms, artifacts (e.g., stories, arrangements, rituals, and language, Schein 1986), and behaviors (Homburg and Pflesser, 2000). On the other hand, to make sense of national culture, we use the definition by Hostede and Hofstede (1991), according to whom national culture is the "collective programming of the mind which distinguishes the members of one group or category of people from another" (p. 5). Paralleling the reasoning applied above, extending Schein (1986) contribution and consistent with previous work in the area (see e.g., Tellis et al. 2009), we adopt a holistic view of national culture referring to it as a core set of attitudes and practices that are shared by the citizens of a country (Tellis et al., 2009), including shared basic values, behavioral norms, artifacts, and behaviors (Homburg and Pflesser, 2000).

2.3.3 Corporate Culture and National Culture

Building on our earlier considerations, we hereinafter define Culture of Innovation as a culture (corporate or national) (i.e., shared basic values, behavioral norms, artifacts, and behaviors) that fosters relentless (product or process) innovation – or, we add, creativity and entrepreneurship - ensuring that the firm stays constantly at the leading edge of innovation (Govindarajan and Kopalle, 2006; Golder and Tellis, 2001; Tellis et al., 2009).

2.4 Methodology

This section elaborates on our approach to paper retrieval, selection, and classification.

2.4.1 Paper Retrieval and Selection: Manual

We first identified papers to be included in the review using Web of Science. Specifically, we searched for research papers whose titles, abstracts, or keywords (i.e., topic) contained the following two combinations of stem words: innov* AND cultur* or new product* AND cultur*[2]. We used stem words, i.e., innov*, cultur*, and new product* (vs. innovation, culture, and new product, respectively), as this allows us to also retrieve papers using words such as innovativeness, cultural, and new products, among others. Following Rubera and Kirca (2012), Baumgartner and Pieters (2003) and Palich et al. (2000), we searched fifteen well-renowned management, marketing, and international business journals. The journals we searched are: *Academy of Management Journal, Industrial Marketing Management, International Journal of Research in Marketing, Journal of Business Research, Journal of International Business Studies, Journal of Management, Journal of Management Studies, Journal of Marketing, Journal of Marketing Research, Journal of Product Innovation Management, Journal of the Academy of Marketing Science, Management Science, Marketing Science, Organization Science, and Strategic Management Journal.* Such initial search resulted in 568 papers. In order to only focus on relatively recent and impactful papers, we excluded papers that were published before 1996 or that received, on average, no more than five yearly citations since first publication[3]. We averaged the number of citations to partially alleviate the problem that the absolute number

[2]AND is a Boolean operator signifying that both search terms must be present.
[3]We also excluded editorial materials such as introductions to special issues.

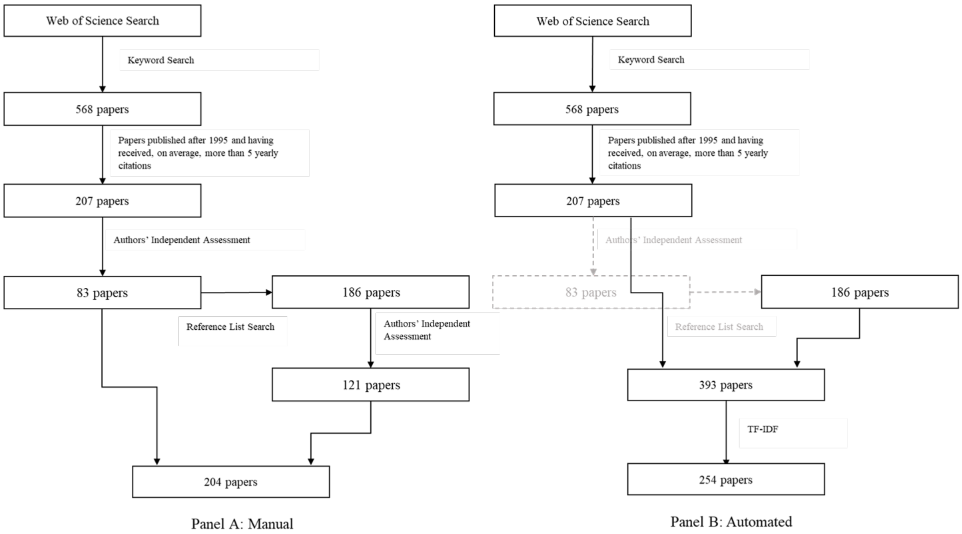


Figure 2.1: Alternative Approaches to Paper Selection: Manual vs. Automated

of citations depends on the paper's age (Biemans et al., 2010). This process resulted in 207 papers (out of the 568 initially retrieved papers). Next, two of the authors independently evaluated the relevance of each paper and its consistency with the topic under investigation (see definitions above). We note here that we adopted a loose-knit criterion to select papers, which matches well our broad definitions of innovation, culture, and culture of innovation. This resulted in the selection of 83 relevant papers. As a third step, we examined the reference sections of all these 83 papers to identify other potentially relevant papers that our keyword search might have missed. This resulted in 186 additional potentially relevant papers, all published in or after 1996 and having received on average more than five yearly citations since their first publication. This was followed by two of the authors again independently evaluating the relevance of each paper and its consistency with the topic under investigation. This evaluation led to the selection of additional 121 papers. Hence, the manual process eventually resulted in a corpus of 204 academic papers (i.e., 83 papers from the initial Web of Science search and 121 papers from the reference list search (a flowchart of our manual approach to paper retrieval and selection is reported in Panel A, Figure 2.1)).

2.4.2 Alternative Approach to Paper Selection: Automated

In order to identify papers to be included in the review, we also adopted an automated approach. At this stage we focused on the 393 papers deemed potentially relevant (i.e., 207 papers from the initial Web of Science search and 186 papers from the reference list search; a flowchart of our automatic approach to paper selection is reported in Panel B, Figure 2.1). In the following paragraphs, we detail the steps involved.

Text Extraction

We first converted our papers, which were in Portable Document Format (i.e., .pdf extension), to images, using the Pdf2Image[4] module for Python. Then, we extracted the text from images using Python Tesseract[5], an Optical Character Recognition (OCR) tool. We treated papers with .pdf extension as images because older papers with .pdf extension are often scans, preventing us from extracting text using a pdf reading tool. At this stage we also separated the core content of papers from their reference lists, as including reference lists might inflate word frequencies and bias our subsequent analyses, as we detail later.

Pre-processing

Removal of "noise" - We removed punctuation, numbers, figures, uppercases (replaced with the corresponding lowercases), and stop words, i.e., words that are extremely frequent in a language but do not convey any specific meaning as, for example, prepositions, conjunctions, and articles. The removal of stop words is a function available in Python for several languages. Here we used the function available in Gensim[6].
Lemmatization and Stemming - We reduced each word to its lemma. This process removes word endings such as "-ing", "-ed", "-s" to reduce each word to its dictionary form

[4]https://pypi.org/project/pdf2image/
[5]https://pypi.org/project/pytesseract/
[6]https://radimrehurek.com/gensim/

(i.e., the form that one could find if they were searching for a word in the dictionary). We then reduced each lemma to its stem. As an example, the word “innovations” becomes “innovation” after lemmatization, and “innov” after stemming. This allowed us to group all variations of the concept of “innovation” (i.e., “innovate”, “innovative”, “innovativeness” were all eventually reduced to “innov”) and evaluate how often the concept occurs in the text. For lemmatization and stemming we used Spacy[7] and NLTK[8], respectively.

Collocation - Since one of the concepts we are interested in, i.e., “new product”, is made up by two words, we collocated words. Collocation is used when two words have a different meaning when they are adjacent compared to the meaning they would have if in isolation. For example, the words “social” and “media” have a different meaning when they are adjacent like in “social media” compared to when used in isolation. In order to treat “new product” as a distinct concept, we identified repeated couples of adjacent words and joined their tokens with “_”, so that “new product” becomes “new_product”. In order to automatically identify couples of words that needed to be collocated we used the BigramCollocationFinder[9] function available in NLTK.

Paper Selection

After pre-processing, we used Term Frequency - Inverse Document Frequency (TF-IDF, hereinafter; Spärck Jones 1972) to select papers relevant to our literature review, out of the collection of 393 potentially relevant papers. TF-IDF is an information retrieval statistic that captures how typical a term is of a certain document given a certain collection of documents. For each document in a collection of n documents, one can compute the TF-IDF score of a certain search term t as follows:

1. $TF(t) = \textit{number of times t occurs in the document}$

[7] https://spacy.io/
[8] https://www.nltk.org/
[9] https://www.nltk.org/

2. $IDF(t) = \ log\ \frac{1+n}{1+DF(t)}\ +1$

3. $TF - IDF(t) = \ TF(t) \times IDF(t)$

where t is a search term, n is the number of documents, and DF(t) is the number of documents among the n documents that contain at least one instance of t (Pedregosa et al., 2011). Given that TF-IDF(t) results from multiplying TF(t) with IDF(t), a higher TF-IDF(t) score can be obtained only if one of the two terms, i.e., TF(t) or IDF(t), increases. The TF(t) component increases when the search term t is very frequent in the focal document while the IDF(t) component increases when the search term is infrequent across other documents in the collection. Hence, the more a search term t is frequent in the focal document and infrequent across the whole collection of documents, i.e., typical of the focal document, the higher the TF-IDF(t) score for the focal document will be.
We used TF-IDF as an alternative approach to paper selection focusing on the 393 papers deemed potentially relevant (see Panel B, Figure 2.1). Specifically, in order to distinguish between relevant and irrelevant papers, we focused on the TF-IDF scores for our search terms, i.e., the stems “innov”, “cultur”, and “new_product”, for each of the above-mentioned potentially relevant 393 papers. We first focused on each search term in isolation and labelled as relevant, according to TF-IDF, those papers whose TF-IDF score for that search term was above or equal to the mean[10] TF-IDF score of the search term in our entire collection of papers. For example, Hurley and Hult’s (1998) TF-IDF score for “innov” is 0.041. Given that this value is above the mean TF-IDF score for “innov” in our collection (0.021), Hurley and Hult (1998) was automatically labelled as relevant when using “innov” as the only search term in TF-IDF. We alternatively considered search terms in pairs and labelled as relevant those papers whose TF-IDF scores were above or equal to the mean TF-IDF score for at least one of them

[10]We relied on means rather than medians as medians could not be used consistently for all the search terms. Specifically, the stem “new_product” is absent from most papers, resulting in a null median tf-idf score. Hence, every paper would have been classified as relevant if we had used the median as the threshold.

(i.e., “innov” >= mean(“innov”) OR[11] “cultur” >= mean (“cultur”); “new_product” >= mean(“new_product”) OR “cultur”>= mean (“cultur”); “innov”>= mean(“innov”) OR “new_product” >= mean(“new_product”)).

Considering our manual classification as the correct classification and comparing the outcomes of TF-IDF against it, we then assessed the predictive ability of the alternative six TF-IDF criteria detailed above. Based on the match/mismatch between the outcomes of the manual vs. automated classification, we computed, for each TF-IDF criterion, the number of: (a) True Positives, i.e., papers classified as relevant according to both the manual classification and the automated (TF-IDF) classification; (b) True Negatives, i.e., papers classified as irrelevant according to both the manual classification and the automated (TF-IDF) classification; (c) False Positives, i.e., papers classified as irrelevant according to the manual classification and relevant according to the automated (TF-IDF) classification; (d) False Negatives, i.e., papers classified as relevant according to the manual classification and irrelevant according to the automated (TF-IDF) classification. Based on the above, we then calculated the sensitivity, specificity, and accuracy of the alternative TF-IDF criteria. We define TF-IDF Sensitivity as its ability to correctly classify relevant papers over the total number of relevant papers (Swift et al., 2020) and compute it as follows:

4. $Sensitivity = \frac{Number\ of\ True\ Positives}{(Number\ of\ True\ Positives + Number\ of\ False\ Negatives)}$

We define TF-IDF Specificity as its ability to correctly classify irrelevant papers over the total number of irrelevant papers (Swift et al., 2020) and compute it as follows:

5. $Specificity = \frac{Number\ of\ True\ Negatives}{(Number\ of\ True\ Negatives + Number\ of\ False\ Positives)}$

Last, we define TF-IDF Accuracy as its ability to correctly classify relevant and irrelevant papers over the total number of papers (Hovy, 2020) and compute it as follows:

6. $Accuracy = \frac{(Number\ of\ True\ Positives + Number\ of\ True\ Negatives)}{Total\ Number}$

[11]OR is a Boolean operator signifying that at least one of the search terms must be present

	Column 1 "innov"[1]	Column 2 "cultur"[1]	Column 3 "new_product"[1]	Column 4 "innov" OR "cultur"[2]	Column 5 "new_product" OR "cultur"[2]	Column 6 "innov" OR "new_product"[2]
Sensitivity	0.716	0.534	0.461	0.858	0.784	**0.868**
Specificity	0.677	0.471	0.836	0.307	0.349	**0.593**
Accuracy	0.697	0.504	0.641	0.593	0.575	**0.735**

[1] A paper is classified as relevant if its TF-IDF score for the search term (e.g., "innov") is equal or above the mean in the collection of papers.
[2] A paper is classified as relevant if its TF-IDF score for at least one of the search terms (e.g., "innov" OR "cultur") is equal or above the mean in the collection of papers.

Table 2.1: Comparison across Alternative TF-IDF Criteria

		Automated Classification	
		Automatic "Relevant"	**Automatic "Irrelevant"**
Manual Classification	**Manual "Relevant"**	True Positives 177 (45%)	False Negatives 27 (6.9%)
	Manual "Irrelevant"	False Positives 77 (19.6%)	True Negatives 112 (28.5%)

Table 2.2: Confusion Matrix

In Table 2.1, we compare sensitivity, specificity, and accuracy of the alternative TF-IDF criteria, considering our manual classification as the correct one (baseline). Having a recurrence equal or above the mean for at least one notion of innovation (i.e., "innov" OR "new_product") is the most accurate criterion (Accuracy = 73.5%) to distinguish relevant and irrelevant papers (see Column 6, Table 2.1). Using this criterion resulted in the selection of 254 papers (see Panel B, Figure 2.1) to be included in the literature review. In Table 2.2, we report the confusion matrix obtained when using the best-performing TF-IDF criterion. Table 2.2 cross-tabulates the manual classification of the 393 potentially relevant papers with the TF-IDF classification and reports the numbers for true positives (177), true negatives (112), false positives (77), and false negatives (27). We note here that, out of 77 false positives, i.e., papers classified as irrelevant according to the manual selection and relevant according to TF-IDF, 72 were manually excluded for reasons other than not being about culture of innovation such as being case studies whose scope was judged as narrow by the authors. We also note that 14 out of 27 false negatives, i.e., papers classified as relevant according to the manual selection and not relevant according to TF-IDF, focus on relevant constructs but using different terms such as, as an example, NPD instead of new product development. Based on the high level of accuracy (73.5%) of the best-performing TF-IDF criterion and on our objective to produce a fully automated, replicable review of the literature, in the next sections we apply LDA to the set of 254 automatically selected papers. We note here that we alternatively applied LDA to the set of 204 manually selected papers. The results, available upon request from the authors, are generally consistent with those obtained using the set of automatically selected papers. In Appendix A, we report on possible alternative approaches to automated paper selection and on why we deem TF-IDF as the most appropriate one for the present study.

2.4.3 Paper Classification: Latent Dirichlet Allocation (LDA)

Pre-processing

We once again pre-processed each paper by removing sources of noise and lemmatizing words. We pre-processed papers again as, at this stage, due to the different purpose of LDA vs. TF-IDF, we did not apply stemming. This results from the fact that, in order to run LDA, one should preferably tag each word with its part of speech (e.g., pronoun, adverb, conjunction, etc.) and keep only words classified as nouns, verbs, adjectives, or adverbs. Stemming words in the pre-processing of text for LDA would have made tagging impossible (e.g., the stem "innov" could have come from the verb "innovate", the noun "innovation", or the adjective "innovative").

Topic Modelling

Topic modelling is a text-mining technique used for discovering an abstract subject from a set of documents (i.e., a corpus). Topic modelling allows to locate hidden topics from a large, unstructured corpus, by clustering words with similar meanings (Griffiths and Steyvers, 2004; Wang et al., 2012). The most used topic modelling technique is Latent Dirichlet Allocation (LDA; Blei et al. 2003). There are several benefits associated with using LDA for literature reviews over other topic modelling techniques. First, LDA does not make assumptions about the structure of the text nor about the syntactical or grammatical properties of the language, being therefore more suitable to extract latent topics from research papers (Tirunillai and Tellis, 2014). The method is also not dependent on assumptions about the underlying distribution of the words, nor it is based on the structure of relationships between words (Tirunillai and Tellis, 2014). Second, LDA uses an unsupervised Bayesian learning algorithm, allowing to adopt a clean slate, assumption-free approach to literature reviews (Tirunillai and Tellis, 2014). Lastly, LDA allows to deal with a large number of documents efficiently (Tirunillai and Tellis, 2014). Formally, LDA represents a collection of documents as a random mixture of latent topics. Given a collection of M documents, formed by N words about K topics, the joint distribution of words and topics within those documents is given by the following equation:

7. $p(w, z, \theta, \phi, \alpha, \beta) = \prod_{i=1}^{K} p(\phi_i | \beta) \prod_{j=1}^{M} p\left(\theta_j \middle| \alpha\right) \prod_{t=1}^{N} p\left(z_{j,t} \middle| \theta_j\right) p\left(w_{j,t} \middle| \phi_{z_{j,t}}\right)$

where w and z are the vectors of all words and topics; θ is a vector of prior topic probabilities that follows a Dirichlet distribution, ϕ is a vector of word probabilities for a given topic that follows a Poisson distribution, and α and β are the hyperparameters on θ and ϕ, respectively.

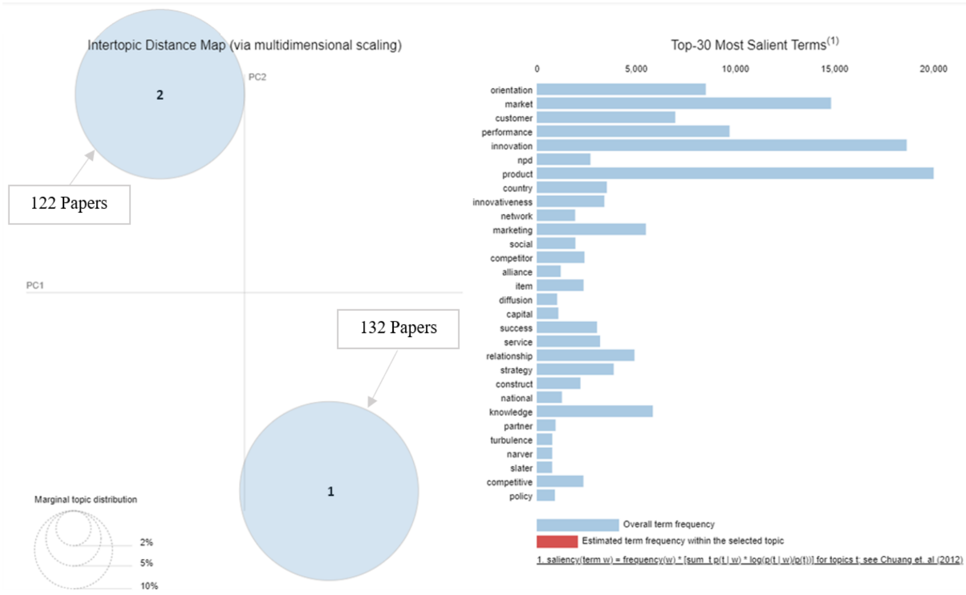
LDA has been recently used in management and marketing research (see e.g., Büschken and Allenby 2016, 2020; Tirunillai and Tellis 2014; Toubia et al. 2019; Zhong and Schweidel 2020). Such LDA applications, however, especially within the marketing field, focus on topic modelling of reviews or other user-generated content (Tirunillai and Tellis, 2014; Toubia et al., 2019). Here we propose, to the best of our knowledge, the first attempt at using LDA to inform a literature review process in the management, marketing, and international business literatures.

LDA Implementation

As evident from equation (7), LDA requires the number of topics as an input. However, the optimal number of topics is generally unknown before running the LDA analysis. For this reason, after pre-processing the text, we ran several LDA analyses setting alternative numbers of topics (between 2 and 20) and eventually selected the best-performing number of topics to represent our corpus (i.e., the entire set of our papers). A common practice to evaluate alternative models' performance in machine learning is to divide the sample data in 2 sets, i.e., training data and testing data. The training data is used to fit the model and the parameters obtained in the fitting stage are used to predict the output of the testing data (Gareth et al., 2013). To evaluate the performance of alternative numbers of topics, we therefore examined the log-likelihood and perplexity (i.e., normalized log-likelihood) of the models on testing data using 10-fold cross-validation (i.e., after dividing the sample in 10 random groups, we iteratively took each group as testing data and the remaining 9 as training data). In this context, log-likelihood and perplexity essentially capture how probable are unseen papers (i.e., testing data) given what the model learned from the training data. Results show that the best number of topics given our corpus of research papers is 2. Hence, each paper in our corpus was allocated by LDA to one of the two topics. Figure 2.2 describes the two identified topics. Panel A, Figure 2.2, represents each topic as a bubble. The larger the bubble, the higher the number of papers allocated to that topic. The further the bubbles are away from each other, the more different the topics they represent. Panel A, Figure 2.2 shows that the two identified topics are substantially equivalent in size (132 papers vs. 122 papers, respectively) and clearly distinct.

LDA also automatically returns a list of terms (Panel B, Figure 2.2) to represent the corpus of papers. By default, LDA returns the 30 most salient terms, ranked in descending order of saliency. As an example, the 5 most salient terms in our corpus are, in descending order, "orientation", "market", "customer", "performance", and "innovation". The blue bars in Panel B, Figure 2.2, represent overall term frequencies, which are used to compute term saliency, as we explain next.

Saliency - Panel B, Figure 2.2, shows the most salient terms in our corpus of papers. The saliency of terms is determined based on their (a) probability of occurrence in the



Panel A: Topics Panel B: Most Salient Terms in the Corpus

Figure 2.2: Topics and Most Salient Terms across the Entire Corpus of Automatically Selected Papers. Panel A represents each topic as a bubble. The larger the bubble, the higher the number of papers allocated to that topic. The further the bubbles are away from each other, the more different the topics they represent. Panel B represents the most salient terms within our corpus. The blue bars represent overall term frequencies.

whole corpus and (2) distinctiveness, as follows:

8. $saliency(w) = p(w) \times distinctiveness(w)$

where p(w) is the observed probability of term w in the corpus of papers (i.e., the frequency of term w over the total number of words in the corpus) and distinctiveness(w) is its distinctiveness. Following Chuang et al. (2012), each term can be considered more or less distinctive depending on how informative it is for determining a certain topic, compared to any other randomly selected term. Distinctiveness(w) is mathematically defined as the Kullback-Leibler divergence (Kullback and Leibler 1951) between the likelihood that the observed term w was generated by topic t (i.e., p(t|w)) and the likelihood that any randomly selected term w' was generated by topic t (i.e., p(t)) as follows:

9. $distinctiveness(w) = \sum_t p(t|w) log \frac{p(t|w)}{p(t)}$

Saliency provides us with terms that are potentially relevant to describe the research topics covered in the papers in our corpus. However, salient terms, which are returned automatically by LDA, are more informative about the general topic of culture of innovation than about distinct topics within it (Sievert and Shirley, 2014). To understand the content of the individual topics, LDA returns the most frequent terms per topic. However, as all our papers are within the same field, target the same audience, and are written by authors with similar backgrounds, this may increase the probability that multiple topics have frequent words in common. Since our aim is to use LDA to distinguish different streams of research, hereinafter we use *relevance* (Sievert and Shirley, 2014) to identify the most representative terms in each topic and describe it.

Relevance - Relevance (Sievert and Shirley, 2014) is computed as follows:

10. $Relevance(w,\ t \mid \lambda) = \lambda\ log(\phi_{tw})\ +\ (1\ -\ \lambda)\ log\left(\frac{\phi_{tw}}{p_w}\right)$

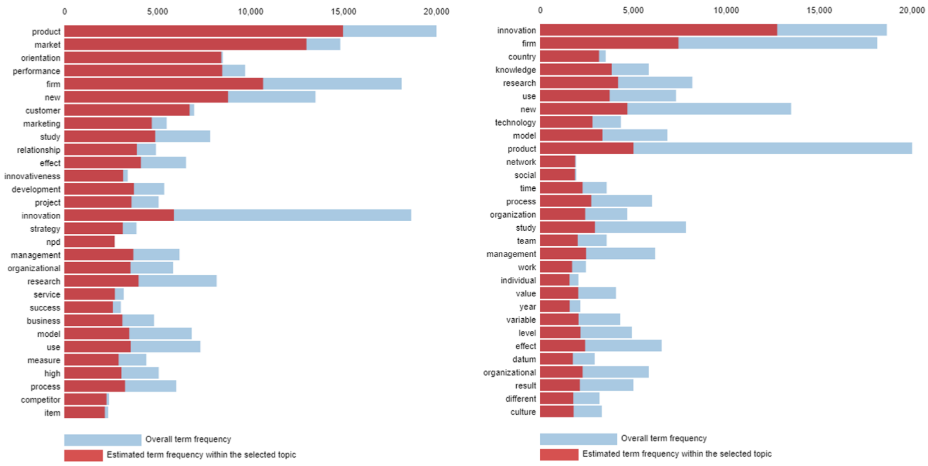
where ϕ_{tw} is the probability of term $w \in \{1, ..., V\}$ for topic $t \in \{1, ..., T\}$, p_w is the marginal probability of term w in the set of documents, and λ is a weight parameter ranging between 0 and 1 given to the probability of term w under topic t relative to its lift (i.e., the ratio of the probability of term w within topic t over its marginal probability across the corpus). Sievert and Shirley (2014) find the optimal value of λ to be around 0.6, hence we set λ to 0.6 when evaluating term relevance within a certain topic. We also note here that setting λ =1 causes frequency by topic and relevance to overlap.

2.4.4 Results

As mentioned above, Panel A, Figure 2.2, shows that the two identified topics are substantially equivalent in size (132 papers vs. 122 papers, respectively) and clearly distinct. The 5 most salient terms in our corpus are, in descending order of saliency, “orientation”, “market”, “customer”, “performance”, and “innovation” (Figure 2.2, Panel B).

Figure 2.3 display the 30 most relevant terms in Topic 1 (Panel A, Figure 2.3) and Topic

2 (Panel B, Figure 2.3), ranked in descending order of relevance. The figure also reports, for each term, its frequency in the topic (i.e., red bar) and its overall frequency across the whole corpus (i.e., blue bar). The interpretation of the results from LDA was performed



Panel A: Topic 1

Panel B: Topic 2

Figure 2.3: Most Relevant Terms in Topic 1 and Topic 2

by the authors starting from the top relevant terms assigned to each topic, and after careful consideration of the papers, until a consensus was reached.

The 10 most relevant terms assigned to the first topic, Topic 1 (132 papers), are “product”, “market”, “orientation”, “performance”, “firm”, “new”, “customer”, “marketing”, “study”, and “relationship”, based on which it was possible to provide the topic with the label “Market Orientation and Innovation Performance”, driven primarily by terms such as “market”, “orientation”, “customer”, “marketing”, and “relationship” (Figure 2.3, Panel A). The 10 most relevant terms assigned to the second topic, Topic 2 (122 papers), are “innovation”, “firm”, “country”, “knowledge”, “research”, “use”, “new”, “technology”, “model”, and “product”, based on which it was possible to provide the topic with the label “Cultural Advantages and Innovation Performance”, driven primarily by terms such as “firm”, “country”, “knowledge”, “research”, and “technology” (Figure 2.3, Panel B).

We subsequently re-ran LDA to identify sub-topics within each of the above-mentioned main topics. This resulted in two sub-topics being identified for each main topic. Figures

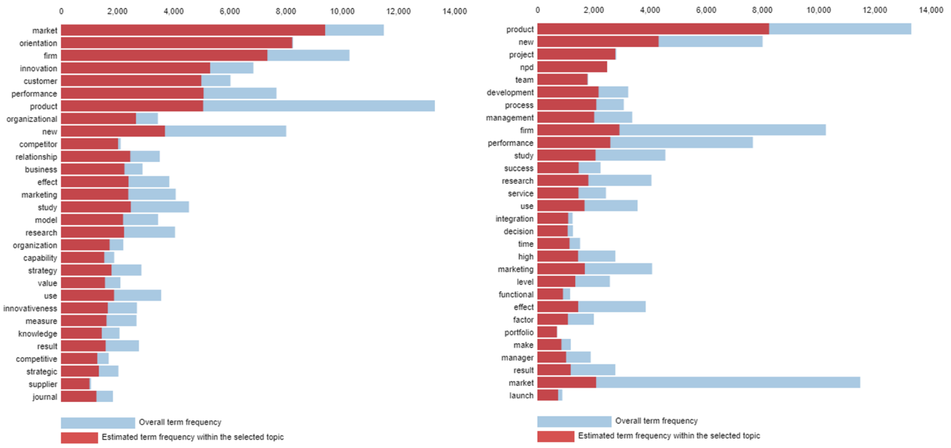
2.4 to 2.6 report the identified sub-topics and their most relevant terms. Specifically, we identified two sub-topics within Topic 1, “Market Orientation and Innovation Performance”, i.e., Sub-topic 1.1, “Customer and Competitor Orientation” (83 papers), driven by terms such as “market”, “orientation”, “customer”, and “competitor”, and Sub-topic 1.2, “Inter-functional Coordination and Cultural Cohesiveness” (49 papers), driven by terms such a “project”, “npd”, “team”, “development”, “process”, “research” and “integration”. Figure 2.4, Panel A, describes the two sub-topics, i.e., subtopic 1.1 and sub-topic 1.2, which appear to be clearly distinct. The most relevant terms for sub-topic 1.1 and sub-topic 1.2 are shown in Figure 2.5, Panels A and B, respectively.

Further, we identified two sub-topics within Topic 2, “Cultural Advantages and Innovation Performance”, i.e., Sub-topic 2.1, “Corporate Culture Advantages” (64 papers), driven by terms such as “innovation”, “knowledge”, “organization”, “capability”, and “resource”, and Sub-topic 2.2, “National Culture Advantages” (58 papers), driven by terms such as “firm”, “product”, “innovation”, “country”, “national”, and “diffusion”. Figure 2.4, Panel B, describes the two sub-topics, i.e., subtopic 2.1 and sub-topic 2.2, which,

Panel A: Topic 1: Sub-topic 1.1 and Sub-topic 1.2

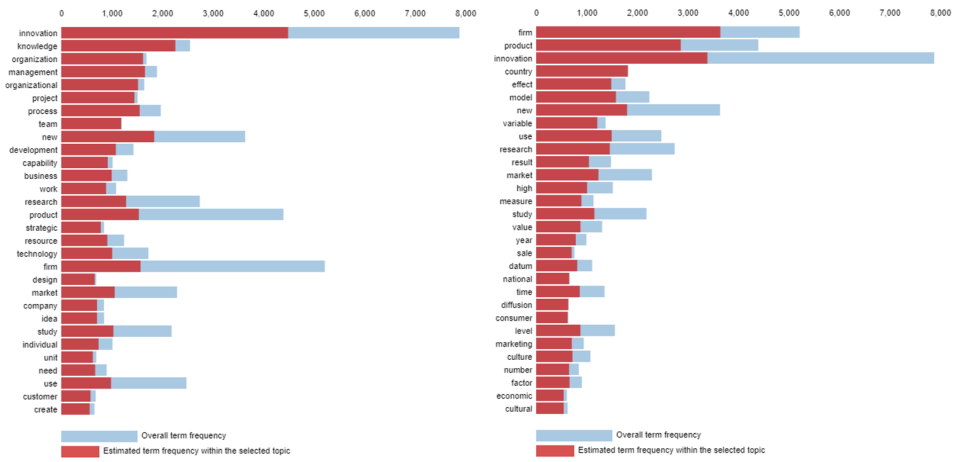
Panel B: Topic 2: Sub-topic 2.1 and Sub-topic 2.2

Figure 2.4: Sub-topics within Topic 1 and Topic 2



Panel A: Sub-topic 1.1 Panel B: Sub-topic 1.2

Figure 2.5: Most Relevant Terms in Sub-topic 1.1 and Sub-topic 1.2



Panel A: Sub-topic 2.1 Panel B: Sub-topic 2.2

Figure 2.6: Most Relevant Terms in Sub-topic 2.1 and Sub-topic 2.2

once again, appear to be clearly distinct. The most relevant terms for sub-topic 2.1 and sub-topic 2.2 are shown in Figure 2.6, Panels A and B, respectively.

Panel A: Relevant terms per Topic/Sub-Topic

Term	Column 1 **Topic 1**	Column 2 **Sub-Topic 1.1**	Column 3 **Sub-Topic 1.2**	Column 4 **Topic 2**	Column 5 **Sub-Topic 2.1**	Column 6 **Sub-Topic 2.2**
1	product	market	product	innovation	innovation	firm
2	market	orientation	new	firm	knowledge	product
3	orientation	firm	project	country	organization	innovation
4	performance	innovation	npd	knowledge	management	country
5	firm	customer	team	research	organizational	effect
6	new	performance	development	use	project	model
7	customer	product	process	new	process	new
8	marketing	organizational	management	technology	team	variable
9	study	new	firm	model	new	use
10	relationship	competitor	performance	product	development	research
Number of Papers (N)	132	83	49	122	64	58

Panel B: Papers per Year by Sub-Topic

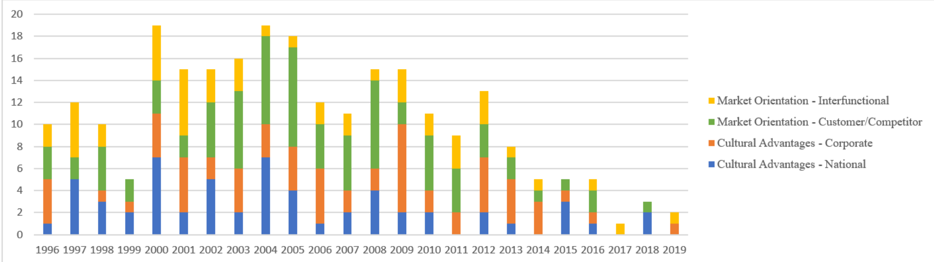


Figure 2.7: LDA: Summary of Results

Panel A, Figure 2.7, reports a summary of the results from our LDA analysis including, for each topic/sub-topic, the 10 most relevant terms and the number of assigned papers. We also report research trends over time in Panel B, Figure 2.7. In Appendix B, we report the full list of papers allocated to each topic/sub-topic in the manuscript.

2.5 Discussion

In the following sections, we discuss the topics and sub-topics identified with LDA more in detail. An overview of topics and sub-topics is also provided in Figure 2.8.

2.5.1 Topic 1: Market Orientation and Innovation Performance

Driven primarily by terms such as "market", "orientation", "customer", "marketing", and "relationship" (Figure 2.3, Panel A), and after careful examination of the assigned papers, we labelled Topic 1 as "Market Orientation and Innovation Performance". According to the seminal paper by Narver and Slater (1990), customer orientation, competitor orientation, and inter-functional coordination are the key tenets of a firm's market orientation, i.e., “the organizational culture that most effectively and efficiently creates the necessary behaviors for the creation of superior value for buyers and, thus, continuous superior performance for the business” (p. 21). On the one hand, “Customer Orientation is the sufficient understanding of one’s target buyers to be able to create superior value for them continuously” (p. 21) while Competitor Orientation “means that a seller understands the short-term strengths and weaknesses and long-term capabilities of both the key current and the key potential competitors” (p. 22). Customer and competitor orientation provide firms with a superior knowledge of customers and competitors, respectively, which facilitates the development and launch of innovation (Li and Calantone, 1998). On the other hand, inter-functional coordination is “the coordinated utilization of company resources in creating superior value for target customers” (p. 22) and requires coordinated efforts of the various groups and sub-groups within the company. According to De Luca et al. (2010), “Inter-functional coordination includes open communication between scientists and business development personnel, sharing of projects’ goals and responsibilities, and common perspectives on innovation priorities by different departments” (p. 308). Such close cooperation and communication among different functions has also been identified in the NPD literature as an important antecedent to success (see e.g., Atuahene-Gima 1996). Examining the seminal papers in this area, Deshpandé et al. (1993) conceptualize market orientation as an aspect of corporate culture which provides “norms for behaviour” (Day 1994, p. 43), i.e., underlying beliefs and values, while Deshpandé and Farley (1998) and Kohli and Jaworski (1990) treat it as a certain set of behaviors and/or processes. Research in this area has primarily focused on the consequences of market orientation for innovativeness and new product performance Kirca et al. 2005. Related to this, one important debate is whether a market orientation fosters innovation or rather causes incremental improvements. Despite the debate going on for decades (Atuahene-Gima, 1996), an examination of papers allocated to Topic 1 seems to confirm the positive relationship between market orientation and innovation (see e.g., Baker and Sinkula 1999; Zhou et al.

2005). We discuss sub-topic 1.1 and sub-topic 1.2 in more detail in the following sections.

2.5.2 Sub-topic 1.1: Customer and Competitor Orientation

Driven primarily by terms such as “market”, “orientation”, “customer”, and “competitor” (Figure 2.5, Panel A), and after careful examination of the assigned papers, we labelled Sub-topic 1.1 as “Customer and Competitor Orientation”. The journals in our sample that have published the most in this area are Journal of Product Innovation Management (18 papers), Journal of Business Research (15), and Journal of Marketing (12). Upon examination of the 83 papers allocated to sub-topic 1.1, we identified two streams of research focusing on (1) customer and competitor orientation as a set of behaviors and/or processes and (2) customer and competitor orientation as a set of underlying beliefs and values, respectively.

Most papers allocated to sub-topic 1.1 treat customer and competitor orientation as a set of behaviors and/or processes, consistent with Deshpandé and Farley (1998) and Kohli and Jaworski (1990). As an example, Frambach et al. (2003) adopt a behavioural view of customer and competitor orientation, measured using scales based on previous studies (Kohli et al., 1993; Narver and Slater, 1990), as the extent to which the firm engages in behaviors related to understanding and responding to customers and competitors, respectively, and find negative effects for competitor orientation and positive effects for customer orientation on new product development and introduction. In the same stream,Harmancioglu et al. (2010) find that a firm’s market information collection effort, a key behavioral manifestation of customer and competitor orientation (Kohli and Jaworski, 1990), has a positive effect on innovativeness. Most of the papers in this stream use scales, such as the MARKOR (Kohli et al., 1993) and MKTOR (Narver and Slater, 1990) scales, to operationalize customer and competitor orientation. In the second, and narrower, stream of research, Noble et al. (2002), who also look at inter-functional coordination, show that firms with higher levels of competitor orientation, measured starting from text analysis of statements in letters to shareholders, exhibit superior performance via increased innovation. Overall, both streams of research generally treat customer and competitor orientation as antecedents to innovation performance.

2.5.3 Sub-topic 1.2: Inter-functional Coordination and Cultural Cohesiveness

Driven primarily by terms such as “project”, “npd”, “team”, “development”, “process”, “research”, and “integration” (Figure 2.5, Panel B), and after careful examination of the assigned papers, we labelled Sub-topic 1.2 as “Inter-functional Coordination and Cultural Cohesiveness”. The journals in our sample that have published the most in this area are Journal of Product Innovation Management (27 papers), Journal of Marketing Research (15) and Industrial Marketing Management (3). Upon examination of the 49 papers allocated to sub-topic 1.2, we identified two streams of research focusing on (1) inter-functional coordination and (2) cultural cohesiveness, respectively.

Within the first stream, integration of R&D with marketing is a frequent topic (Calantone and Rubera, 2012; Danneels, 2002). This is consistent with the view that market competence, which resides within the marketing department, and technology competence, which resides within the R& D department (Danneels, 2002), are two key sources of new product success. In this stream, Ayers et al. (1997) find that integration between marketing and R&D, measured by marketing involvement, engineering involvement, and information exchange, affects new product success. Within the second stream, Brockman and Morgan (2006) find that cultural cohesiveness, an organizational mindset, has a moderating influence on both an organization’s use of its existing knowledge to develop innovative new products and the resulting performance of those products, while Sethi et al. (2001) find a negative effect of social cohesiveness on level of innovation and performance. Overall, both streams of research generally treat inter-functional coordination and cultural cohesiveness as antecedents to innovation performance.

2.5.4 Topic 2: Cultural Advantages and Innovation Performance

Driven primarily by terms such as “firm”, “country”, “knowledge”, “research”, and “technology (Figure 2.3, Panel B), and after careful examination of the assigned papers, we labelled Topic 2 as “Cultural Advantages and Innovation Performance”.
According to Tellis et al. (2009), corporate culture and national culture, which include both beliefs and values and behaviors and processes, are key drivers of (radical) innovation. While corporate culture can be used to explain differences in innovation outcomes across firms in a country, national culture can be used to explain differences in innovation outcomes across countries. The dichotomy between corporate and national culture is also consistent with seminal research on firm-specific vs. country-specific advantages in the international business domain (Rugman, 1981). Extending developments in the literature on firm-specific vs. country-specific advantages (Rugman, 1981), we define Corporate Culture Advantages as those unique cultural resources and capabilities as well as cultural resources’ and capabilities’ combinations of a firm that positively affect a firm’s innovative outcomes. Further, we define National Culture Advantages as the unique cultural factors in a country that positively affect the innovative outcomes of the country as well as the innovative outcomes of the firms based or operating in it. While corporate culture advantages refer to a firm’s cultural competitive strength, e.g., advantages in upstream (R&D) and/or downstream (marketing/customization) capabilities, national culture advantages refer primarily to exogenous country-level cultural advantages that ultimately influence firm strategy, structure, and performance (Rugman, 1981; Rugman et al., 1985). We discuss sub-topic 2.1 and sub-topic 2.2 in more detail in the following sections.

2.5.5 Sub-topic 2.1: Corporate Culture Advantages

Driven primarily by terms such as “innovation”, “knowledge”, “organization”, “capability”, and “resource” (Figure 2.6, Panel A), and after careful examination of the assigned papers, we labelled Sub-topic 2.1 as “Corporate Culture Advantages”. The journals in our sample that have published the most in this area are Journal of Product Innovation Management (17 papers), Organization Science (8), and Strategic Management Journal (12). Upon examination of the 64 papers allocated to sub-topic 2.1, we identified three streams of research focusing on (1) leadership, (2) individuals and teams, and (3) general approach to innovation and innovation-related activities, respectively.

The first stream of research focuses on how leaders in a firm can drive innovative outcomes. Numerous studies have indicated that managerial-level factors, including leadership, represent the main antecedents of corporate innovation (Damanpour, 1991). As an example, Elenkov et al. (2005) investigate the effect of strategic leadership behaviors on executive influence on process and product innovation while Gumusluoğlu and Ilsev (2009) focus on how transformational leadership affects creativity both at the individual-level and the organizational-level. The second stream of research in this area focuses on how sub-units (e.g., individuals, teams) in a firm can drive innovative outcomes. As an example, Bharadwaj and Menon (2000) focus on individual- and organization-level creativity and show that the presence of both individual and organizational creativity mechanisms leads to the highest level of innovation performance (vs. individual creativity only and organizational creativity only). In the same stream and again focusing on individuals, Griffiths-Hemans and Grover (2006) look at individual (and organizational) factors that can boost the creation and harnessing of worthwhile ideas in firms, while Vera and Crossan (2005) show that improvisation has a positive effect on team-level innovation. Last, the third stream of research in this area focuses on the firm’s more general approach to innovation and innovation-related activities. Papers in this stream are heterogeneous in nature and focus on how the firm's approach to innovation can drive innovative outcomes. In this stream, Carmona-Lavado et al. (2010) focus on a firm's social capital while Leal-Rodríguez et al. (2014) focus on absorptive capacity. Overall, all the streams of research generally treat corporate culture advantages as antecedents to innovation performance.

2.5.6 Sub-topic 2.2: National Culture Advantages

Driven primarily by terms such as "firm", "product", "innovation", "country", "national", and "diffusion" (Figure 2.6, Panel B), and after careful examination of the assigned papers, we labelled Sub-topic 2.2 as "National Culture Advantages". The journals in our sample that have published the most in this area are Marketing Science (10 papers), Journal of Marketing (8), Academy of Management Journal (6), and Journal of Business Research (6). Upon examination of the 64 papers allocated to sub-topic 2.2, we identified two streams of research focusing on (1) innovation adoption and diffusion and (2) national culture and firm strategies, respectively.

The first stream of research in this area focuses on how consumers' cultural characteristics in a country may drive innovation adoption and diffusion, with clear implication for companies. As an example,Tellis et al. (2003) look at how, among others, high activity rate of women, economic openness, uncertainty avoidance, masculinity, and need for achievement affect the take-off of new products. On a similar note, Sundqvist et al. (2005) investigate the roles of the Hofstede's (1991) values of power distance, uncertainty avoidance, and individualism on innovation adoption in a country. The second stream of research in this area looks at how a country's culture may drive an organization strategy. In this area, Steensma et al. (2000) look at the effect that national culture (i.e., uncertainty avoidance, cooperative values) has on the propensity for small, independent manufacturing enterprises to cooperate with other firms for technological innovation. Luk et al. (2008) investigate how Guanxi with government officials positively affects administrative and product-related innovativeness in a transition economy but not in a market

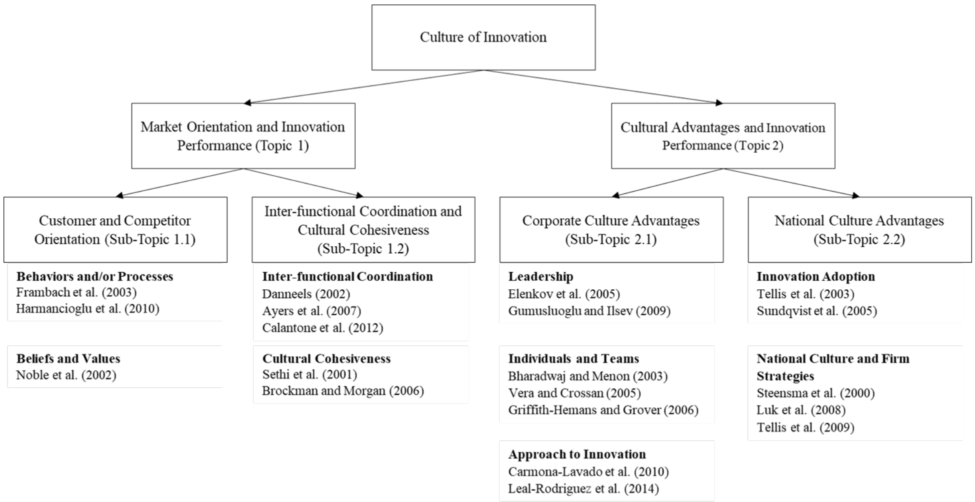


Figure 2.8: Research Streams in Culture of Innovation

economy. Overall, most of the research in this area has used Hofstede's (1991) values to measure national culture. We consider this to be a limitation, as we detail later. Overall, both streams of research generally treat corporate culture advantages as antecedents to innovation performance. We report an overview of the various research streams in Figure 2.8.

2.6 Towards an Agenda on Culture of Innovation

We now identify directions for further research in the culture of innovation field.

Shedding Light on Process Innovation - Research on innovation, in general, and culture of innovation, in particular, have primarily focused on product innovation, neglecting process innovation. Firms, nonetheless, can also innovate through process innovation (e.g., Elenkov et al. 2005), which can, among others, be instrumental to product innovation (Cillo et al., 2018). Future researchers should try to investigate whether and how some of the established cultural drivers of innovation in the product innovation domain can also drive process innovation. This is particularly urgent as the relevance of process innovation is dramatically increasing over time due, among others, to the growing phenomenon of servitization (Visnjic et al., 2016). Further, due to the continuous acceleration in the pace of technological developments, in addition to managing existing products, firms must innovate in technology and processes to outpace global competition (Westkämper and Walter, 2014).

Leveraging Alternative Sources of Data to Operationalize Market Orientation - As mentioned above, most papers focusing on market orientation in the field of culture of innovation are surveys using scales such as the MARKOR or MKTOR scales in their operationalizations. However, as Berger et al. (2020) state: "A primary challenge [...] is to obtain reliable and generalizable survey or field data about factors that lie deep in the firm's culture and structure or that are housed in the mental models and beliefs of marketing leaders and employees" (p. 7). Future research should therefore consider leveraging alternative sources of data to operationalize market orientation. Related to this, text analysis offers an objective and systematic solution to assess constructs in naturally occurring data (e.g., letters to shareholders, press releases, patent text, marketing messages, conference calls with analysts) that may be more valid. Specifically, text-analyses of 10-Ks and letters to shareholders available from SEC EDGAR (see e.g.,Noble et al. 2002; Yadav et al. 2007), press releases, social media posts, patent text, and even conference calls transcripts available from Capital IQ or FactSet could be useful.

Broadening the Scope of Studies on Market Orientation - Most of the research we reviewed in sub-topic 1.1., i.e., "Customer and Competitor Orientation", focused on customers and competitors as key stakeholders, therefore neglecting other potentially relevant constituencies. Future research should adopt a stakeholder perspective (Ferrell et al., 2010), investigating whether and how the firm's cultural focus on a broader set of stakeholders, including suppliers, employees, regulators, shareholders, and the local community (Greenley and Foxall, 1997), may drive innovative outcomes. Further, most research we reviewed in sub-topic 1.2, i.e., "Inter-functional Coordination and Cultural Cohesiveness", focused on the R&D-Marketing interface (see Olson et al. 2001 and Tatikonda and Montoya-Weiss 2001 for two exceptions). Future research should look more closely at other functions, i.e., sales, operations, and finance, and at how their integration with marketing and R&D could benefit/hinder innovative outcomes.

Exploring Innovation as an Antecedent to Corporate Culture - Innovation and culture are intricately intertwined. While, as shown in this review, a burgeoning corpus of research has studied how corporate culture drives innovation, innovation may also affect corporate culture, that is, while corporate culture inspires innovation, innovation outcomes may also inspire cultural change within a firm (Galambos, 1997). Related to this, as an example, focusing on how failures in innovation drive a firm's culture would represent a substantial extension to the literature, answering calls for more research "into the mechanisms by which organizations deal with and learn from failure" (Madsen and Desai 2010, p. 472). More in general, future researchers should try to unveil whether and how a firm's past innovative efforts and outcomes, either positive or negative, may affect its corporate culture, with a particular focus, we argue, on shared basic values, behavioral norms, and artifacts.

Challenging Conventions in the Study of National Culture - Most of the papers we reviewed in sub-topic 2.2., i.e., "National Culture Advantages", suffer from many of the limitations traditionally associated with conventional interpretations and operationalizations of national culture. First, national culture has thus far been treated in the literature on culture of innovation as a stable, immutable characteristic. Future research should consider whether and how national culture may, although slowly, evolve over time (Shenkar et al., 2020). Second, while most research on national culture has used Hofstede's (1991) measures, future research should use, in the absence of time-varying measures for national culture, alternatives to Hofstede's values such as Schwartz' value survey (Schwartz, 1994), Inglehart and Associates' World Value Survey (Inglehart, 1997), Thompenaars and Hampden-Turner (1998), Global Leadership and Organizational Behavior Effectiveness (GLOBE) project's cultural dimensions (Steenkamp and Geyskens, 2014) and cultural tightness and looseness (Gelfand et al., 2011). Finally, related to the previous point, national culture enters the literature on culture of innovation primarily as a set of underlying values and beliefs. Future research should consider cultural artifacts (Schein, 1986) stemming from national cultural values (e.g., local infrastructure, R&D spending, private/public interactions, etc.) and whether and how they drive innovative outcomes. Such artifacts have been studied in other fields (see e.g., Archibugi and Coco 2005), but surprisingly neglected by the fields covered in this review.

Investigating Positive Societal Outcomes of a Culture of Innovation - Finally, the literature on culture of innovation we reviewed looks at a broad range of proximal (e.g., number of new products introduced, time to new product launch, conflict in NPD) and ultimate (e.g., sales, market share, ROA, Tobin's Q, stock market responses, etc.) innovation-related outcomes. We argue that the literature on culture of innovation would benefit from focusing on a broader and not purely economic set of outcomes and, in particular, on the "positive externalities and win-win situations in which innovation can benefit both commercial and non-commercial stakeholders" (Moorman, 2018). While innovation has been shown to improve standards of living over time (Wilkie and Moore, 1999), we see an opportunity to study culture of innovation in emerging areas such as frugal innovation (Radjou et al., 2015), social innovation aimed at improving the livelihoods of vulnerable, impoverished, and subsistence marketplace consumers (Lee et al., 2019; Vassallo et al., 2019), and environmentally-sustainable innovation (Olson, 2013).

2.7 Contributions

This research makes at least four contributions to the literature. First, from a theoretical perspective, we conduct a literature review on culture of innovation within the broader fields of management, marketing, and international business. To the best of our knowledge, we are the first to provide almost two and half decades of research on culture of innovation with a structure. Given the constantly evolving nature of this research field across various disciplines and its broad scope, we expect this study to serve as a fundamental stepping-stone for readers interested in culture of innovation and to act as a "one stop shop" to provide them with a glimpse into the culture of innovation literature. Second, again from a theoretical perspective, the extant literature on culture of innovation is fragmented, as research on this topic has proceeded in parallel in many academic fields, namely management, marketing, and international business, with little theoretical and empirical integration (Rubera and Kirca, 2012). The lack of integration across these research domains makes it difficult for researchers to foresee future studies on culture of innovation. In this study, in addition to integrating contributions from several domains, we provide an extensive future research agenda to further advance the body of scholarship on culture of innovation. In so doing, we lay the groundwork for the next generation of research in the field, by identifying areas in which further research on culture of innovation needs to be pursued. Third, from a methodological perspective, we are the first to bring to innovation scholarship the approach of conducting a literature review using NLP. In this study, NLP allowed us to identify the key topics and sub-topics in the literature on culture of innovation in an objective and efficient manner. Finally, again from a methodological perspective, this study offered a step-by-step tutorial detailing how NLP can be employed in a literature review. In the TF-IDF section, as an example, we provide definitions and Python tools that can be used to pre-process the text and explain how to use this measure to select relevant papers. The LDA section offers numerous suggestions about, among others, the pre-processing steps to be applied and the measures that can be used to better interpret results (e.g., saliency, relevance). We offered this with the aim of making it possible for future researchers to replicate our steps when conducting their own literature reviews, with clear benefits in terms of objectivity, replicability, and efficiency.

2.8 Limitations

This article has several limitations, some of which represent opportunities for further research. First, our study uses LDA on the set of papers judged relevant by TF-IDF. This was justified by our decision to adopt a fully automated and replicable approach, and it also allows us to cast a wide net and include a vast number of papers. In so doing, inadvertently we may have induced some biases in our paper selection, which may have led to the inclusion of a small number of only marginally relevant papers and to the exclusion of a small number of relevant papers. While results (available upon request from the authors) obtained using the set of manually selected papers were generally consistent with those reported in this paper, this should be borne in mind when interpreting our findings. Second, given the wealth of research across different fields, our analysis of the various sub-topics identified by LDA privileged breadth over depth. While this is a by-product of our approach, other narrower and therefore deeper reviews (see e.g., Büschgens et al. 2013; Eisend et al. 2016) could be used to complement our findings. Finally, we opted for a systematic literature review, over a meta-analytic review, in order to be able to adopt a clean slate, assumption-free approach, which was more consistent with our objectives. Nonetheless, future research adopting a meta-analytic approach would be a valuable extension to the present work. In sum, we view this study as a useful step in promoting a better understanding of culture of innovation and hope it stimulates further research in the area.

A Alternative Automated Approaches to Paper Selection

Two alternative NLP approaches might be used to predict whether a certain paper is relevant or not: (1) training a classifier and (2) looking at semantic similarity. Here we explain how these processes work and why, in our opinion, TF-IDF was the best approach in this case.

Training a classifier

This approach entails selecting a set of papers' features and using them to predict relevance leveraging a classification algorithm ranging from logistic regression to Support Vector Machine or KNN. The algorithm would have been then trained on, as an example, 80% percent of papers in a manually prelabelled (relevant vs. not) corpus and tested on the remaining 20%. The weights of the best-performing algorithm would have been stored and eventually reapplied to a larger corpus of unlabelled papers. This makes the whole process of conducting the literature review supervised, thereby undermining our aim of demonstrating the process of conducting unsupervised literature review. This approach is inconsistent with our aim in two ways. First, this approach is feasible only after the researcher has manually label enough papers to train a reliable classifier. On the other hand, TF-IDF does not require any initial labelling, therefore serving our goal of using a fully automated approach. Though we labelled our papers to have a benchmark to evaluate the performance of TF-IDF, it would have been able to generate a sample of relevant papers even without our annotation. Second, in this approach, features' selection would be arbitrary. Any information about the document could be used as a feature (e.g., number of citations, number of words, authors' location), relying too much on the researcher's own and potentially biased judgment. On the other hand, TF-IDF allows us to be more consistent with our aim of offering an objective and unsupervised literature review process by determining relevance based on an objective, quantifiable measure.

	Requires labelling	Subjective features	Confirmatory selection
Classifier	Yes	Yes	Potentially yes, if trained on a biased sample
Semantic Similarity	No	No	Yes
TF-IDF	**No**	**No**	**No**

Table 3: Alternative Automated Approaches to Paper Selection

Semantic Similarity

The second approach suggested by the NLP literature to identify relevant papers is looking at the semantic similarity between two documents, i.e., the similarity of their meanings. To implement this approach, we could prepare a list of papers that, as expert in the field, we already know are relevant. Then we could use these papers as benchmark for new papers and automatically measure how similar are the new papers to the original set. Finally retain only papers that are similar enough to the benchmark papers. Although this approach overcomes the pre-labelling and features' selection problems posed by a classifier, it would require starting from one or more pre-selected relevant papers. This approach might limit our capacity to detect unknown streams of research. In other words, this approach would judge relevance only in a confirmatory fashion, limiting our ability to include in our sample streams of research we might not be aware of at the beginning of the literature review process. Again, using TF-IDF overcomes this problem because it allows us to judge relevance only based on the core concepts of our literature review, without the risk that a specific stream of research within the domain we are interested in biases our sample of papers.

Table 3 summarizes the pros and cons of each of the above-mentioned alternative approaches.

B Papers classified with LDA

The following table provides the full list of paper included, their Topic probabilities computed with LDA, and the topic they were assigned to.

Table 4: Studies per Topic

Beginning of Table 4

Study	**p(1)**	**p(2)**	**p(1.1)**	**p(1.2)**	**p(2.1)**	**p(2.2)**	**Topic**	**Sub-Topic**	**Topic Label**
Hurley and Hult (1998)	0.77	0.23	0.84	0.16			1	1.1	orientation - competitor/customer
Han et al. (1998)	0.93	0.07	0.99	0.01			1	1.1	orientation - competitor/customer
Knight and Cavusgil (2004)	0.8	0.2	0.61	0.39			1	1.1	orientation - competitor/customer
Cooke et al. (1997)	0	1			0.4	0.6	2	2.2	cultural advantages - national
Hitt et al. (2001)	0.19	0.81			0.84	0.16	2	2.1	cultural advantages - corporate
West and Bogers (2014)	0.02	0.98			0.7	0.3	2	2.1	cultural advantages - corporate
Steenkamp et al. (1999)	0.23	0.77			0	1	2	2.2	cultural advantages - national
Hargadon and Douglas (2001)	0	1			0.85	0.15	2	2.1	cultural advantages - corporate
Cooper (2011)	0.37	0.63			0.79	0.21	2	2.1	cultural advantages - corporate
Slater and Narver (1998)	0.91	0.09	0.98	0.02			1	1.1	orientation - competitor/customer
Moorman and Miner (1997)	0.52	0.48	0.24	0.76			1	1.2	orientation - interfunctional
Meyer-Krahmer and Schmoch (1998)	0.01	0.99			0.31	0.69	2	2.2	cultural advantages - national
Taylor and Greve (2006)	0.02	0.98			0.59	0.41	2	2.1	cultural advantages - corporate
Tellis et al. (2009)	0.15	0.85			0.03	0.97	2	2.2	cultural advantages - national
Zheng et al. (2010)	0.87	0.13	0.8	0.2			1	1.1	orientation - competitor/customer
Ritter and Gemünden (2003)	0.53	0.47	0.36	0.64			1	1.2	orientation - interfunctional
McDermott and O'connor (2002)	0.2	0.8			0.97	0.03	2	2.1	cultural advantages - corporate

Continuation of Table 4

Study	p(1)	p(2)	p(1.1)	p(1.2)	p(2.1)	p(2.2)	Topic	Sub-Topic	Topic Label
Terziovski (2010)	0.47	0.53			0.49	0.51	2	2.2	cultural advantages - national
Nakata and Sivakumar (1996)	0.36	0.64			0.5	0.5	2	2.1	cultural advantages - corporate
Verganti (2008)	0.01	0.99			0.97	0.03	2	2.1	cultural advantages - corporate
García-Morales et al. (2012)	0.67	0.33	0.95	0.05			1	1.1	orientation - competitor/customer
De Brentani (2001)	0.97	0.03	0.01	0.99			1	1.2	orientation - interfunctional
Rubera and Kirca (2012)	0.75	0.25	0.96	0.04			1	1.1	orientation - competitor/customer
Srinivasan et al. (2002)	0.58	0.42	0.99	0.01			1	1.1	orientation - competitor/customer
George et al. (2012)	0.01	0.99			0.73	0.27	2	2.1	cultural advantages - corporate
Fagerberg et al. (2012)	0	1			0.22	0.78	2	2.2	cultural advantages - national
Steensma et al. (2000)	0.03	0.97			0.11	0.89	2	2.2	cultural advantages - national
Vanloqueren and Baret (2009)	0	1			0.59	0.41	2	2.1	cultural advantages - corporate
Langerak et al. (2004)	1	0	0.58	0.42			1	1.1	orientation - competitor/customer
Deshpandé and Farley (2004)	0.7	0.3	0.81	0.19			1	1.1	orientation - competitor/customer
Vera and Crossan (2005)	0.12	0.88			0.94	0.06	2	2.1	cultural advantages - corporate
Hogan and Coote (2014)	0.59	0.41	0.54	0.46			1	1.1	orientation - competitor/customer
Hansen and Løvås (2004)	0	1			0.58	0.42	2	2.1	cultural advantages - corporate
Slater and Narver (1999)	0.99	0.01	1	0			1	1.1	orientation - competitor/customer
Van den Bulte and Stremersch (2004)	0	1			0	1	2	2.2	cultural advantages - national

Continuation of Table 4

Study	p(1)	p(2)	p(1.1)	p(1.2)	p(2.1)	p(2.2)	Topic	Sub-Topic	Topic Label
Tellis et al. (2003)	0	1			0	1	2	2.2	cultural advantages - national
Elenkov et al. (2005)	0.24	0.76			0.64	0.36	2	2.1	cultural advantages - corporate
Jacob et al. (2003)	0	1			0.53	0.47	2	2.1	cultural advantages - corporate
Slater et al. (2014)	0.66	0.34	0.58	0.42			1	1.1	orientation - competitor/customer
Richard et al. (2004)	0.42	0.58			0.23	0.77	2	2.2	cultural advantages - national
Ireland et al. (2003)	0.24	0.76			0.99	0.01	2	2.1	cultural advantages - corporate
Kirca et al. (2005)	1	0	0.98	0.02			1	1.1	orientation - competitor/customer
Grinstein (2008)	0.84	0.16	1	0			1	1.1	orientation - competitor/customer
Evanschitzky et al. (2012)	0.81	0.19	0.17	0.83			1	1.2	orientation - interfunctional
Bartel and Garud (2009)	0	1			1	0	2	2.1	cultural advantages - corporate
Büschgens et al. (2013)	0.31	0.69			0.63	0.37	2	2.1	cultural advantages - corporate
Bock et al. (2012)	0.25	0.75			0.72	0.28	2	2.1	cultural advantages - corporate
Elenkov et al. (2005)	0.28	0.72			0.47	0.53	2	2.2	cultural advantages - national
Chen et al. (2005)	0.59	0.41	0	1			1	1.2	orientation - interfunctional
Laforet (2008)	0.76	0.24	0.88	0.12			1	1.1	orientation - competitor/customer
Luk et al. (2008)	0.46	0.54			0.2	0.8	2	2.2	cultural advantages - national
Hult et al. (2003)	0.92	0.08	0.99	0.01			1	1.1	orientation - competitor/customer
Zhou et al. (2005)	0.9	0.1	0.98	0.02			1	1.1	orientation - competitor/customer

Continuation of Table 4

Study	p(1)	p(2)	p(1.1)	p(1.2)	p(2.1)	p(2.2)	Topic	Sub-Topic	Topic Label
Kleinschmidt et al. (2007)	1	0	0.12	0.88			1	1.2	orientation - interfunctional
De Brentani and Kleinschmidt (2004)	0.92	0.08	0	1			1	1.2	orientation - interfunctional
Kotha et al. (2001)	0.19	0.81			0.17	0.83	2	2.2	cultural advantages - national
O'Connor (2008)	0.15	0.85			1	0	2	2.1	cultural advantages - corporate
Kuhlmann (2001)	0	1			0.23	0.77	2	2.2	cultural advantages - national
Hult et al. (2002)	0.66	0.34	0.85	0.15			1	1.1	orientation - competitor/customer
Lin et al. (2013)	0.42	0.58			0.82	0.18	2	2.1	cultural advantages - corporate
Sherman et al. (2000)	0.86	0.14	0.04	0.96			1	1.2	orientation - interfunctional
Moenaert et al. (2000)	0.33	0.67			0.87	0.13	2	2.1	cultural advantages - corporate
De Boer et al. (1999)	0.13	0.87			1	0	2	2.1	cultural advantages - corporate
Wijnberg and Gemser (2000)	0	1			0.8	0.2	2	2.1	cultural advantages - corporate
Frambach et al. (2003)	1	0	0.99	0.01			1	1.1	orientation - competitor/customer
Ahluwalia (2008)	0.9	0.1	0	1			1	1.2	orientation - interfunctional
Rothenberg (2003)	0.04	0.96			1	0	2	2.1	cultural advantages - corporate
Chua et al. (2015)	0	1			0.21	0.79	2	2.2	cultural advantages - national
Markham and Lee (2013)	0.65	0.35	0	1			1	1.2	orientation - interfunctional
Jean et al. (2010)	0.81	0.19	0.86	0.14			1	1.1	orientation - competitor/customer
Stremersch and Tellis (2004)	0	1			0	1	2	2.2	cultural advantages - national

Continuation of Table 4

Study	p(1)	p(2)	p(1.1)	p(1.2)	p(2.1)	p(2.2)	Topic	Sub-Topic	Topic Label
Martin and Grbac (2003)	1	0	1	0			1	1.1	orientation - competitor/customer
Mueller et al. (2013)	0.18	0.82			0.52	0.48	2	2.1	cultural advantages - corporate
Gao et al. (2007)	0.99	0.01	1	0			1	1.1	orientation - competitor/customer
Menguc et al. (2007)	0.96	0.04	0.99	0.01			1	1.1	orientation - competitor/customer
Storey et al. (2016)	0.9	0.1	0.16	0.84			1	1.2	orientation - interfunctional
Herrera (2015)	0.19	0.81			0.94	0.06	2	2.1	cultural advantages - corporate
Xie et al. (1998)	0.91	0.09	0.01	0.99			1	1.2	orientation - interfunctional
Vicente-Saez and Martinez-Fuentes (2018)	0	1			0.44	0.56	2	2.2	cultural advantages - national
Wei and Morgan (2004)	1	0	0.65	0.35			1	1.1	orientation - competitor/customer
Shujahat et al. (2019)	0.28	0.72			0.92	0.08	2	2.1	cultural advantages - corporate
Fang (2011)	0.45	0.55			0.65	0.35	2	2.1	cultural advantages - corporate
Kahn et al. (2012)	0.84	0.16	0	1			1	1.2	orientation - interfunctional
Souder and Song (1997)	0.96	0.04	0.08	0.92			1	1.2	orientation - interfunctional
Menon et al. (2002)	0.61	0.39	0.13	0.87			1	1.2	orientation - interfunctional
De Brentani et al. (2010)	0.98	0.02	0.11	0.89			1	1.2	orientation - interfunctional
Corsaro et al. (2012)	0.05	0.95			0.72	0.28	2	2.1	cultural advantages - corporate
Hoffman et al. (2010)	0.44	0.56			0.15	0.85	2	2.2	cultural advantages - national
Xie et al. (2003)	0.86	0.14	0	1			1	1.2	orientation - interfunctional

Continuation of Table 4

Study	p(1)	p(2)	p(1.1)	p(1.2)	p(2.1)	p(2.2)	Topic	Sub-Topic	Topic Label
Bello et al. (2004)	0.02	0.98			0.67	0.33	2	2.1	cultural advantages - corporate
Zhao and Cavusgil (2006)	1	0	1	0			1	1.1	orientation - competitor/customer
Story et al. (2017)	0.66	0.34	0.44	0.56			1	1.2	orientation - interfunctional
Ali and Park (2016)	0.67	0.33	0.92	0.08			1	1.1	orientation - competitor/customer
Kester et al. (2011)	0.78	0.22	0.01	0.99			1	1.2	orientation - interfunctional
Sundqvist et al. (2005)	0.01	0.99			0	1	2	2.2	cultural advantages - national
Deshpandé et al. (2000)	0.64	0.36	0.82	0.18			1	1.1	orientation - competitor/customer
Dougherty and Dunne (2012)	0	1			1	0	2	2.1	cultural advantages - corporate
Buck and Shahrim (2005)	0	1			0.09	0.91	2	2.2	cultural advantages - national
Castaño et al. (2015)	0.14	0.86			0.09	0.91	2	2.2	cultural advantages - national
Lemola (2002)	0	1			0.3	0.7	2	2.2	cultural advantages - national
Griffiths-Hemans and Grover (2006)	0.2	0.8			0.9	0.1	2	2.1	cultural advantages - corporate
Kumar and Krishnan (2002)	0	1			0	1	2	2.2	cultural advantages - national
Cooper (2019)	0.82	0.18	0.03	0.97			1	1.2	orientation - interfunctional
Theoharakis and Hooley (2008)	0.91	0.09	0.99	0.01			1	1.1	orientation - competitor/customer
Lazzeretti and Capone (2016)	0	1			0.24	0.76	2	2.2	cultural advantages - national
Zien and Buckler (1997)	0.23	0.77			0.98	0.02	2	2.1	cultural advantages - corporate
Bstieler and Hemmert (2010)	0.55	0.45	0.05	0.95			1	1.2	orientation - interfunctional

Continuation of Table 4

Study	p(1)	p(2)	p(1.1)	p(1.2)	p(2.1)	p(2.2)	Topic	Sub-Topic	Topic Label
Johnsen and Ford (2006)	0.62	0.38	0.77	0.23			1	1.1	orientation - competitor/customer
Calantone and Rubera (2012)	0.77	0.23	0.43	0.57			1	1.2	orientation - interfunctional
Paswan and Wittmann (2009)	0.22	0.78			1	0	2	2.1	cultural advantages - corporate
Wright et al. (2012)	0	1			0.98	0.02	2	2.1	cultural advantages - corporate
Leal-Rodríguez et al. (2014)	0.41	0.59			0.56	0.44	2	2.1	cultural advantages - corporate
Ozkaya et al. (2015)	1	0	1	0			1	1.1	orientation - competitor/customer
Brockman and Morgan (2006)	0.97	0.03	0.29	0.71			1	1.2	orientation - interfunctional
Kumar et al. (1998)	0	1			0	1	2	2.2	cultural advantages - national
Arundel et al. (2015)	0.01	0.99			0.14	0.86	2	2.2	cultural advantages - national
Dibrell et al. (2011)	0.95	0.05	1	0			1	1.1	orientation - competitor/customer
Souder and Song (1998)	0.94	0.06	0	1			1	1.2	orientation - interfunctional
Stock et al. (2013)	0.8	0.2	0.72	0.28			1	1.1	orientation - competitor/customer
Agarwal and Bayus (2002)	0.08	0.92			0	1	2	2.2	cultural advantages - national
Ahuja (2000)	0	1			0.09	0.91	2	2.2	cultural advantages - national
Anderson et al. (2014)	0.11	0.89			0.8	0.2	2	2.1	cultural advantages - corporate
Anning-Dorson (2018)	0.89	0.11	0.79	0.21			1	1.1	orientation - competitor/customer
Atuahene-Gima (1996)	1	0	0.74	0.26			1	1.1	orientation - competitor/customer
Atuahene-Gima and Ko (2001)	1	0	0.74	0.26			1	1.1	orientation - competitor/customer

Continuation of Table 4

Study	p(1)	p(2)	p(1.1)	p(1.2)	p(2.1)	p(2.2)	Topic	Sub-Topic	Topic Label
Atuahene-Gima et al. (2005)	1	0	0.83	0.17			1	1.1	orientation - competitor/customer
Atuahene-Gima and Wei (2011)	1	0	0.33	0.67			1	1.2	orientation - interfunctional
Augusto and Coelho (2009)	1	0	0.97	0.03			1	1.1	orientation - competitor/customer
Ayers et al. (1997)	0.93	0.07	0	1			1	1.2	orientation - interfunctional
Baker and Sinkula (2005)	1	0	0.96	0.04			1	1.1	orientation - competitor/customer
Baker and Sinkula (1999)	1	0	1	0			1	1.1	orientation - competitor/customer
Baker and Sinkula (2007)	0.99	0.01	1	0			1	1.1	orientation - competitor/customer
Bartholomew (1997)	0.01	0.99			0.27	0.73	2	2.2	cultural advantages - national
Bharadwaj and Menon (2000)	0.46	0.54			0.63	0.37	2	2.1	cultural advantages - corporate
Bhuian et al. (2005)	1	0	0.96	0.04			1	1.1	orientation - competitor/customer
Bonner et al. (2002)	0.99	0.01	0	1			1	1.2	orientation - interfunctional
Burrus et al. (2018)	0.01	0.99			0.1	0.9	2	2.2	cultural advantages - national
Calantone et al. (2002)	0.94	0.06	0.98	0.02			1	1.1	orientation - competitor/customer
Calantone et al. (2006)	0.96	0.04	0.29	0.71			1	1.2	orientation - interfunctional
Calantone et al. (2003)	1	0	0.48	0.52			1	1.2	orientation - interfunctional
Calantone et al. (2010)	0.94	0.06	0.56	0.44			1	1.1	orientation - competitor/customer
Calantone et al. (1996)	0.95	0.05	0.09	0.91			1	1.2	orientation - interfunctional
Capaldo (2007)	0.04	0.96			0.62	0.38	2	2.1	cultural advantages - corporate

Continuation of Table 4

Study	p(1)	p(2)	p(1.1)	p(1.2)	p(2.1)	p(2.2)	Topic	Sub-Topic	Topic Label
Carmona-Lavado et al. (2010)	0.33	0.67			0.72	0.28	2	2.1	cultural advantages - corporate
Chandy and Tellis (2000)	0.01	0.99			0.13	0.87	2	2.2	cultural advantages - national
Chandy and Tellis (1998)	0.32	0.68			0.21	0.79	2	2.2	cultural advantages - national
Christensen (2006)	0.05	0.95			0.84	0.16	2	2.1	cultural advantages - corporate
Christensen and Bower (1996)	0.12	0.88			0.87	0.13	2	2.1	cultural advantages - corporate
Cui and O'Connor (2012)	0.13	0.87			0.41	0.59	2	2.2	cultural advantages - national
Damanpour et al. (2009)	0.32	0.68			0.56	0.44	2	2.1	cultural advantages - corporate
Danneels (2002)	0.53	0.47	0.49	0.51			1	1.2	orientation - interfunctional
Danneels and Kleinschmidtb (2001)	1	0	0.05	0.95			1	1.2	orientation - interfunctional
Dekimpe et al. (2000)	0	1			0	1	2	2.2	cultural advantages - national
DeSarbo et al. (2005)	0.93	0.07	0.78	0.22			1	1.1	orientation - competitor/customer
Dess et al. (1997)	0.97	0.03	0.54	0.46			1	1.1	orientation - competitor/customer
Dougherty and Hardy (1996)	0.06	0.94			0.98	0.02	2	2.1	cultural advantages - corporate
Droge et al. (2008)	1	0	0.8	0.2			1	1.1	orientation - competitor/customer
Dutta et al. (1999)	0.43	0.57			0.18	0.82	2	2.2	cultural advantages - national
Gallego et al. (2013)	0.01	0.99			0.3	0.7	2	2.2	cultural advantages - national
Ganesh et al. (1997)	0.03	0.97			0	1	2	2.2	cultural advantages - national
Gatignon and Xuereb (1997)	0.97	0.03	0.94	0.06			1	1.1	orientation - competitor/customer

Continuation of Table 4

Study	p(1)	p(2)	p(1.1)	p(1.2)	p(2.1)	p(2.2)	Topic	Sub-Topic	Topic Label
Gielens and Steenkamp (2007)	0.37	0.63			0	1	2	2.2	cultural advantages - national
Golder and Tellis (1997)	0	1			0	1	2	2.2	cultural advantages - national
Golder and Tellis (2004)	0	1			0	1	2	2.2	cultural advantages - national
Gruner and Homburg (2000)	0.9	0.1	0.16	0.84			1	1.2	orientation - interfunctional
Gumusluoğlu and Ilsev (2009)	0.39	0.61			0.63	0.37	2	2.1	cultural advantages - corporate
Harmancioglu et al. (2010)	0.98	0.02	0.83	0.17			1	1.1	orientation - competitor/customer
Hauser et al. (2006)	0.34	0.66			0.52	0.48	2	2.1	cultural advantages - corporate
Henard and Szymanski (2001)	0.98	0.02	0.1	0.9			1	1.2	orientation - interfunctional
Hirst et al. (2009)	1	0	0.61	0.39			1	1.1	orientation - competitor/customer
Hughes and Morgan (2007)	0.92	0.08	0.61	0.39			1	1.1	orientation - competitor/customer
Hult and Ketchen Jr (2001)	0.9	0.1	1	0			1	1.1	orientation - competitor/customer
Hult et al. (2004)	1	0	0.99	0.01			1	1.1	orientation - competitor/customer
Hultink et al. (1997)	0.96	0.04	0	1			1	1.2	orientation - interfunctional
Jansen et al. (2006)	0.36	0.64			0.88	0.12	2	2.1	cultural advantages - corporate
Jansen et al. (2009)	0.22	0.78			0.99	0.01	2	2.1	cultural advantages - corporate
Kahn (2001)	1	0	0.78	0.22			1	1.1	orientation - competitor/customer
Katila (2002)	0.05	0.95			0.31	0.69	2	2.2	cultural advantages - national
Katila and Shane (2005)	0.11	0.89			0.39	0.61	2	2.2	cultural advantages - national

Continuation of Table 4

Study	p(1)	p(2)	p(1.1)	p(1.2)	p(2.1)	p(2.2)	Topic	Sub-Topic	Topic Label
Kester et al. (2014)	0.99	0.01	0.08	0.92			1	1.2	orientation - interfunctional
Kochhar and David (1996)	0.02	0.98			0.07	0.93	2	2.2	cultural advantages - national
Kyriakopoulos and Moorman (2004)	0.8	0.2	0.87	0.13			1	1.1	orientation - competitor/customer
Laursen and Salter (2006)	0.02	0.98			0.35	0.65	2	2.2	cultural advantages - national
Li and Calantone (1998)	0.96	0.04	0.62	0.38			1	1.1	orientation - competitor/customer
Lisboa et al. (2011)	0.8	0.2	0.97	0.03			1	1.1	orientation - competitor/customer
De Luca et al. (2010)	0.82	0.18	0.76	0.24			1	1.1	orientation - competitor/customer
Lukas and Ferrell (2000)	1	0	0.93	0.07			1	1.1	orientation - competitor/customer
Lumpkin and Dess (1996)	0.64	0.36	0.4	0.6			1	1.2	orientation - interfunctional
Manu and Sriram (1996)	0.85	0.15	0.59	0.41			1	1.1	orientation - competitor/customer
Matsuno and Mentzer (2000)	1	0	1	0			1	1.1	orientation - competitor/customer
Matsuno et al. (2002)	1	0	0.98	0.02			1	1.1	orientation - competitor/customer
Mirvis et al. (2016)	0	1			0.92	0.08	2	2.1	cultural advantages - corporate
Morgan and Berthon (2008)	0.71	0.29	0.99	0.01			1	1.1	orientation - competitor/customer
Narver et al. (2004)	1	0	0.99	0.01			1	1.1	orientation - competitor/customer
Ngo and O'Cass (2012)	0.99	0.01	0.97	0.03			1	1.1	orientation - competitor/customer
Noble et al. (2002)	0.86	0.14	0.97	0.03			1	1.1	orientation - competitor/customer
O'Connor (1998)	0.46	0.54			0.99	0.01	2	2.1	cultural advantages - corporate

Continuation of Table 4

Study	p(1)	p(2)	p(1.1)	p(1.2)	p(2.1)	p(2.2)	Topic	Sub-Topic	Topic Label
O'Connor and Veryzer (2001)	0.21	0.79			1	0	2	2.1	cultural advantages - corporate
O'Connor and Rice (2013)	0.15	0.85			1	0	2	2.1	cultural advantages - corporate
Olson et al. (2001)	0.99	0.01	0	1			1	1.2	orientation - interfunctional
Olson et al. (2005)	1	0	0.93	0.07			1	1.1	orientation - competitor/customer
Paladino (2007)	1	0	0.99	0.01			1	1.1	orientation - competitor/customer
Paladino (2008)	1	0	1	0			1	1.1	orientation - competitor/customer
Parry et al. (2009)	0.84	0.16	0	1			1	1.2	orientation - interfunctional
Pauwels et al. (2004)	0.29	0.71			0	1	2	2.2	cultural advantages - national
Pelham and Wilson (1995)	1	0	0.99	0.01			1	1.1	orientation - competitor/customer
Putsis Jr et al. (1997)	0	1			0	1	2	2.2	cultural advantages - national
Raisch et al. (2009)	0.06	0.94			1	0	2	2.1	cultural advantages - corporate
Rodríguez-Pinto et al. (2011)	0.99	0.01	0.52	0.48			1	1.1	orientation - competitor/customer
Sandvik and Sandvik (2003)	1	0	0.97	0.03			1	1.1	orientation - competitor/customer
Semrau et al. (2016)	0.54	0.46	0.68	0.32			1	1.1	orientation - competitor/customer
Sethi (2000)	0.94	0.06	0.1	0.9			1	1.2	orientation - interfunctional
Sethi et al. (2001)	0.66	0.34	0.03	0.97			1	1.2	orientation - interfunctional
Sharma and Lacey (2004)	0.42	0.58			0.32	0.68	2	2.2	cultural advantages - national
Siguaw et al. (2006)	0.7	0.3	0.92	0.08			1	1.1	orientation - competitor/customer

Continuation of Table 4

Study	p(1)	p(2)	p(1.1)	p(1.2)	p(2.1)	p(2.2)	Topic	Sub-Topic	Topic Label
Simpson et al. (2006)	0.65	0.35	0.78	0.22			1	1.1	orientation - competitor/customer
Sivadas and Dwyer (2000)	0.43	0.57			0.82	0.18	2	2.1	cultural advantages - corporate
Smith and Tushman (2005)	0.07	0.93			1	0	2	2.1	cultural advantages - corporate
Song and Parry (1997)	0.98	0.02	0.03	0.97			1	1.2	orientation - interfunctional
Song and Xie (2000)	1	0	0.02	0.98			1	1.2	orientation - interfunctional
Sood and Tellis (2009)	0.12	0.88			0.11	0.89	2	2.2	cultural advantages - national
Sorescu and Spanjol (2008)	0.1	0.9			0.02	0.98	2	2.2	cultural advantages - national
Sorescu et al. (2003)	0.14	0.86			0.04	0.96	2	2.2	cultural advantages - national
Stam and Elfring (2008)	0.31	0.69			0.39	0.61	2	2.2	cultural advantages - national
Swink (2000)	0.93	0.07	0.01	0.99			1	1.2	orientation - interfunctional
Szymanski et al. (2007)	0.96	0.04	0.3	0.7			1	1.2	orientation - interfunctional
Talke et al. (2011)	0.73	0.27	0.62	0.38			1	1.1	orientation - competitor/customer
Talukdar et al. (2002)	0.01	0.99			0	1	2	2.2	cultural advantages - national
Tatikonda and Montoya-Weiss (2001)	0.87	0.13	0.04	0.96			1	1.2	orientation - interfunctional
Teece (2007)	0.15	0.85			0.95	0.05	2	2.1	cultural advantages - corporate
Theoharakis and Hooley (2008)	0.91	0.09	0.99	0.01			1	1.1	orientation - competitor/customer
Thomas and Mueller (2000)	0.04	0.96			0.17	0.83	2	2.2	cultural advantages - national
Tsai (2001)	0.08	0.92			0.88	0.12	2	2.1	cultural advantages - corporate

Continuation of Table 4

Study	p(1)	p(2)	p(1.1)	p(1.2)	p(2.1)	p(2.2)	Topic	Sub-Topic	Topic Label
Tyler and Gnyawali (2002)	0.91	0.09	0.78	0.22			1	1.1	orientation - competitor/customer
Van den Bulte (2000)	0	1			0	1	2	2.2	cultural advantages - national
Van der Vegt et al. (2005)	0.07	0.93			0.13	0.87	2	2.2	cultural advantages - national
Wei and Atuahene-Gima (2009)	1	0	0.44	0.56			1	1.2	orientation - interfunctional
Wiklund and Shepherd (2003)	0.6	0.4	0.88	0.12			1	1.1	orientation - competitor/customer
Woodside (2005)	0.88	0.12	0.93	0.07			1	1.1	orientation - competitor/customer
Yadav et al. (2007)	0.08	0.92			0.22	0.78	2	2.2	cultural advantages - national
Yli-Renko et al. (2001)	0.34	0.66			0.6	0.4	2	2.1	cultural advantages - corporate
Zhou et al. (2005)	0.9	0.1	0.98	0.02			1	1.1	orientation - competitor/customer
Zhou and Wu (2010)	0.28	0.72			0.78	0.22	2	2.1	cultural advantages - corporate
Zhou et al. (2009)	1	0	1	0			1	1.1	orientation - competitor/customer
Amabile et al. (1996)	0.46	0.54			0.85	0.15	2	2.1	cultural advantages - corporate
Brockman et al. (2010)	0.73	0.27	0	1			1	1.2	orientation - interfunctional
Gatignon et al. (2002)	0.26	0.74			0.74	0.26	2	2.1	cultural advantages - corporate
Reid and De Brentani (2004)	0.46	0.54			1	0	2	2.1	cultural advantages - corporate
Subramaniam and Youndt (2005)	0.04	0.96			0.9	0.1	2	2.1	cultural advantages - corporate
Terwiesch and Xu (2008)	0	1			0.39	0.61	2	2.2	cultural advantages - national
Wei and Morgan (2004)	1	0	0.63	0.37			1	1.1	orientation - competitor/customer

Continuation of Table 4

Study	**p(1)**	**p(2)**	**p(1.1)**	**p(1.2)**	**p(2.1)**	**p(2.2)**	**Topic**	**Sub-Topic**	**Topic Label**
He and Wong (2004)	0.25	0.75			0.49	0.51	2	2.2	cultural advantages - national
Poskela and Martinsuo (2009)	0.91	0.09	0	1			1	1.2	orientation - interfunctional
Christiansen and Varnes (2009)	0.49	0.51			0.97	0.03	2	2.1	cultural advantages - corporate

References

Agarwal, R. and Bayus, B. L. (2002). The market evolution and sales takeoff of product innovations. *Management Science*, 48(8):1024–1041.

Ahluwalia, R. (2008). How far can a brand stretch? understanding the role of self-construal. *Journal of Marketing Research*, 45(3):337–350.

Ahuja, G. (2000). Collaboration networks, structural holes, and innovation: A longitudinal study. *Administrative science quarterly*, 45(3):425–455.

Ali, M. and Park, K. (2016). The mediating role of an innovative culture in the relationship between absorptive capacity and technical and non-technical innovation. *Journal of business research*, 69(5):1669–1675.

Amabile, T. M., Conti, R., Coon, H., Lazenby, J., and Herron, M. (1996). Assessing the work environment for creativity. *Academy of management journal*, 39(5):1154–1184.

Anderson, N., Potočnik, K., and Zhou, J. (2014). Innovation and creativity in organizations: A state-of-the-science review, prospective commentary, and guiding framework. *Journal of management*, 40(5):1297–1333.

Anning-Dorson, T. (2018). Customer involvement capability and service firm performance: The mediating role of innovation. *Journal of Business Research*, 86:269–280.

Archibugi, D. and Coco, A. (2005). Measuring technological capabilities at the country level: A survey and a menu for choice. *Research policy*, 34(2):175–194.

Arundel, A., Casali, L., and Hollanders, H. (2015). How european public sector agencies innovate: The use of bottom-up, policy-dependent and knowledge-scanning innovation methods. *Research Policy*, 44(7):1271–1282.

Atuahene-Gima, K. (1996). Market orientation and innovation. *Journal of business research*, 35(2):93–103.

Atuahene-Gima, K. and Ko, A. (2001). An empirical investigation of the effect of market orientation and entrepreneurship orientation alignment on product innovation. *Organization science*, 12(1):54–74.

Atuahene-Gima, K., Slater, S. F., and Olson, E. M. (2005). The contingent value of responsive and proactive market orientations for new product program performance. *Journal of product innovation management*, 22(6):464–482.

Atuahene-Gima, K. and Wei, Y. (2011). The vital role of problem-solving competence in new product success. *Journal of Product Innovation Management*, 28(1):81–98.

Augusto, M. and Coelho, F. (2009). Market orientation and new-to-the-world products: Exploring the moderating effects of innovativeness, competitive strength, and environmental forces. *Industrial marketing management*, 38(1):94–108.

Ayers, D., Dahlstrom, R., and Skinner, S. J. (1997). An exploratory investigation of organizational antecedents to new product success. *Journal of Marketing Research*, 34(1):107–116.

Baker, W. E. and Sinkula, J. M. (1999). The synergistic effect of market orientation and learning orientation on organizational performance. *Journal of the academy of marketing science*, 27(4):411–427.

Baker, W. E. and Sinkula, J. M. (2005). Market orientation and the new product paradox. *Journal of Product Innovation Management*, 22(6):483–502.

Baker, W. E. and Sinkula, J. M. (2007). Does market orientation facilitate balanced innovation programs? an organizational learning perspective. *Journal of product innovation management*, 24(4):316–334.

Barney, J. B. (1986). Organizational culture: can it be a source of sustained competitive advantage? *Academy of management review*, 11(3):656–665.

Bartel, C. A. and Garud, R. (2009). The role of narratives in sustaining organizational innovation. *Organization science*, 20(1):107–117.

Bartholomew, S. (1997). National systems of biotechnology innovation: complex interdependence in the global system. *Journal of international business studies*, 28(2):241–266.

Baumgartner, H. and Pieters, R. (2003). The structural influence of marketing journals: A citation analysis of the discipline and its subareas over time. *Journal of marketing*, 67(2):123–139.

Bello, D. C., Lohtia, R., and Sangtani, V. (2004). An institutional analysis of supply chain innovations in global marketing channels. *Industrial Marketing Management*, 33(1):57–64.

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

Bharadwaj, S. and Menon, A. (2000). Making innovation happen in organizations: individual creativity mechanisms, organizational creativity mechanisms or both? *Journal of Product Innovation Management: An International Publication of the Product Development & Management Association*, 17(6):424–434.

Bhuian, S. N., Menguc, B., and Bell, S. J. (2005). Just entrepreneurial enough: the moderating effect of entrepreneurship on the relationship between market orientation and performance. *Journal of business research*, 58(1):9–17.

Biemans, W., Griffin, A., and Moenaert, R. (2010). In search of the classics: A study of the impact of jpim papers from 1984 to 2003. *Journal of Product Innovation Management*, 27(4):461–484.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022.

Bock, A. J., Opsahl, T., George, G., and Gann, D. M. (2012). The effects of culture and structure on strategic flexibility during business model innovation. *Journal of Management studies*, 49(2):279–305.

Bonner, J. M., Ruekert, R. W., and Walker Jr, O. C. (2002). Upper management control of new product development projects and project performance. *Journal of Product Innovation Management: An International Publication of the Product Development & Management Association*, 19(3):233–245.

Brockman, B. K. and Morgan, R. M. (2006). The moderating effect of organizational cohesiveness in knowledge use and new product development. *Journal of the Academy of Marketing Science*, 34(3):295–307.

Brockman, B. K., Rawlston, M. E., Jones, M. A., and Halstead, D. (2010). An exploratory model of interpersonal cohesiveness in new product development teams. *Journal of Product Innovation Management*, 27(2):201–219.

Bstieler, L. and Hemmert, M. (2010). Increasing learning and time efficiency in interorganizational new product development teams. *Journal of Product Innovation Management*, 27(4):485–499.

Buck, T. and Shahrim, A. (2005). The translation of corporate governance changes across national cultures: The case of germany. *Journal of International Business Studies*, 36(1):42–61.

Burrus, R. T., Graham, J. E., and Jones, A. T. (2018). Regional innovation and firm performance. *Journal of Business Research*, 88:357–362.

Büschgens, T., Bausch, A., and Balkin, D. B. (2013). Organizational culture and innovation: A meta-analytic review. *Journal of product innovation management*, 30(4):763–781.

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

Büschken, J. and Allenby, G. M. (2020). Improving text analysis using sentence conjunctions and punctuation. *Marketing Science*, 39(4):727–742.

Calantone, R., Garcia, R., and Dröge, C. (2003). The effects of environmental turbulence on new product development strategy planning. *Journal of product innovation management*, 20(2):90–103.

Calantone, R. and Rubera, G. (2012). When should rd & e and marketing collaborate? the moderating role of exploration–exploitation and environmental uncertainty. *Journal of Product Innovation Management*, 29(1):144–157.

Calantone, R. J., Cavusgil, S. T., and Zhao, Y. (2002). Learning orientation, firm innovation capability, and firm performance. *Industrial marketing management*, 31(6):515–524.

Calantone, R. J., Chan, K., and Cui, A. S. (2006). Decomposing product innovativeness and its effects on new product success. *Journal of Product Innovation Management*, 23(5):408–421.

Calantone, R. J., Harmancioglu, N., and Droge, C. (2010). Inconclusive innovation "returns": A meta-analysis of research on innovation in new product development. *Journal of Product Innovation Management*, 27(7):1065–1081.

Calantone, R. J., Schmidt, J. B., and Song, X. M. (1996). Controllable factors of new product success: A cross-national comparison. *Marketing science*, 15(4):341–358.

Capaldo, A. (2007). Network structure and innovation: The leveraging of a dual network as a distinctive relational capability. *Strategic management journal*, 28(6):585–608.

Carmona-Lavado, A., Cuevas-Rodríguez, G., and Cabello-Medina, C. (2010). Social and organizational capital: Building the context for innovation. *Industrial Marketing Management*, 39(4):681–690.

Castaño, M.-S., Méndez, M.-T., and Galindo, M.-Á. (2015). The effect of social, cultural, and economic factors on entrepreneurship. *Journal of business research*, 68(7):1496–1500.

Cerf, V. G. (2020). Growing a culture of innovation: 5 lessons from google.

Chandy, R. K. and Tellis, G. J. (1998). Organizing for radical product innovation: The overlooked role of willingness to cannibalize. *Journal of marketing research*, 35(4):474–487.

Chandy, R. K. and Tellis, G. J. (2000). The incumbent's curse? incumbency, size, and radical product innovation. *Journal of marketing*, 64(3):1–17.

Chen, G., Liu, C., and Tjosvold, D. (2005). Conflict management for effective top management teams and innovation in china. *Journal of Management Studies*, 42(2):277–300.

Christensen, C. M. (2006). The ongoing process of building a theory of disruption. *Journal of Product innovation management*, 23(1):39–55.

Christensen, C. M. and Bower, J. L. (1996). Customer power, strategic investment, and the failure of leading firms. *Strategic management journal*, 17(3):197–218.

Christiansen, J. K. and Varnes, C. J. (2009). Formal rules in product development: Sensemaking of structured approaches. *Journal of Product Innovation Management*, 26(5):502–519.

Chua, R. Y., Roth, Y., and Lemoine, J.-F. (2015). The impact of culture on creativity: How cultural tightness and cultural distance affect global innovation crowdsourcing work. *Administrative Science Quarterly*, 60(2):189–227.

Cillo, P., Griffith, D. A., and Rubera, G. (2018). The new product portfolio innovativeness–stock returns relationship: The role of large individual investors' culture. *Journal of Marketing*, 82(6):49–70.

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., and Kuksa, P. (2011). Natural language processing (almost) from scratch. *Journal of machine learning research*, 12(ARTICLE):2493–2537.

Cooke, P., Uranga, M. G., and Etxebarria, G. (1997). Regional innovation systems: Institutional and organisational dimensions. *Research policy*, 26(4-5):475–491.

Cooper, R. G. (2011). Perspective: The innovation dilemma: How to innovate when the market is mature. *Journal of Product Innovation Management*, 28(s1):2–27.

Cooper, R. G. (2019). The drivers of success in new-product development. *Industrial Marketing Management*, 76:36–47.

Corsaro, D., Cantù, C., and Tunisini, A. (2012). Actors' heterogeneity in innovation networks. *Industrial Marketing Management*, 41(5):780–789.

Crossan, M. M. and Apaydin, M. (2010). A multi-dimensional framework of organizational innovation: A systematic review of the literature. *Journal of management studies*, 47(6):1154–1191.

Cui, A. S. and O'Connor, G. (2012). Alliance portfolio resource diversity and firm innovation. *Journal of Marketing*, 76(4):24–43.

Damanpour, F. (1991). Organizational innovation: A meta-analysis of effects of determinants and moderators. *Academy of management journal*, 34(3):555–590.

Damanpour, F., Walker, R. M., and Avellaneda, C. N. (2009). Combinative effects of innovation types and organizational performance: A longitudinal study of service organizations. *Journal of management studies*, 46(4):650–675.

Danneels, E. (2002). The dynamics of product innovation and firm competences. *Strategic management journal*, 23(12):1095–1121.

Danneels, E. and Kleinschmidtb, E. J. (2001). Product innovativeness from the firm's perspective: Its dimensions and their relation with project selection and performance. *Journal of Product Innovation Management: An International Publication of the Product Development & Management Association*, 18(6):357–373.

Day, G. S. (1994). The capabilities of market-driven organizations. *Journal of marketing*, 58(4):37–52.

De Boer, M., Van Den Bosch, F. A., and Volberda, H. W. (1999). Managing organizational knowledge integration in the emerging multimedia complex. *Journal of management studies*, 36(3):379–398.

De Brentani, U. (2001). Innovative versus incremental new business services: Different keys for achieving success. *Journal of Product Innovation Management: An International Publication of the Product Development & Management Association*, 18(3):169–187.

De Brentani, U. and Kleinschmidt, E. J. (2004). Corporate culture and commitment:

impact on performance of international new product development programs. *Journal of product innovation management*, 21(5):309–333.

De Brentani, U., Kleinschmidt, E. J., and Salomo, S. (2010). Success in global new product development: Impact of strategy and the behavioral environment of the firm. *Journal of product innovation management*, 27(2):143–160.

De Luca, L. M., Verona, G., and Vicari, S. (2010). Market orientation and r&d effectiveness in high-technology firms: An empirical investigation in the biotechnology industry. *Journal of Product Innovation Management*, 27(3):299–320.

Dekimpe, M. G., Parker, P. M., and Sarvary, M. (2000). Global diffusion of technological innovations: A coupled-hazard approach. *Journal of Marketing Research*, 37(1):47–59.

DeSarbo, W. S., Anthony Di Benedetto, C., Song, M., and Sinha, I. (2005). Revisiting the miles and snow strategic framework: uncovering interrelationships between strategic types, capabilities, environmental uncertainty, and firm performance. *Strategic management journal*, 26(1):47–74.

Deshpandé, R. and Farley, J. U. (1998). Measuring market orientation: generalization and synthesis. *Journal of market-focused management*, 2(3):213–232.

Deshpandé, R. and Farley, J. U. (2004). Organizational culture, market orientation, innovativeness, and firm performance: an international research odyssey. *International Journal of research in Marketing*, 21(1):3–22.

Deshpandé, R., Farley, J. U., and Webster Jr, F. E. (1993). Corporate culture, customer orientation, and innovativeness in japanese firms: a quadrad analysis. *Journal of marketing*, 57(1):23–37.

Deshpandé, R., Farley, J. U., and Webster Jr, F. E. (2000). Triad lessons: generalizing results on high performance firms in five business-to-business markets. *International Journal of Research in Marketing*, 17(4):353–362.

Dess, G. G., Lumpkin, G. T., and Covin, J. G. (1997). Entrepreneurial strategy making and firm performance: Tests of contingency and configurational models. *Strategic management journal*, 18(9):677–695.

Dibrell, C., Craig, J. B., and Hansen, E. N. (2011). How managerial attitudes toward the natural environment affect market orientation and innovation. *Journal of Business Research*, 64(4):401–407.

Dougherty, D. and Dunne, D. D. (2012). Digital science and knowledge boundaries in complex innovation. *Organization Science*, 23(5):1467–1484.

Dougherty, D. and Hardy, C. (1996). Sustained product innovation in large, mature organizations: Overcoming innovation-to-organization problems. *Academy of management journal*, 39(5):1120–1153.

Droge, C., Calantone, R., and Harmancioglu, N. (2008). New product success: is it really controllable by managers in highly turbulent environments? *Journal of Product Innovation Management*, 25(3):272–286.

Dutta, S., Narasimhan, O., and Rajiv, S. (1999). Success in high-technology markets: Is marketing capability critical? *Marketing science*, 18(4):547–568.

Eisend, M., Evanschitzky, H., and Gilliland, D. I. (2016). The influence of organizational and national culture on new product performance. *Journal of Product Innovation Management*, 33(3):260–276.

Elenkov, D. S., Judge, W., and Wright, P. (2005). Strategic leadership and executive innovation influence: an international multi-cluster comparative study. *Strategic management journal*, 26(7):665–682.

Evanschitzky, H., Eisend, M., Calantone, R. J., and Jiang, Y. (2012). Success factors of product innovation: An updated meta-analysis. *Journal of product innovation management*, 29:21–37.

Fagerberg, J., Fosaas, M., and Sapprasert, K. (2012). Innovation: Exploring the knowledge base. *Research policy*, 41(7):1132–1153.

Fang, E. (2011). The effect of strategic alliance knowledge complementarity on new product innovativeness in china. *Organization Science*, 22(1):158–172.

Ferrell, O., Gonzalez-Padron, T. L., Hult, G. T. M., and Maignan, I. (2010). From market orientation to stakeholder orientation. *Journal of Public Policy & Marketing*, 29(1):93–96.

Frambach, R. T., Prabhu, J., and Verhallen, T. M. (2003). The influence of business strategy on new product activity: The role of market orientation. *International journal of research in marketing*, 20(4):377–397.

Galambos, L. (1997). Global perspectives on modern business. *Business History Review*, 71(2):287–290.

Gallego, J., Rubalcaba, L., and Suárez, C. (2013). Knowledge for innovation in europe: The role of external knowledge on firms' cooperation strategies. *Journal of Business Research*, 66(10):2034–2041.

Ganesh, J., Kumar, V., and Subramaniam, V. (1997). Learning effect in multinational diffusion of consumer durables: An exploratory investigation. *Journal of the Academy of Marketing Science*, 25(3):214–228.

Gao, G. Y., Zhou, K. Z., and Yim, C. K. B. (2007). On what should firms focus in transitional economies? a study of the contingent value of strategic orientations in china. *International journal of research in marketing*, 24(1):3–15.

García-Morales, V. J., Jiménez-Barrionuevo, M. M., and Gutiérrez-Gutiérrez, L. (2012). Transformational leadership influence on organizational performance through organizational learning and innovation. *Journal of business research*, 65(7):1040–1050.

Gareth, J., Daniela, W., Trevor, H., and Robert, T. (2013). *An introduction to statistical learning: with applications in R*. Spinger.

Gatignon, H., Tushman, M. L., Smith, W., and Anderson, P. (2002). A structural approach to assessing innovation: Construct development of innovation locus, type, and characteristics. *Management science*, 48(9):1103–1122.

Gatignon, H. and Xuereb, J.-M. (1997). Strategic orientation of the firm and new product performance. *Journal of marketing research*, 34(1):77–90.

Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., Duan, L., Almaliach, A., Ang, S., Arnadottir, J., et al. (2011). Differences between tight and loose cultures: A 33-nation study. *science*, 332(6033):1100–1104.

George, G., McGahan, A. M., and Prabhu, J. (2012). Innovation for inclusive growth: Towards a theoretical framework and a research agenda. *Journal of management studies*, 49(4):661–683.

Gielens, K. and Steenkamp, J.-B. E. (2007). Drivers of consumer acceptance of new packaged goods: An investigation across products and countries. *International Journal of Research in Marketing*, 24(2):97–111.

Golder, P. and Tellis, G. J. (2001). Will and vision: How latecomers grow to dominate markets.

Golder, P. N. and Tellis, G. J. (1997). Will it ever fly? modeling the takeoff of really new consumer durables. *Marketing Science*, 16(3):256–270.

Golder, P. N. and Tellis, G. J. (2004). Growing, growing, gone: Cascades, diffusion, and turning points in the product life cycle. *Marketing Science*, 23(2):207–218.

Govindarajan, V. and Kopalle, P. K. (2006). The usefulness of measuring disruptiveness of innovations ex post in making ex ante predictions. *Journal of product innovation management*, 23(1):12–18.

Green, G. M. (2012). *Pragmatics and natural language understanding*. Routledge.

Greenley, G. E. and Foxall, G. R. (1997). Multiple stakeholder orientation in uk companies and the implications for company performance. *Journal of Management Studies*, 34(2):259–284.

Griffiths, T. L. and Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National academy of Sciences*, 101(suppl 1):5228–5235.

Griffiths-Hemans, J. and Grover, R. (2006). Setting the stage for creative new products: investigating the idea fruition process. *Journal of the Academy of Marketing Science*, 34(1):27–39.

Grinstein, A. (2008). The effect of market orientation and its components on innovation consequences: a meta-analysis. *Journal of the academy of Marketing science*, 36(2):166–173.

Gruner, K. E. and Homburg, C. (2000). Does customer interaction enhance new product success? *Journal of business research*, 49(1):1–14.

Gumusluoğlu, L. and Ilsev, A. (2009). Transformational leadership and organizational innovation: The roles of internal and external support for innovation. *Journal of Product Innovation Management*, 26(3):264–277.

Han, J. K., Kim, N., and Srivastava, R. K. (1998). Market orientation and organizational performance: is innovation a missing link? *Journal of marketing*, 62(4):30–45.

Hansen, M. T. and Løvås, B. (2004). How do multinational companies leverage technological competencies? moving from single to interdependent explanations. *Strategic Management Journal*, 25(8-9):801–822.

Hargadon, A. B. and Douglas, Y. (2001). When innovations meet institutions: Edison and the design of the electric light. *Administrative science quarterly*, 46(3):476–501.

Harmancioglu, N., Grinstein, A., and Goldman, A. (2010). Innovation and performance outcomes of market information collection efforts: The role of top management team involvement. *International Journal of Research in Marketing*, 27(1):33–43.

Hauser, J., Tellis, G. J., and Griffin, A. (2006). Research on innovation: A review and agenda for marketing science. *Marketing science*, 25(6):687–717.

He, Z.-L. and Wong, P.-K. (2004). Exploration vs. exploitation: An empirical test of the ambidexterity hypothesis. *Organization science*, 15(4):481–494.

Henard, D. H. and Szymanski, D. M. (2001). Why some new products are more successful than others. *Journal of marketing Research*, 38(3):362–375.

Herrera, M. E. B. (2015). Creating competitive advantage by institutionalizing corporate social innovation. *Journal of business research*, 68(7):1468–1474.

Hirst, G., Van Knippenberg, D., and Zhou, J. (2009). A cross-level perspective on employee creativity: Goal orientation, team learning behavior, and individual creativity. *Academy of management journal*, 52(2):280–293.

Hitt, M. A., Ireland, R. D., Camp, S. M., and Sexton, D. L. (2001). Strategic entrepreneurship: Entrepreneurial strategies for wealth creation. *Strategic management journal*, 22(6-7):479–491.

Hoffman, D. L., Kopalle, P. K., and Novak, T. P. (2010). The “right” consumers for better concepts: Identifying consumers high in emergent nature to develop new product concepts. *Journal of Marketing Research*, 47(5):854–865.

Hogan, S. J. and Coote, L. V. (2014). Organizational culture, innovation, and performance: A test of schein’s model. *Journal of business research*, 67(8):1609–1621.

Homburg, C. and Pflesser, C. (2000). A multiple-layer model of market-oriented organizational culture: Measurement issues and performance outcomes. *Journal of marketing research*, 37(4):449–462.

Hostede, G. and Hofstede, G. (1991). Cultures and organizations: Software of the mind. *New York: McGrawHill*, 1997:23–47.

Hovy, D. (2020). *Text Analysis in Python for Social Scientists: Discovery and Exploration*. Cambridge University Press.

Hughes, M. and Morgan, R. E. (2007). Deconstructing the relationship between entrepreneurial orientation and business performance at the embryonic stage of firm growth. *Industrial marketing management*, 36(5):651–661.

Hult, G. T. M., Hurley, R. F., and Knight, G. A. (2004). Innovativeness: Its antecedents and impact on business performance. *Industrial marketing management*, 33(5):429–438.

Hult, G. T. M. and Ketchen Jr, D. J. (2001). Does market orientation matter?: A test of the relationship between positional advantage and performance. *Strategic management journal*, 22(9):899–906.

Hult, G. T. M., Ketchen Jr, D. J., and Nichols Jr, E. L. (2002). An examination of cultural competitiveness and order fulfillment cycle time within supply chains. *Academy of management Journal*, 45(3):577–586.

Hult, G. T. M., Snow, C. C., and Kandemir, D. (2003). The role of entrepreneurship in building cultural competitiveness in different organizational types. *Journal of management*, 29(3):401–426.

Hultink, E. J., Griffin, A., Hart, S., and Robben, H. S. (1997). Industrial new product launch strategies and product development performance. *Journal of product innovation management*, 14(4):243–257.

Hurley, R. F. and Hult, G. T. M. (1998). Innovation, market orientation, and organizational learning: an integration and empirical examination. *Journal of marketing*, 62(3):42–54.

Inglehart, R. (1997). *Modernization and Postmodernization: Cultural, Economic, and Political Change in 43 Societies*. Princeton University Press.

Ireland, R. D., Hitt, M. A., and Sirmon, D. G. (2003). A model of strategic entrepreneurship: The construct and its dimensions. *Journal of management*, 29(6):963–989.

Jacob, M., Lundqvist, M., and Hellsmark, H. (2003). Entrepreneurial transformations in the swedish university system: the case of chalmers university of technology. *Research policy*, 32(9):1555–1568.

Jansen, J. J., Tempelaar, M. P., Van den Bosch, F. A., and Volberda, H. W. (2009). Structural differentiation and ambidexterity: The mediating role of integration mechanisms. *Organization science*, 20(4):797–811.

Jansen, J. J., Van Den Bosch, F. A., and Volberda, H. W. (2006). Exploratory innovation, exploitative innovation, and performance: Effects of organizational antecedents and environmental moderators. *Management science*, 52(11):1661–1674.

Jean, R.-J. B., Sinkovics, R. R., and Cavusgil, S. T. (2010). Enhancing international customer–supplier relationships through it resources: A study of taiwanese electronics suppliers. *Journal of International Business Studies*, 41(7):1218–1239.

Johnsen, R. E. and Ford, D. (2006). Interaction capability development of smaller suppliers in relationships with larger customers. *Industrial marketing management*, 35(8):1002–1015.

Kahn, K. B. (2001). Market orientation, interdepartmental integration, and product development performance. *Journal of Product Innovation Management: AN INTERNATIONAL PUBLICATION OF THE PRODUCT DEVELOPMENT & MANAGEMENT ASSOCIATION*, 18(5):314–323.

Kahn, K. B., Barczak, G., Nicholas, J., Ledwith, A., and Perks, H. (2012). An examination of new product development best practice. *Journal of product innovation management*, 29(2):180–192.

Kang, Y., Cai, Z., Tan, C.-W., Huang, Q., and Liu, H. (2020). Natural language processing (nlp) in management research: A literature review. *Journal of Management Analytics*, 7(2):139–172.

Katila, R. (2002). New product search over time: past ideas in their prime? *Academy of Management journal*, 45(5):995–1010.

Katila, R. and Shane, S. (2005). When does lack of resources make new firms innovative? *Academy of Management Journal*, 48(5):814–829.

Kester, L., Griffin, A., Hultink, E. J., and Lauche, K. (2011). Exploring portfolio decision-making processes. *Journal of product innovation management*, 28(5):641–661.

Kester, L., Hultink, E. J., and Griffin, A. (2014). An empirical investigation of the antecedents and outcomes of npd portfolio success. *Journal of Product Innovation Management*, 31(6):1199–1213.

Kirca, A. H., Jayachandran, S., and Bearden, W. O. (2005). Market orientation: A meta-analytic review and assessment of its antecedents and impact on performance. *Journal of marketing*, 69(2):24–41.

Kleinschmidt, E. J., De Brentani, U., and Salomo, S. (2007). Performance of global new product development programs: a resource-based view. *Journal of Product Innovation Management*, 24(5):419–441.

Knight, G. A. and Cavusgil, S. T. (2004). Innovation, organizational capabilities, and the born-global firm. *Journal of international business studies*, 35(2):124–141.

Kochhar, R. and David, P. (1996). Institutional investors and firm innovation: A test of competing hypotheses. *Strategic management journal*, 17(1):73–84.

Kohli, A. K. and Jaworski, B. J. (1990). Market orientation: the construct, research propositions, and managerial implications. *Journal of marketing*, 54(2):1–18.

Kohli, A. K., Jaworski, B. J., and Kumar, A. (1993). Markor: a measure of market orientation. *Journal of Marketing research*, 30(4):467–477.

Kotha, S., Rindova, V. P., and Rothaermel, F. T. (2001). Assets and actions: Firm-specific factors in the internationalization of us internet firms. *Journal of International Business Studies*, 32(4):769–791.

Kuhlmann, S. (2001). Future governance of innovation policy in europe—three scenarios. *Research policy*, 30(6):953–976.

Kumar, K., Subramanian, R., and Yauger, C. (1998). Examining the market orientation-performance relationship: a context-specific study. *Journal of management*, 24(2):201–233.

Kumar, V. and Krishnan, T. V. (2002). Multinational diffusion models: An alternative framework. *Marketing Science*, 21(3):318–330.

Kyriakopoulos, K. and Moorman, C. (2004). Tradeoffs in marketing exploitation and exploration strategies: The overlooked role of market orientation. *International Journal of Research in Marketing*, 21(3):219–240.

Laforet, S. (2008). Size, strategic, and market orientation affects on innovation. *Journal of business Research*, 61(7):753–764.

Langerak, F., Hultink, E. J., and Robben, H. S. (2004). The impact of market orientation, product advantage, and launch proficiency on new product performance and organizational performance. *Journal of product innovation management*, 21(2):79–94.

Laursen, K. and Salter, A. (2006). Open for innovation: the role of openness in explaining innovation performance among uk manufacturing firms. *Strategic management journal*, 27(2):131–150.

Lazzeretti, L. and Capone, F. (2016). How proximity matters in innovation networks dynamics along the cluster evolution. a study of the high technology applied to cultural goods. *Journal of Business Research*, 69(12):5855–5865.

Leal-Rodríguez, A. L., Ariza-Montes, J. A., Roldán, J. L., and Leal-Millán, A. G. (2014). Absorptive capacity, innovation and cultural barriers: A conditional mediation model. *Journal of Business Research*, 67(5):763–768.

Lee, R. P., Spanjol, J., and Sun, S. L. (2019). Social innovation in an interconnected world: Introduction to the special issue.

Lemola, T. (2002). Convergence of national science and technology policies: the case of finland. *Research policy*, 31(8-9):1481–1490.

Li, T. and Calantone, R. J. (1998). The impact of market knowledge competence on new product advantage: conceptualization and empirical examination. *Journal of marketing*, 62(4):13–29.

Lin, H.-E., McDonough III, E. F., Lin, S.-J., and Lin, C. Y.-Y. (2013). Managing the exploitation/exploration paradox: The role of a learning capability and innovation ambidexterity. *Journal of Product Innovation Management*, 30(2):262–278.

Lisboa, A., Skarmeas, D., and Lages, C. (2011). Innovative capabilities: Their drivers and effects on current and future performance. *Journal of Business Research*, 64(11):1157–1161.

Luk, C.-L., Yau, O. H., Sin, L. Y., Alan, C., Chow, R. P., and Lee, J. S. (2008). The effects of social capital and organizational innovativeness in different institutional contexts. *Journal of International Business Studies*, 39(4):589–612.

Lukas, B. A. and Ferrell, O. C. (2000). The effect of market orientation on product innovation. *Journal of the academy of marketing science*, 28(2):239–247.

Lumpkin, G. T. and Dess, G. G. (1996). Clarifying the entrepreneurial orientation construct and linking it to performance. *Academy of management Review*, 21(1):135–172.

Madsen, P. M. and Desai, V. (2010). Failing to learn? the effects of failure and success on organizational learning in the global orbital launch vehicle industry. *Academy of management journal*, 53(3):451–476.

Manning, C. and Schutze, H. (1999). *Foundations of statistical natural language processing*. MIT press.

Manu, F. and Sriram, V. (1996). Innovation, marketing and performance strategy, environment. *Journal of Business Research*, 35:79–91.

Markham, S. K. and Lee, H. (2013). P roduct d evelopment and m anagement a ssociation's 2012 c omparative p erformance a ssessment s tudy. *Journal of Product Innovation Management*, 30(3):408–429.

Martin, J. H. and Grbac, B. (2003). Using supply chain management to leverage a firm's market orientation. *Industrial marketing management*, 32(1):25–38.

Matsuno, K. and Mentzer, J. T. (2000). The effects of strategy type on the market orientation–performance relationship. *Journal of marketing*, 64(4):1–16.

Matsuno, K., Mentzer, J. T., and Özsomer, A. (2002). The effects of entrepreneurial proclivity and market orientation on business performance. *Journal of marketing*, 66(3):18–32.

McDermott, C. M. and O'connor, G. C. (2002). Managing radical innovation: an overview of emergent strategy issues. *Journal of Product Innovation Management: an international publication of the product development & management association*, 19(6):424–438.

Menguc, B., Auh, S., and Shih, E. (2007). Transformational leadership and market orientation: Implications for the implementation of competitive strategies and business unit performance. *Journal of business research*, 60(4):314–321.

Menon, A., Chowdhury, J., and Lukas, B. A. (2002). Antecedents and outcomes of new product development speed: An interdisciplinary conceptual framework. *Industrial Marketing Management*, 31(4):317–328.

Meyer-Krahmer, F. and Schmoch, U. (1998). Science-based technologies: university–industry interactions in four fields. *Research policy*, 27(8):835–851.

Mirvis, P., Herrera, M. E. B., Googins, B., and Albareda, L. (2016). Corporate social innovation: How firms learn to innovate for the greater good. *Journal of Business Research*, 69(11):5014–5021.

Moenaert, R. K., Caeldries, F., Lievens, A., and Wauters, E. (2000). Communication flows in international product innovation teams. *Journal of Product Innovation Management:*

An International Publication of the Product Development & Management Association, 17(5):360–377.

Moorman, C. (2018). Call for papers | journal of marketing special issue: Better marketing for a better world.

Moorman, C. and Miner, A. S. (1997). The impact of organizational memory on new product performance and creativity. *Journal of marketing research*, 34(1):91–106.

Morgan, R. E. and Berthon, P. (2008). Market orientation, generative learning, innovation strategy and business performance inter-relationships in bioscience firms. *Journal of management studies*, 45(8):1329–1353.

Mueller, V., Rosenbusch, N., and Bausch, A. (2013). Success patterns of exploratory and exploitative innovation: A meta-analysis of the influence of institutional factors. *Journal of Management*, 39(6):1606–1636.

Nakata, C. and Sivakumar, K. (1996). National culture and new product development: An integrative review. *Journal of Marketing*, 60(1):61–72.

Narver, J. C. and Slater, S. F. (1990). The effect of a market orientation on business profitability. *Journal of marketing*, 54(4):20–35.

Narver, J. C., Slater, S. F., and MacLachlan, D. L. (2004). Responsive and proactive market orientation and new-product success. *Journal of product innovation management*, 21(5):334–347.

Ngo, L. V. and O'Cass, A. (2012). In search of innovation and customer-related performance superiority: The role of market orientation, marketing capability, and innovation capability interactions. *Journal of Product Innovation Management*, 29(5):861–877.

Noble, C. H., Sinha, R. K., and Kumar, A. (2002). Market orientation and alternative strategic orientations: a longitudinal assessment of performance implications. *Journal of marketing*, 66(4):25–39.

Noble, C. H., Spanjol, J., and Kirca, A. H. (2021). Advancing broad and deep understanding in innovation management: Meta-analyses and literature reviews.

O'Connor, G. C. (1998). Market learning and radical innovation: A cross case comparison of eight radical innovation projects. *Journal of Product Innovation Management: An International Publication of The Product Development & Management Association*, 15(2):151–166.

O'Connor, G. C. (2008). Major innovation as a dynamic capability: A systems approach. *Journal of product innovation management*, 25(4):313–330.

O'Connor, G. C. and Rice, M. P. (2013). A comprehensive model of uncertainty associated with radical innovation. *Journal of Product Innovation Management*, 30:2–18.

O'Connor, G. C. and Veryzer, R. W. (2001). The nature of market visioning for technology-based radical innovation. *Journal of Product Innovation Management: An International Publication of the Product Development & Management Association*, 18(4):231–246.

Olson, E. L. (2013). Perspective: the green innovation value chain: a tool for evaluating the diffusion prospects of green products. *Journal of Product Innovation Management*, 30(4):782–793.

Olson, E. M., Slater, S. F., and Hult, G. T. M. (2005). The performance implications of fit among business strategy, marketing organization structure, and strategic behavior. *Journal of marketing*, 69(3):49–65.

Olson, E. M., Walker Jr, O. C., Ruekerf, R. W., and Bonnerd, J. M. (2001). Patterns of cooperation during new product development among marketing, operations and r&d: Implications for project performance. *Journal of Product Innovation Management: An International Publication of The Product Development & Management Association*, 18(4):258–271.

Ozkaya, H. E., Droge, C., Hult, G. T. M., Calantone, R., and Ozkaya, E. (2015). Market orientation, knowledge competence, and innovation. *International Journal of Research in Marketing*, 32(3):309–318.

Paladino, A. (2007). Investigating the drivers of innovation and new product success: a comparison of strategic orientations. *Journal of Product Innovation Management*, 24(6):534–553.

Paladino, A. (2008). Analyzing the effects of market and resource orientations on innovative outcomes in times of turbulence. *Journal of Product Innovation Management*, 25(6):577–592.

Palich, L. E., Cardinal, L. B., and Miller, C. C. (2000). Curvilinearity in the diversification–performance linkage: an examination of over three decades of research. *Strategic management journal*, 21(2):155–174.

Parry, M. E., Song, M., De Weerd-Nederhof, P. C., and Visscher, K. (2009). The impact of npd strategy, product strategy, and npd processes on perceived cycle time. *Journal of Product Innovation Management*, 26(6):627–639.

Paswan, A. K. and Wittmann, C. M. (2009). Knowledge management and franchise systems. *Industrial Marketing Management*, 38(2):173–180.

Pauwels, K., Silva-Risso, J., Srinivasan, S., and Hanssens, D. M. (2004). New products, sales promotions, and firm value: The case of the automobile industry. *Journal of Marketing*, 68(4):142–156.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830.

Pelham, A. M. and Wilson, D. T. (1995). A longitudinal study of the impact of market structure, firm structure, strategy, and market orientation culture on dimensions of small-firm performance. *Journal of the academy of marketing science*, 24(1):27–43.

Poskela, J. and Martinsuo, M. (2009). Management control and strategic renewal in the front end of innovation. *Journal of Product Innovation Management*, 26(6):671–684.

Putsis Jr, W. P., Balasubramanian, S., Kaplan, E. H., and Sen, S. K. (1997). Mixing behavior in cross-country diffusion. *Marketing Science*, 16(4):354–369.

Radjou, N., Prabhu, J., et al. (2015). *Frugal Innovation: How to do more with less.* The Economist.

Raisch, S., Birkinshaw, J., Probst, G., and Tushman, M. L. (2009). Organizational ambidexterity: Balancing exploitation and exploration for sustained performance. *Organization science*, 20(4):685–695.

Reid, S. E. and De Brentani, U. (2004). The fuzzy front end of new product development for discontinuous innovations: A theoretical model. *Journal of product innovation management*, 21(3):170–184.

Richard, O. C., Barnett, T., Dwyer, S., and Chadwick, K. (2004). Cultural diversity in management, firm performance, and the moderating role of entrepreneurial orientation dimensions. *Academy of management journal*, 47(2):255–266.

Ritter, T. and Gemünden, H. G. (2003). Network competence: Its impact on innovation success and its antecedents. *Journal of business research*, 56(9):745–755.

Rodríguez-Pinto, J., Carbonell, P., and Rodríguez-Escudero, A. I. (2011). Speed or quality? how the order of market entry influences the relationship between market orientation and new product performance. *International Journal of Research in Marketing*, 28(2):145–154.

Rothenberg, S. (2003). Knowledge content and worker participation in environmental management at nummi. *Journal of management studies*, 40(7):1783–1802.

Rubera, G. and Kirca, A. H. (2012). Firm innovativeness and its performance outcomes: A meta-analytic review and theoretical integration. *Journal of Marketing*, 76(3):130–147.

Rugman, A. M. (1981). *Inside the multinationals: The economics of internal markets*. Columbia University Press, New York.

Rugman, A. M., Lecraw, D. J., Booth, L. D., et al. (1985). *International business: Firm and environment*. McGraw-Hill College.

Sandvik, I. L. and Sandvik, K. (2003). The impact of market orientation on product innovativeness and business performance. *International journal of Research in Marketing*, 20(4):355–376.

Schein, E. H. (1986). Organizational culture and leadership: A dynamic view 1985. *San Francisco: Jossey-Bass. Organization Studies*, 7(2):199–201.

Schwartz, S. H. (1994). Are there universal aspects in the structure and contents of human values? *Journal of social issues*, 50(4):19–45.

Semrau, T., Ambos, T., and Kraus, S. (2016). Entrepreneurial orientation and sme performance across societal cultures: An international study. *Journal of Business Research*, 69(5):1928–1932.

Sethi, R. (2000). New product quality and product development teams. *Journal of marketing*, 64(2):1–14.

Sethi, R., Smith, D. C., and Park, C. W. (2001). Cross-functional product development teams, creativity, and the innovativeness of new consumer products. *Journal of marketing research*, 38(1):73–85.

Sharma, A. and Lacey, N. (2004). Linking product development outcomes to market valuation of the firm: The case of the us pharmaceutical industry. *Journal of Product Innovation Management*, 21(5):297–308.

Shenkar, O., Tallman, S. B., Wang, H., and Wu, J. (2020). National culture and international business: A path forward. *Journal of International Business Studies*, pages 1–18.

Sherman, J. D., Souder, W. E., and Jenssen, S. A. (2000). Differential effects of the primary forms of cross functional integration on product development cycle time. *Journal of Product Innovation Management: An International Publication of the Product Development & Management Association*, 17(4):257–267.

Shujahat, M., Sousa, M. J., Hussain, S., Nawaz, F., Wang, M., and Umer, M. (2019). Translating the impact of knowledge management processes into knowledge-based innovation: The neglected and mediating role of knowledge-worker productivity. *Journal of Business Research*, 94:442–450.

Sievert, C. and Shirley, K. (2014). LDAvis: A method for visualizing and interpreting topics. In *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, pages 63–70, Baltimore, Maryland, USA. Association for Computational Linguistics.

Siguaw, J. A., Simpson, P. M., and Enz, C. A. (2006). Conceptualizing innovation orientation: A framework for study and integration of innovation research. *Journal of product innovation management*, 23(6):556–574.

Simpson, P. M., Siguaw, J. A., and Enz, C. A. (2006). Innovation orientation outcomes: The good and the bad. *Journal of Business Research*, 59(10-11):1133–1141.

Sivadas, E. and Dwyer, F. R. (2000). An examination of organizational factors influencing new product success in internal and alliance-based processes. *Journal of marketing*, 64(1):31–49.

Slater, S. F., Mohr, J. J., and Sengupta, S. (2014). Radical product innovation capability: Literature review, synthesis, and illustrative research propositions. *Journal of Product Innovation Management*, 31(3):552–566.

Slater, S. F. and Narver, J. C. (1998). Customer-led and market-oriented: let's not confuse the two. *Strategic management journal*, 19(10):1001–1006.

Slater, S. F. and Narver, J. C. (1999). Market-oriented is more than being customer-led. *Strategic management journal*, 20(12):1165–1168.

Smith, W. K. and Tushman, M. L. (2005). Managing strategic contradictions: A top management model for managing innovation streams. *Organization science*, 16(5):522–536.

Song, M. and Xie, J. (2000). Does innovativeness moderate the relationship between cross-functional integration and product performance? *Journal of international marketing*, 8(4):61–89.

Song, X. M. and Parry, M. E. (1997). A cross-national comparative study of new product development processes: Japan and the united states. *Journal of marketing*, 61(2):1–18.

Sood, A. and Tellis, G. J. (2009). Do innovations really pay off? total stock market returns to innovation. *Marketing Science*, 28(3):442–456.

Sorescu, A. B., Chandy, R. K., and Prabhu, J. C. (2003). Sources and financial consequences of radical innovation: Insights from pharmaceuticals. *Journal of marketing*, 67(4):82–102.

Sorescu, A. B. and Spanjol, J. (2008). Innovation's effect on firm value and risk: Insights from consumer packaged goods. *Journal of Marketing*, 72(2):114–132.

Souder, W. E. and Song, X. M. (1997). Contingent product design and marketing strategies influencing new product success and failure in us and japanese electronics firms. *Journal of Product Innovation Management: An International Publication of the Product Development & Management Association*, 14(1):21–34.

Souder, W. E. and Song, X. M. (1998). Analyses of us and japanese management processes associated with new product success and failure in high and low familiarity markets. *Journal of Product Innovation Management: AN INTERNATIONAL PUBLICATION OF THE PRODUCT DEVELOPMENT & MANAGEMENT ASSOCIATION*, 15(3):208–223.

Srinivasan, R., Lilien, G. L., and Rangaswamy, A. (2002). Technological opportunism and radical technology adoption: An application to e-business. *Journal of marketing*, 66(3):47–60.

Stam, W. and Elfring, T. (2008). Entrepreneurial orientation and new venture performance: The moderating role of intra-and extraindustry social capital. *Academy of management journal*, 51(1):97–111.

Steenkamp, J.-B. E. and Geyskens, I. (2014). Manufacturer and retailer strategies to impact store brand share: Global integration, local adaptation, and worldwide learning. *Marketing Science*, 33(1):6–26.

Steenkamp, J.-B. E., Ter Hofstede, F., and Wedel, M. (1999). A cross-national investigation into the individual and national cultural antecedents of consumer innovativeness. *Journal of marketing*, 63(2):55–69.

Steensma, H. K., Marino, L., Weaver, K. M., and Dickson, P. H. (2000). The influence of national culture on the formation of technology alliances by entrepreneurial firms. *Academy of management journal*, 43(5):951–973.

Stock, R. M., Six, B., and Zacharias, N. A. (2013). Linking multiple layers of innovation-oriented corporate culture, product program innovativeness, and business performance: A contingency approach. *Journal of the Academy of Marketing Science*, 41(3):283–299.

Storey, C., Cankurtaran, P., Papastathopoulou, P., and Hultink, E. J. (2016). Success factors for service innovation: A meta-analysis. *Journal of Product Innovation Management*, 33(5):527–548.

Story, V. M., Raddats, C., Burton, J., Zolkiewski, J., and Baines, T. (2017). Capabilities for advanced services: A multi-actor perspective. *Industrial Marketing Management*, 60:54–68.

Stremersch, S. and Tellis, G. J. (2004). Understanding and managing international growth of new products. *International journal of Research in Marketing*, 21(4):421–438.

Subramaniam, M. and Youndt, M. A. (2005). The influence of intellectual capital on the types of innovative capabilities. *Academy of Management journal*, 48(3):450–463.

Sundqvist, S., Frank, L., and Puumalainen, K. (2005). The effects of country characteristics, cultural similarity and adoption timing on the diffusion of wireless communications. *Journal of business research*, 58(1):107–110.

Swift, A., Heale, R., and Twycross, A. (2020). What are sensitivity and specificity? *Evidence-Based Nursing*, 23(1):2–4.

Swink, M. (2000). Technological innovativeness as a moderator of new product design integration and top management support. *Journal of Product Innovation Management: An International Publication of the Product Development & Management Association*, 17(3):208–220.

Szymanski, D. M., Kroff, M. W., and Troy, L. C. (2007). Innovativeness and new product success: Insights from the cumulative evidence. *Journal of the academy of Marketing Science*, 35(1):35–52.

Talke, K., Salomo, S., and Kock, A. (2011). Top management team diversity and strategic innovation orientation: The relationship and consequences for innovativeness and performance. *Journal of product innovation management*, 28(6):819–832.

Talukdar, D., Sudhir, K., and Ainslie, A. (2002). Investigating new product diffusion across products and countries. *Marketing Science*, 21(1):97–114.

Tatikonda, M. V. and Montoya-Weiss, M. M. (2001). Integrating operations and marketing perspectives of product innovation: The influence of organizational process factors and capabilities on development performance. *Management science*, 47(1):151–172.

Taylor, A. and Greve, H. R. (2006). Superman or the fantastic four? knowledge combination and experience in innovative teams. *Academy of management journal*, 49(4):723–740.

Teece, D. J. (2007). Explicating dynamic capabilities: the nature and microfoundations of (sustainable) enterprise performance. *Strategic management journal*, 28(13):1319–1350.

Tellis, G. J., Prabhu, J. C., and Chandy, R. K. (2009). Radical innovation across nations: The preeminence of corporate culture. *Journal of marketing*, 73(1):3–23.

Tellis, G. J., Stremersch, S., and Yin, E. (2003). The international takeoff of new products: The role of economics, culture, and country innovativeness. *Marketing Science*, 22(2):188–208.

Terwiesch, C. and Xu, Y. (2008). Innovation contests, open innovation, and multiagent problem solving. *Management science*, 54(9):1529–1543.

Terziovski, M. (2010). Innovation practice and its performance implications in small and medium enterprises (smes) in the manufacturing sector: a resource-based view. *Strategic management journal*, 31(8):892–902.

Theoharakis, V. and Hooley, G. (2008). Customer orientation and innovativeness: Differing roles in new and old europe. *International Journal of Research in Marketing*, 25(1):69–79.

Thomas, A. S. and Mueller, S. L. (2000). A case for comparative entrepreneurship: Assessing the relevance of culture. *Journal of international business studies*, 31(2):287–301.

Thompenaars, F. and Hampden-Turner, C. (1998). *Riding the waves of culture: Understanding cultural diversity in global business*. McGraw-Hill.

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

Toubia, O., Iyengar, G., Bunnell, R., and Lemaire, A. (2019). Extracting features of entertainment products: A guided latent dirichlet allocation approach informed by the psychology of media consumption. *Journal of Marketing Research*, 56(1):18–36.

Tsai, W. (2001). Knowledge transfer in intraorganizational networks: Effects of network position and absorptive capacity on business unit innovation and performance. *Academy of management journal*, 44(5):996–1004.

Tyler, B. B. and Gnyawali, D. R. (2002). Mapping managers' market orientations regarding new product success. *Journal of Product Innovation Management: AN INTERNATIONAL PUBLICATION OF THE PRODUCT DEVELOPMENT & MANAGEMENT ASSOCIATION*, 19(4):259–276.

Van den Bulte, C. (2000). New product diffusion acceleration: Measurement and analysis. *Marketing Science*, 19(4):366–380.

Van den Bulte, C. and Stremersch, S. (2004). Social contagion and income heterogeneity in new product diffusion: A meta-analytic test. *Marketing Science*, 23(4):530–544.

Van der Vegt, G. S., Van de Vliert, E., and Huang, X. (2005). Location-level links between diversity and innovative climate depend on national power distance. *Academy of Management Journal*, 48(6):1171–1182.

Vanloqueren, G. and Baret, P. V. (2009). How agricultural research systems shape a technological regime that develops genetic engineering but locks out agroecological innovations. *Research Policy*, 6(38):971–983.

Vassallo, J. P., Prabhu, J. C., Banerjee, S., and Voola, R. (2019). The role of hybrid organizations in scaling social innovations in bottom-of-the-pyramid markets: Insights from microfinance in india. *Journal of Product Innovation Management*, 36(6):744–763.

Vera, D. and Crossan, M. (2005). Improvisation and innovative performance in teams. *Organization science*, 16(3):203–224.

Verganti, R. (2008). Design, meanings, and radical innovation: A metamodel and a research agenda. *Journal of product innovation management*, 25(5):436–456.

Vicente-Saez, R. and Martinez-Fuentes, C. (2018). Open science now: A systematic literature review for an integrated definition. *Journal of business research*, 88:428–436.

Visnjic, I., Wiengarten, F., and Neely, A. (2016). Only the brave: Product innovation, service business model innovation, and their impact on performance. *Journal of product innovation management*, 33(1):36–52.

Wang, C., Blei, D., and Heckerman, D. (2012). Continuous time dynamic topic models. *arXiv preprint arXiv:1206.3298*.

Wei, Y. and Morgan, N. A. (2004). Supportiveness of organizational climate, market orientation, and new product performance in chinese firms. *Journal of product innovation management*, 21(6):375–388.

Wei, Y. S. and Atuahene-Gima, K. (2009). The moderating role of reward systems in the relationship between market orientation and new product performance in china. *International Journal of Research in Marketing*, 26(2):89–96.

West, J. and Bogers, M. (2014). Leveraging external sources of innovation: a review of research on open innovation. *Journal of product innovation management*, 31(4):814–831.

Westkämper, E. and Walter, F. (2014). Towards the re-industrialization of europe. *A Concept for Manufacturing for*, 2030:1–111.

Wijnberg, N. M. and Gemser, G. (2000). Adding value to innovation: Impressionism and the transformation of the selection system in visual arts. *Organization science*, 11(3):323–329.

Wiklund, J. and Shepherd, D. (2003). Knowledge-based resources, entrepreneurial orientation, and the performance of small and medium-sized businesses. *Strategic management journal*, 24(13):1307–1314.

Wilkie, W. L. and Moore, E. S. (1999). Marketing's contributions to society. *Journal of marketing*, 63(4_suppl1):198–218.

Woodside, A. G. (2005). Firm orientations, innovativeness, and business performance: Advancing a system dynamics view following a comment on hult, hurley, and knight's 2004 study. *Industrial marketing management*, 34(3):275–279.

Wright, C., Sturdy, A., and Wylie, N. (2012). Management innovation through standardization: Consultants as standardizers of organizational practice. *Research Policy*, 41(3):652–662.

Xie, J., Song, M., and Stringfellow, A. (2003). Antecedents and consequences of goal incongruity on new product development in five countries: A marketing view. *Journal of Product Innovation Management*, 20(3):233–250.

Xie, J., Song, X. M., and Stringfellow, A. (1998). Interfunctional conflict, conflict resolution styles, and new product success: A four-culture comparison. *Management science*, 44(12-part-2):S192–S206.

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

Yli-Renko, H., Autio, E., and Sapienza, H. J. (2001). Social capital, knowledge acquisition, and knowledge exploitation in young technology-based firms. *Strategic management journal*, 22(6-7):587–613.

Zhao, Y. and Cavusgil, S. T. (2006). The effect of supplier's market orientation on manufacturer's trust. *Industrial Marketing Management*, 35(4):405–414.

Zheng, W., Yang, B., and McLean, G. N. (2010). Linking organizational culture, structure, strategy, and organizational effectiveness: Mediating role of knowledge management. *Journal of Business research*, 63(7):763–771.

Zhong, N. and Schweidel, D. A. (2020). Capturing changes in social media content: a multiple latent changepoint topic model. *Marketing Science*, 39(4):827–846.

Zhou, K. Z., Brown, J. R., and Dev, C. S. (2009). Market orientation, competitive advantage, and performance: A demand-based perspective. *Journal of business research*, 62(11):1063–1070.

Zhou, K. Z. and Wu, F. (2010). Technological capability, strategic flexibility, and product innovation. *Strategic management journal*, 31(5):547–561.

Zhou, K. Z., Yim, C. K., and Tse, D. K. (2005). The effects of strategic orientations on technology-and market-based breakthrough innovations. *Journal of marketing*, 69(2):42–60.

Zien, K. A. and Buckler, S. A. (1997). From experience dreams to market: crafting a culture of innovation. *Journal of Product Innovation Management: An International Publication of the Product Development & Management Association*, 14(4):274–287.

Chapter 3

Disentangling the “echoverse” for brand communication

Serena Pugliese, Gaia Rubera, Giovanna Padula

3.1 Abstract

Kuksov et al. (2013) suggest that consumers’ communication about the brand does not initiate without a brand first communication that provides novel information about the product. This "first-mover advantage" in communication should allow the brand to set conversation content. However, consumers tend to diverge from brand content (Kozinets et al., 2010). This consumers’ tendency creates a constant tension for brands between the need to keep up with its consumers to keep them engaged, encouraging the brand to follow the conversation direction proposed by consumers, and the need to keep control of its own narrative. A similar tension has been observed in politics (Barberá et al., 2019) where politicians equally face the need to convince people of their point of view while at the same time address relevant issues for their voters. In this research, we try to understand which actor, between brand and consumers, is more effective in determining

the content of the interaction. To understand the dynamic of their interaction, we plan to observe the content of brands' social media posts and consumers' social media posts for ten years in terms of the evolution of the topic discussed. In particular, we are interested in determining whether the topic of the conversation tends to be set by consumers or brands.

3.2 Introduction and Theoretical Background

The abrupt growth of social media in the last decade has substantially reshaped the communication environment of the firms (Hewett et al., 2016). A key aspect of the rise of social media has been the emergence of new ways for consumers to engage with brands and interact with other consumers (Lamberton and Stephen, 2016). Indeed, consumers have been increasingly engaging in new forms of online WOM, giving rise to a networked co-production of narratives (Kozinets et al., 2010) that help consumers make emotional and personal connections with brands, thus contributing to shaping brand personality (Lee et al., 2018) and affect brand reputation (Rust et al., 2021). At the same time, firms have been increasingly using social media as new communication platforms seeding large-scale online WOM marketing and viral campaigns as well as implementing targeted digital advertising (Lamberton and Stephen, 2016). In this context, social network sites have become new places for interactivity and content-delivery spreading conversations about brands, where consumers have accrued their social influence power, and firms have been increasingly using consumers' online activities and content generation as communications tools (Lamberton and Stephen, 2016). As a consequence, social media have been creating a reverberating "echoverse" for brand communication, with each actor contributing and being influenced by each other's actions and with online WOM becoming increasingly central (Hewett et al., 2016).

This shift in the communication environment, brought by the advent of social media, paved the way for a new line of inquiry investigating the impact of consumers' online content generation and WOM on firm performance (Moe and Trusov, 2011; Stephen and Galak, 2012; Tirunillai and Tellis, 2012; Stephen and Galak, 2012; Nam and Kannan, 2014; Srinivasan et al., 2016; Babić Rosario et al., 2016; Kumar et al., 2016; Srinivasan et al., 2016). In their attempt to provide a more nuanced theory on the social media – firm performance relationship, other scholars have called for more emphasis on the relationship between social media and consumer-related metrics, such as consumer acquisition (De Vries et al., 2017) or consumer mindset metrics (Colicev et al., 2018, 2019), for their impact on firm terminal financial performance (Petersen et al., 2018).

Though assessing the social media–firm performance relationship has spurred a great deal of interest among scholars, our knowledge of successfully managing consumers' online activities and content generation is still in its infancy. This paper aims to fill this gap by shifting the attention from the outcome to the process occurring within the social media communication environment. Investigating the consumers–firms' interactions within the brand conversation environment is of the utmost importance to understand how firms could better harness consumer power in social media to support their performance.

While it is well known that firm communications and online WOM spreading across consumers mutually influence each other (Hewett et al., 2016), it is still unknown whether firms devote more time to an issue before or after attention to that issue by consumers increases. In fact, on one side, the relevance of online WOM has urged marketing scholars to suggest the need for the firms to induce brand-related social media conversation either by participating in brand conversation through firms' generation contents(Kozinets et al., 2010; Babić Rosario et al., 2016) or through targeted advertising campaigns (Lambrecht et al., 2018). Emphasis on this firm-created or firm-seeded online WOM suggests that firms lead the brand conversation in the social media communication environment. On the other side, the general wisdom in marketing communication literature has always emphasized that firms' communication starts from the understanding of the target the communication is addressed to (Lasswell, 2017; John and Steven, 2002; Schramm, 1954). Research on social media argues that consumers use online WOM as a way of individual expression (for a review of this literature, see Lamberton and Stephen 2016). More recent empirical studies on consumers' online behavior find that the extent to which firm-seeded online WOM helps spur conversation around a brand is contingent on whether the message conveyed by the firms is consistent with the "trending" topics and helps enhance a user's self-presentation (Lambrecht et al., 2018). This alternative perspective suggests that firms follow the brand conversation in the social media communication environment. Understanding who leads and who follows the brand conversation in the social media communication environment is relevant in order to help the firms better exploit the social influence power of consumers on online social media.

Interesting evidence about how conversation content evolves on social media can be found within the political domain (Barberá et al., 2019), where it is equally important for certain users (i.e., politicians) to be responsive to their audience needs. To pursue this line of inquiry, we readapt their procedure to the industry domain. We mined Twitter data to assess how brand conversation–related topics emerged and evolved. We chose the Twitter platform because of its well-acknowledged opinion leadership role in shaping the minds of the general population (Hewett et al., 2016). Furthermore, Twitter is characterized by fast-paced and short-lived information flows so that new topics continuously emerge and fade (Lambrecht et al., 2018). Consequently, as we are interested in understanding who leads the emergence of new topics and how these topics evolve, Twitter represents a very fruitful social media platform to investigate our research topic.

3.3 Data Description

3.3.1 Brands

We focus on tweets from brands in the soft drinks industry. We collected new products from Mintel and brand market shares from Passport from 2008 to 2018. New products provide a reference list for new topics the brand might discuss to evaluate later the topics we will find. However, Mintel has a vast pool of brands. To narrow it down and focus on brands whose market presence is relevant enough to make them more likely to be also discussed on social media, we collected sales from Passport. We searched for the Twitter accounts of brands available on both Mintel and Passport. For these brands, we identified their English Twitter accounts. Given that a brand can have multiple national accounts, we identified between 1 and 8 accounts per brand. Our final sample of brands includes 268 brands and 460 accounts. For these 460 accounts, we scraped their entire user timeline (i.e., everything they shared on Twitter) until the end of 2018. These accounts account for 3.204.843 tweets.

3.3.2 Consumers

For consumers, we replicate the distinction suggested from Barberá et al. (2019) based on how involved they are with the industry we are interested in.

General consumers:

For consumers, we could not fully follow the procedure proposed by (Barberá et al., 2019). First, given that their domain is politics, they can assume that every citizen in the US can potentially be interested in topics relevant to politics and vice versa. So they randomly generated a sample of 25000 users' IDs and got their tweets if they were residents in the US. We cannot assume that any randomly generated Twitter user could be interested in our brands. Hence, to identify these accounts' relevant public, we focused on tweets mentioning the brands' accounts in our sample. From these tweets, we got a list of users (almost 6.3 mln users). Out of this list, we got a random sample of 10000 users that we treat as our general public. Out of these 10000 users, we collected the user timeline of 7788 users. The remaining 2212 were impossible to scrape because the account was discontinued or private. These users account for 52.584.113 tweets. The tweets retrieved were already labeled by language. The language seemed to be automatically detected, so it was not fully accurate. However, tweets incorrectly classified as non-English tended to be tweets with very few English words that would be mostly deleted after cleaning, for example, "YES!!! @mention @mention @mention". Hence, we filtered tweets by language and dropped all the non-English tweets. The remaining tweets are 42.385.228.

Attentive consumers:

As mentioned, Barberá et al. (2019) distinguishes between the general public and the attentive public. In their study, the attentive public includes 10000 randomly generated users that follow at least five major media outlets in the US. We can replicate this distinction by focusing on users in our list that tweeted about multiple accounts in our sample. Given our different time span (10 years instead of 2), we could not aim for the same sample size in terms of users, and we aimed for 10000 users in total.

To identify attentive users in our domain, we sorted the complete list of users from those who tweeted at the highest number of brands to the lowest and then removed the users that tweeted at less than 23 brands in our sample. We used 23 as the threshold because it would give a list of 2369 users, which would allow us to get to our desired sample size. This process gave us a sample of 2202 users, 119.917.585 tweets (109.795.259 without non-English tweets).

Table 3.1 summarizes the number of tweets per category of users in our sample. Note that the last three sets of tweets are partially overlapping: tweets mentioning the brand from users in the other two samples were collected twice because they must also be in their user timeline. Moreover, given that we have the location of brands' accounts, we

Category	Scraping criterium	N accounts	N tweets
Brands	Brand's user timeline	460	3.204.843
Tweets about brand	All tweets with "@brand" in text	6.297.684	21.309.972
General Public	User timeline of a random sample of users	7788	42.385.228
Attentive public	User timeline of users mentioning more than 22 brands	2205	109.795.259

Table 3.1: Tweets by user category

might consider to group brands and users by country, given that Twitter trending topics are generally local.

3.4 Tweets cleaning

After filtering the tweets for language, we pre-processed the text of the tweets. (Barberá et al., 2019) have a sample size in terms of tweets much smaller (39.692.180 in total, collected only between January 2013 and December 2014). Though they do not state it explicitly, they kept both hashtags and mentions as evident from the words for topics they reported. Given our sample size in terms of the number of tweets, including hashtags and mentions might be too dispersive in terms of dictionary also because hashtags can represent local trending topics. At the same time, we did not impose any restrictions on geography. So, we removed punctuation, links, hashtags, and mentions. We left only the brands' mentions, converting them into the brands' names. Keeping the brand names might help distinguish whether a topic originated as a general area of interest among consumers' discussions or was explicitly related to the brand. We also regularized words to reduce misspellings, removed stop words, lemmatized words, and removed words that were not nouns, adjectives, verbs, and adverbs.

3.5 Methodology

To study how conversations between brands and consumers evolve, we focus on topics discussed, whether they are consistent between brands and consumers' tweets, and, in that case, whether they were introduced in the conversation by brands or consumers. To identify the topic discussed, we use LDA, i.e., Latent Dirichlet Association (Blei et al., 2003). LDA has already been used in marketing and management research (Tirunillai and Tellis, 2014; Toubia et al., 2019), and it is a standard technique for topic modeling. It allows identifying the optimal number of different topics in a set of documents and the words with the strongest association with each topic. Documents in our setting are represented by the aggregated text of tweets shared on a specific day from each category of users (i.e., brands, generic consumers, and attentive consumers). Since not all the topics discussed by consumers in our sample are referred to the brands or the industry we are interested in, we first fitted LDA on brands' tweets, and then, always through LDA, we will estimate whether consumers' tweets refer to the same topics or not. We ran our model several times with a different number of topics (K), three levels of learning decay (0.5, 0.7, 0.9), using 10-fold cross-validation to identify the optimal number of topics. Then we evaluate their performance by looking at perplexity and log-likelihood. A preliminary test on tweets from brands from 2018 suggested ten as the optimal number of topics. Table 3.2 reports the most representative words for the topics identified.

These preliminary results already highlight some interesting trends. First of all, it is interesting to notice how "Pepsi" is on the same topic as its historical competitor "coke," suggesting that they are often discussed together. Moreover, the topic identified seems to suggest two main communication strategies. Some topics like topic 1 focus on tangible attributes or benefits of the product like the taste or discounts. Others, like topics 2 and 3, prefer to focus on communicating a certain identity like Kingsfisher's that what to be associated with sports(topic 3) or Mountain Dew with videogames.

	Topic 0	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9
Word 0	check	check	dew	kingfisher	thank	coffee	good	thank	point	tea
Word 1	thank	coupon	throwback	test	good	store	love	tweet	shop	thank
Word 2	day	free	check	cricketer	today	starbuck	flavor	coffee	double	good
Word 3	pepsi	pom	game	beer	great	rockstar	thank	today	nectar	new
Word 4	coke	juice	video	player	make	new	lifewater	day	triple	water
Word 5	good	thank	race	big	answer	today	think	know	collect	bottle
Word 6	great	recipe	good	play	day	card	try	donut	online	check
Word 7	love	pepper	fuel	score	try	good	lizard	ask	end	fiji
Word 8	new	enjoy	new	india	week	sorry	drink	week	voucher	photo
Word 9	weekend	love	awesome	indian	contest	day	make	feedback	code	odwalla
Word 10	happy	wonderful	vote	match	love	thank	glad	good	gift	free
Word 11	nyc	tip	store	cricket	enjoy	great	water	run	offer	green
Word 12	think	thanx	fan	red	recipe	try	know	twitter	great	facebook
Word 13	enjoy	pomegranate	make	record	know	make	new	time	today	code
Word 14	twitter	great	day	day	glad	come	want	say	tonight	delivery

Table 3.2: Words per topics in brands' daily tweets 2018

After identifying the topics discussed by brands, we will infer whether consumers' communication on a particular day contained that topic or not. LDA also allows us to predict whether a new document is about a previously identified topic. To do that, we will also match consumers and brands in terms of geography. Our data are not restricted to a specific country. Though using English data, the majority of them comes obviously from US and UK. This matching is particularly important since Twitter trends tend to be local.

3.6 Expected results

Our research design offers us three key distinctions:

1. which topic is being discussed
2. who introduced it
3. when the brand started talking about it.

Based on the kind of topic, for example, distinguishing between emotional and informational content, we might observe a different level of the ability of the brand to lead the conversation. Informational content is similar to what we observed in topic 1, based on objective attributes and promotions. On the other hand, emotional content is more engaging on average and hence might encourage consumers to stick to the topic already proposed by the brand.

Based on who introduced the topic, it seems more likely that consumers will lead the conversation because of their number and market power. Hence, it might still be more profitable for the brand to start addressing new topics rather than losing engagement. However, topic shifts and increased engagement are not inherently positive. For example, also firestorms (Herhausen et al., 2019), i.e., social media crises, are characterized by topic shifts and increased engagement. Hence it is crucial to understand to what extent the brand can leave control of the conversation to consumers. We also expect to observe differences in terms of categories of consumers. In particular attentive consumers are more likely to be perceived as an expert source of information, so they might also be more likely than the generic consumers to lead the conversation.

Finally, based on when the brand starts including a topic in its communication, it might gain more or less credibility. For example, now that consumers are, on average, more concerned about the environment, brands that started worrying about their environmental impact earlier are more likely to be perceived as credible on that topic, which in turn would increase the effectiveness of their communication.

References

Babić Rosario, A., Sotgiu, F., De Valck, K., and Bijmolt, T. H. (2016). The effect of electronic word of mouth on sales: A meta-analytic review of platform, product, and metric factors. *Journal of Marketing Research*, 53(3):297–318.

Barberá, P., Casas, A., Nagler, J., Egan, P. J., Bonneau, R., Jost, J. T., and Tucker, J. A. (2019). Who leads? who follows? measuring issue attention and agenda setting by legislators and the mass public using social media data. *American Political Science Review*, 113(4):883–901.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022.

Colicev, A., Kumar, A., and O'Connor, P. (2019). Modeling the relationship between firm and user generated content and the stages of the marketing funnel. *International Journal of Research in Marketing*, 36(1):100–116.

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

De Vries, L., Gensler, S., and Leeflang, P. S. (2017). Effects of traditional advertising and social messages on brand-building metrics and customer acquisition. *Journal of Marketing*, 81(5):1–15.

Herhausen, D., Ludwig, S., Grewal, D., Wulf, J., and Schoegel, M. (2019). Detecting, preventing, and mitigating online firestorms in brand communities. *Journal of Marketing*, 83(3):1–21.

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

John, R. R. and Steven, B. (2002). Marketing communications, theory and applications. *Ed. Pretince Hall*.

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

Kuksov, D., Shachar, R., and Wang, K. (2013). Advertising and consumers' communications. *Marketing Science*, 32(2):294–309.

Kumar, A., Bezawada, R., Rishika, R., Janakiraman, R., and Kannan, P. (2016). From social to sale: The effects of firm-generated content in social media on customer behavior. *Journal of Marketing*, 80(1):7–25.

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

Lambrecht, A., Tucker, C., and Wiertz, C. (2018). Advertising to early trend propagators: Evidence from twitter. *Marketing Science*, 37(2):177–199.

Lasswell, H. D. (2017). *Power and personality*. Routledge.

Lee, D., Hosanagar, K., and Nair, H. S. (2018). Advertising content and consumer engagement on social media: Evidence from facebook. *Management Science*, 64(11):5105–5131.

Moe, W. W. and Trusov, M. (2011). The value of social dynamics in online product ratings forums. *Journal of Marketing Research*, 48(3):444–456.

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

Petersen, J. A., Kumar, V., Polo, Y., and Sese, F. J. (2018). Unlocking the power of marketing: understanding the links between customer mindset metrics, behavior, and profitability. *Journal of the Academy of Marketing Science*, 46(5):813–836.

Rust, R. T., Rand, W., Huang, M.-H., Stephen, A. T., Brooks, G., and Chabuk, T. (2021). Real-time brand reputation tracking using social media. *Journal of Marketing*, 85(4):21–43.

Schramm, W. (1954). How communication works. *1954*, pages 3–26.

Srinivasan, S., Rutz, O. J., and Pauwels, K. (2016). Paths to and off purchase: quantifying the impact of traditional marketing and online consumer activity. *Journal of the Academy of Marketing Science*, 44(4):440–453.

Stephen, A. T. and Galak, J. (2012). The effects of traditional and social earned media on sales: A study of a microlending marketplace. *Journal of marketing research*, 49(5):624–639.

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

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

Toubia, O., Iyengar, G., Bunnell, R., and Lemaire, A. (2019). Extracting features of entertainment products: A guided latent dirichlet allocation approach informed by the psychology of media consumption. *Journal of Marketing Research*, 56(1):18–36.