HARVARD | BUSINESS | SCHOOL

Machine Learning Approaches to Facial and Text Analysis: An Application to CEO Oral Communication

Prithwiraj (Raj) Choudhury
Natalie A. Carlson
Dan Wang
Tarun Khanna

Working Paper 18-064

Machine Learning Approaches to Facial and Text Analysis: An Application to CEO Oral Communication

Prithwiraj (Raj) Choudhury
Harvard Business School

Dan Wang
Columbia University

Natalie A. Carlson
Columbia University

Tarun Khanna
Harvard Business School

Working Paper 18-064

Machine Learning Approaches to Facial and Text Analysis:
An Application to CEO Oral Communication[1]

Prithwiraj (Raj) Choudhury, Dan Wang, Natalie A. Carlson and Tarun Khanna

Last Revised: **July 24, 2018**

The advent of machine learning (ML) tools presents researchers with the possibility of using large and new datasets related to text and image repositories. In this paper, we make a methodological contribution to strategy research by documenting a novel synthesis of *two* machine learning methods – the unsupervised topic modeling of textual data and the supervised ML coding of facial images with a neural network algorithm. We employ these novel methods to study CEO oral communication, using videos and corresponding transcripts of emerging market CEO interviews to conduct our analysis. Building on Helfat and Peteraf (2015) who document the importance of "oral language" as an important managerial cognitive capability, we code the topics and sentiments expressed in the text of what the CEOs say (verbal language) and separately code the facial expressions of the CEOs (non-verbal communication). Using the interview text sentiment scores as well as our video-based facial expression sentiment variables, we conducted factor analysis to construct four distinct CEO oral communication "styles", which we label Expressive, Stern, Dour, and Contented. We also reveal that CEOs who communicate with certain styles also tend to focus on specific topics, even controlling for their country-of-origin and gender. For example, CEOs who tend to be more expressive devote more attention to topics related to society at large and avoid topics related to the government. By contrast, dour CEOs are more likely to dwell on topics related to both the government and society. These results suggest that a CEO's communication style reveals a substantial amount of information about their attention to certain aspects of their businesses.

---

[1] Choudhury and Khanna from Harvard Business School; Wang and Carlson from Columbia Business School. Corresponding author: Raj Choudhury (email – pchoudhury@hbs.edu). The authors would like to thank Connie Helfat, and participants at seminars at Duke University and the China Europe International Business School for comments on a prior draft. Patrick Clapp provided excellent research assistance on this project.

INTRODUCTION

With the advent of empirical techniques based on machine learning (ML), research in social sciences is arguably at an inflection point (Athey, 2018). Recent papers in economics such as Mullainathan and Spiess (2017) and Kleinberg et al. (2017) have demonstrated the usefulness of empirical predictive techniques that build on machine learning concepts. Machine learning techniques have been shown to be particularly helpful in analyzing new sources of "big data" that have been previously underutilized for research, such as large textual archives (Antweiler and Frank, 2004) and repositories of images (Glaeser et al., 2018). In this paper, we detail a novel synthesis of state of the art ML methods for coding textual data and facial expressions to shine light on CEO oral communication. While strategy research has started to embrace machine learning tools for predictive analyses (e.g. Menon, Lee and Tabakovic, 2018), to the best of our knowledge, we are the first researchers in strategy to employ ML tools to *simultaneously code text and image data*, to generate inferences for research.

As a proof of concept, we apply these methods to study CEO oral communication in response to the call made by Helfat and Peteraf (2015) to study verbal language and non-verbal communication, *both* important inputs to managerial cognitive capabilities. Communicating well is one of the most important skills in the CEO toolkit. As Bandiera et al. (2018) argue, CEOs need to create organizational alignment, and this requires significant investment in communication across a broad variety of constituencies. In prior research on CEO communication, the workhorse methodological tool has been content analysis of written communication by the CEO, using data such as CEO letters to stakeholders (Watzlawick et al., 1967; Salancik and Meindl, 1984, Barr 1998, Kaplan, 2008). A glaring omission, however, is the analysis of CEO oral communication, i.e. the analysis of what CEOs say (i.e. verbal oral communication) and their facial gestures (i.e. non-verbal oral communication). Helfat and Peteraf (2015) make a persuasive argument for why strategy scholars should study "oral language" (such as CEO oral communication) vis-à-vis "language production" (such as CEO written communication) given that written communication appears to rely on "more controlled mental processing" than speaking. By contrast, both verbal and non-verbal aspects of CEO oral communication are related to "managerial cognitive abilities." However, thus far research on CEO communication has been constrained by the lack of accessible tools that can systematically quantify features related to verbal and non-verbal communication. Nor have there been efforts to identify possible data sources on visual expressions to address this gap.

In this paper, we employ two ML methods as well as sentiment analysis of text (a non-ML method), to code both what CEOs say (verbal communication) and their facial expressions (non-verbal communication). Our first ML method estimates unsupervised topic models through Latent Dirichlet Allocation (Blei et al., 2003). Topic modeling offers a systematic way of quantitatively measuring the distribution of topics that describe the content of a set of documents in the form of sets of keywords (Kaplan and Vakili 2015). In other words, a "topic" is a probability distribution over terms (words). As an example, for a set of documents, a topic with the highest weights on "dog", "leash", "cat", "toy", and "veterinarian" might be judged to be about pets.

Our second method relies on a (non-ML) textual analysis tool to conduct sentiment analysis: this method is a valuable way to get a sense of the emotional valence of a document. The sentiment measures in this paper are calculated using the *syuzhet* R package (Jockers 2015), which employs crowd-sourced lexicons developed by Saif Mohammad at the National Resource Council of Canada (2013). We use this lexicon to code two sentiments in the content of CEO oral communication: 'positive' sentiment and 'negative' sentiment. Topic modeling and sentiment analysis enable us to analyze the content and valence of CEO oral communication.

The third method employed in the paper uses supervised ML to code expressions of facial images. The underlying algorithm (to be explained in detail later) uses convolutional neural networks (Yu and Zhang, 2015) to code facial emotions. At a very high level, the image recognition process comprises taking an image as an input (e.g. static frame of a CEO face) and transforming the image into a field of weighted pixels to code the facial emotions, where the weights are generated by minimizing a loss/error function. The loss function minimization proceeds by comparing the input image to images from a prior training set which have been coded for their facial emotions.

To illustrate our methodology, we use an archive of video interviews with CEOs and founders conducted as part of Harvard Business School's "Creating Emerging Markets" project, publicly available for academic use in teaching and research The archive consists of a collection of oral history transcripts of interviews with the "CEOs" of 59 unique firms – as well as their corresponding video recordings – conducted by researchers at the Harvard Business School between years 2008 and 2017.[2] Each semi-structured interview ranged between one and two hours, and each interview was transcribed and approved by each CEO prior to public distribution through the archive website. The project intentionally sought out "star CEOs from emerging markets" as interviewees; the typical participant was over the age of sixty years. This helped reduce informant bias, given that older informants could be more frank as their words no longer affected their career prospects (Gao, et al 2017). Examples of the firms that the interviewed CEOs represented included the Tata Group, Claro y Cía and the United Bank for Africa. Thirty-four organizations were interviewed from Asia, 9 from Latin America, 8 from the Middle East and Turkey, and 8 from Africa. We used each of the CEO interview transcripts to generate topic models, to code sentiments from the spoken text of the CEO

[2] The individuals interviewed are typically entrepreneur founders or descendants of founders. They may not be formally designated as "CEO" but are regarded as the leaders of iconic companies.

and to code facial expressions. Each of the transcripts is represented by proportions of forty topics. Each proportion measures the extent to which an interview transcript text touches on a topic. Each interview transcript is also assigned sentiment scores for two that we measure (positive, negative). Finally, we coded the video-graphic material to generate facial expression scores for eight emotions (*Anger*, *Contempt*, *Disgust*, *Fear*, *Happiness*, *Neutral*, *Sadness*, and *Surprise*).

Even within our highly selected sample of "star" emerging market CEOs, in our results, we observe meaningful variation in CEO communication styles. To summarize, through a factor analysis, we used our interview text sentiment scores as well as our video-based facial expression sentiment variables to construct four distinct communication styles, which we label *Expressive*, *Stern*, *Dour*, and *Contented*. We find that these four factors describe 75% of the variance among our video and text sentiment variables. We also reveal that CEOs who communicate with certain styles also tend to focus on specific topics, even controlling for their country-of-origin and gender. For example, CEOs who tend to be more *expressive* devote more attention to topics related to society at large and avoid topics related to the government. By contrast, *dour* CEOs are more likely to dwell on topics related to both the government and society. These results suggest that a CEO's communication style reveals a substantial amount of information about their attention to certain aspects of their businesses. To the extent that CEO attention is an important predictor of CEO preferences, attitudes, and behaviors, understanding how style and attention gives new powerful insight into the foundations of how CEOs both prioritize and make decisions.

Our paper contributes to the literatures on managerial cognitive capabilities and CEO communication. Most importantly, our exposition of two ML methods to code text and image data opens up the possibility of strategy researchers embracing these methods and working with large repositories of textual, image and video data across a variety of settings. Indeed, the findings from our analysis are meant to illustrate the promise of our approach with ample room for future investigations.

CEO COMMUNICATION: PRIOR THEORY AND METHODS

The Chief Executive Officer (CEO) arguably occupies the most central and important leadership role at any firm, being principally charged with the obligation to set firm strategy (Hambrick and Mason, 1984). One of the most important ways that a CEO might influence firm strategy is by communicating her ideas to internal and external stakeholders (D'Aveni and MacMillan, 1990; Lefebvre et al., 1997; Yadav et al., 2007). In fact, Bandiera et al. (2011) measure how CEOs spend their time and show that a disproportionate fraction (85%) of CEO time is spent on activities that might involve communication (e.g. activities such as meetings, public speeches, phone calls and conference calls).

From a theoretical standpoint in the strategy literature, CEO communication has been viewed as a core managerial cognitive capability that underpins the firm-level dynamic capability of reconfiguring. In the dynamic capabilities literature, reconfiguring is instrumental in achieving strategic asset alignment and overcoming resistance to change. In other words, as Helfat et al. (2007) argue, in the face of a change in the external environment, "reconfiguring" involves the acquisition of new assets as well as the enhancement and/or reconfiguring of existing assets through innovation. Helfat and Peteraf (2015) establish a link between CEO communication and reconfiguring, outlining several characteristics of oral communication by CEOs and their effects on individual workers and firm strategy: "The communication style of top managers in general, and the way in which they communicate a vision for the organization in particular, can inspire workers, encourage initiative, and drive entrepreneurial growth (Baum, Locke, and Kirkpatrick, 1998; Westley and Mintzberg, 1989). Managerial skill in using language, such as through impromptu talks, flow of words, and articulation in conversation, may affect worker response to change initiatives" (Helfat and Peteraf, 2015; page 843).

The authors also distinguish between "oral language" (i.e. what the CEOs say) and "non-verbal" communication (i.e. how the CEOs say what they say). In fact, Helfat and Peteraf (2015) argue that non-verbal behavior such as facial expressions and gestures can convey a range of information, including that regarding opinions, values, cognitive states such as comprehension or confusion, physical states such as fatigue, and emotions. As the authors state, CEOs can use oral language and non-verbal communication to facilitate strategic change within organizations and drive alignment by orienting members toward common goals (Hill and Levenhagen, 1995).

The empirical literature in strategy has long studied the effect of CEO communication on firm level outcomes, focusing almost entirely on the content of written communication. Yadav et al. (2007) coded CEO communication using letters to shareholders that were featured in firms' annual reports. Using these data, the authors showed that certain features of CEO communication – specifically having greater internal and external focus – can have a "positive and long-term impact on how firms detect, develop and deploy new technologies over time" (Yadav, et al 2007: 84). Similarly, D'Aveni and MacMillan (1990) compared senior managers' letters to shareholders during demand-decline crises for 57 bankrupt firms and 57 matched survivors. The authors found that under environmental uncertainty, not only do the CEOs of surviving firms pay disproportionate attention to the output environment of the firm, but their communication to shareholders also more strongly reflect these structural differences in their attention. CEO communication has also been studied in the strategy literature on cognitive frames: Kaplan (2008) uses CEO letters to shareholders and content analysis to measure managerial cognition.

In fact, in our review of strategy studies on CEO communication, the workhorse methodological tool has been content analysis of written communication by the CEO, such as analysis of CEO letters to stakeholders (Watzlawick et al., 1967; Salancik and Meindl, 1984). Although written communication has several attributes that lend itself well to content analysis, a glaring omission is the analysis using records of oral communication by the CEO. However, from a theoretical standpoint, the analysis of the verbal and non-verbal elements of CEO oral communication is arguably important to understand managerial cognitive capabilities. In fact, Helfat and Peteraf (2015) persuasively argue that it is important to study "oral language" (such as CEO oral communication) vis-à-vis studying "language production" (such as CEO written communication). The authors build on Bialystok and Ryan (1985) and state that "writing, for example, appears to rely more on controlled mental processing than speaking" (Helfat and Peteraf, 2015; page 842).

Arguably, an important reason why researchers have neglected CEO oral communication has been the absence of methodologies hitherto that can perform "unsupervised" analysis using large datasets of CEO oral communication. In fact, Kaplan (2008) acknowledges this constraint and justifies the use of CEO written communication in her analyses by saying, "other kinds of statements by CEOs, such as those obtained through interviews or surveys, might initially appear to be attractive (data) sources, but they are impractical for larger samples of firms over long periods" (Kaplan 2008, page 679).

Recent advances in supervised and unsupervised ML techniques now present strategy scholars with an opportunity to shed light on CEO oral communication. In fact, given that these ML algorithms could be applied to *both text and facial image data*, we now have an opportunity to begin to deliver on what Helfat and Peteraf (2015) have argued: the systematic study of both verbal and non-verbal CEO oral communication. We now outline our methods, dataset and results.

METHODS AND RESULTS

Overview

We develop an approach that synthesizes methods for coding unstructured text data from oral communication and video data from the facial expressions of the corresponding speakers to create a categorization of CEO communication style. We do so in the spirit of Helfat and Peteraf (2015: 837), who identify verbal and non-verbal "oral language" as one of the chief inputs to a manager's cognitive capabilities, which, they argue, can have a profound influence on managerial strategic decision-making. Communication styles have been analyzed in a variety of settings, from physician-to-patient interaction (Buller and Buller 1987), sales pitches (1985), and political speeches (Perloff 2013). Although a variety of definitions exist, Norton's (1978: 99) explanation is arguably the most generalizable across contexts: "the way one verbally and paraverbally interacts to signal how literal meaning should be taken, interpreted, filtered, or understood."

By illustrating a generalizable approach for coding CEO communication style, we emphasize two major implications of our methodological innovation for strategy researchers more broadly. First, because our method brings together verbal language content data and non-verbal facial expression data, we demonstrate how researchers can better understand the relationship between these two dimensions of communication from organizational leaders. Namely, strategy researchers have increasingly begun to draw on machine learning methods to analyze and categorize large corpora of text data (e.g. Menon, Lee and Tabakovic, 2018). However, in the context of CEO communication, text data are codified and written, representing the *edited* thoughts and views of their authors. Intent, attitudes, and views are often conveyed in non-verbal expression, which in some cases, can bring nuance to our understanding of content that is spoken, and in other cases, contribute to surfacing a speaker's 'authentic' perspective on a given matter. Therefore, our approach allows researchers not only to make sense of the topics that receive most of a CEO's attention but also to gain insight into *unedited* attitudes and feelings about those topics directly from the speaker.

Second, we describe our methods in a way that we hope can be generalized across different sources of video data. Specifically, we bring attention to the widespread availability of video data. Although the video data for our sample of CEOs comes from a curated online archive of video interviews, we remind researchers that large online platforms, such as YouTube, and more focused news outlets, such as TechCrunch or CNN, contain searchable video archives of CEO speeches, interviews, and other forms of communication. Given that we increasingly consume information through online video, researchers of leadership should be sensitive to online video platforms as an as yet, untapped data source of managerial communication. Unsupervised topic modeling, combined with automated transcription software and facial expression coding algorithms, constitute a robust set of freely-available tools for the analysis of such widely-available data. We describe how we take advantage of each of these tools in a way that can be generalized to other sources of video and communication data.

Figure 1 reports a general roadmap summary of our methodological approach.

INSERT FIGURE 1 ABOUT HERE

The three tracks in the flowchart represent the three measures we introduce to analyze the text and video: the LDA topic model (A in Figure 1), which models the content of the text; the text sentiment measures (B in Figure 1), which represents the positive and negative sentiment reflected in the text; and the facial emotions gleaned from the video (C in Figure 1). Once these measures have been calculated, we develop four clusters of CEO communication styles based on text sentiment and facial emotion (D in Figure 1) and then examine how these styles relate to the content of the interviews as reflected by topics (E in Figure 1). We describe each of these steps in detail, along with examples from our data, in the following section.

Topic Modeling

To code the content of CEO communication, we first obtain transcripts of the interviews with each of the 59 CEOs in our data. Each interview ranges from 1.5 to 2 hours long, producing 59 transcripts, which on average were 7991 words in length with a standard deviation of 2887 words. We use an unsupervised topic modeling approach to estimate a set of topics into which we can categorize each CEO's interview transcript. By *unsupervised*, we mean that the only input (i.e., text) provided by the researchers in the topic model estimation is the overall number of topics. This approach allows the text to "speak for itself" in that the topics that emerge from the model are not influenced by what semantic subject matter the researchers might expect to find.

There exist other forms of *supervised* topic models, in which, for example, the researchers might specify a document-level covariate of interest because they are specifically interested in analyzing how topics differ along that dimension. In our approach, however, we are most interested in an unbiased representation of the semantic content of the text, which we argue is the approach most researchers should take unless researchers have a specific reason for speculating that a set of texts would be described a known set of topics. Because our text input consists of transcripts of unstructured interviews, we have no clear reason for defining an *ex ante* set of topics in our sample.

In the remainder of our description, we refer to each transcript as a "document" to be consistent with how the field of computational linguistics refers to each separate body of text inputted into a topic model. Below, we first describe the Latent Dirichlet Allocation (LDA) algorithm used to generate topics from a set of documents. From there, we describe the specific pre-processing and cleaning steps we took to prepare our data for topic model estimation. We summarize each step in a way that can be applied to most forms of interview transcript data that might be used as input documents. Finally, we offer an interpretation of the topic model output for our data, as well as several examples to bring greater meaning to the model's results.

LDA Algorithm. The LDA model treats each document as a bag of words, meaning that the word order is not taken into account, and assumes an underlying random generative process in the creation of the "corpus" – or the set of documents being analyzed. It assumes that the collection of documents was generated by an imaginary probabilistic process, word by word, by first sampling a topic from a given document's distribution of topics and then sampling a word from that topic's word distribution. The sampling algorithm takes in the cleaned documents and then works backward, returning the most probable set of topics to have produced the given set of documents, if they had indeed been created in this imaginary way. A researcher can then infer the meaningful subjects represented by these topics, as in the transportation example given above, and calculate the proportions of each document estimated to belong to each topic.

Mathematically, the model assumes each document consists of a random mixture of a finite set of topics, and each topic represents a probabilistic distribution over the terms in the "vocabulary". A "vocabulary" refers to all of the terms used at least once across the entire collection of documents. The model is essentially a Bayesian variant of Latent Semantic Analysis, in which the topic distribution is given a Dirichlet prior (Griffiths and Steyvers 2004). Specifically, the Dirichlet distribution is a probability distribution that samples over a discrete set of categorical events, and is often used as a prior in Bayesian mixture models.

INSERT FIGURE 2 ABOUT HERE

The resulting probabilistic generative process – the hypothetical way in which the documents are assumed to have been created – is graphically represented by the plate diagram in Figure 2. The larger plate indicates that the step is repeated for each of *M* documents, while the smaller plate indicates that within a document, the step is repeated for each of the *N* words. To generate each word in each document, the process consists of selecting a topic *z* over the document's mixture of topics, and then a word *w* from that topic's vocabulary weightings.[3] To determine the most likely set of topics to have generated the collection of documents, we fit the model on the corpus by employing a Gibbs sampling algorithm – a commonly-used method of iteratively sampling until convergence is reached – as the optimal solution cannot be solved for directly.[4] We run the sampling using the *topicmodels* package in R (Hornik et al. 2011).

[3] α and β are the hyperparameters for the Dirichlet priors on the topic distribution per document and the term distribution per topic, respectively. θ_m parameterizes the categorical distribution of the document's topic mixture, while the topic's vocabulary weightings have a categorical distribution with parameter Φ_z.

[4] Because the underlying estimation problem is intractable, a number of approximation methods are typically used in estimating the LDA model, most commonly expectation-maximization algorithms and Markov chain Monte Carlo (MCMC) sampling methods (Yao, Mimno et al. 2009). In this analysis, we employ one of the MCMC methods, collapsed Gibbs sampling. This is a permutation of the standard Gibbs sampling algorithm, a process of iteratively sampling the conditional probabilities of a joint distribution. By collapsing out (i.e. integrating over) the Dirichlet prior distribution, the algorithm encourages faster convergence (van Dyk and Park 2008).

Cleaning and pre-processing. A number of preprocessing steps are necessary to ensure that a LDA model results in coherent topics. In particular, for our oral history transcripts, we only used text that was spoken by the interviewee so that we do not simultaneously model the thoughts and opinions of the interviewer. Also, because several of the CEOs were interviewed in a language other than English (specifically, either Spanish, Portuguese, or Turkish), we utilized the English translations of the interview transcripts as our input data. We acknowledge that this might stand as a limitation of our approach, as our model might be accounting for a translator's own interpretations of a CEO's words rather than capturing the CEO's native tongue expressions. Our regression analysis ultimately attempts to account for this potentially confounding factor by controlling for CEO's national origin, but we recognize that this limitation would not exist if all of the interviews were conducted in the same language.

Segmenting. Document length can have a powerful influence on the interpretability of an LDA model, which is highly relevant to the context of oral histories and interviews. For example, a very long document may contain so many subjects that it is difficult for the algorithm to identify a coherent set of topics, since the document is treated as a single bag of words. A frequent step with longer documents is to break down the document into smaller, semantically coherent segments (commonly 500 or 1000 words), a process for which a number of algorithms exist (Riedl and Biemann 2012). However, the turn-taking design of an oral interview provides a natural structure by which to segment each document. By removing the interviewer's questions and treating each response as its own segment, model performance improves significantly. The model then treats each segment of a transcript (i.e., each response to an interviewer's question) as its own stand-alone document.

Lowercase, punctuation, stopwords. A typically mandatory step in the document cleaning process is the conversion of all text to lowercase and the removal of punctuation and numeric characters. Because word order and sentence structure are not considered in the LDA estimation process, it should not matter whether a word is capitalized (e.g. at the beginning of a sentence) or not; likewise, punctuation symbols are rarely relevant. Another common step is to remove all "stop words" – that is, common words such as "and" or "the" that give no relevant information about the topic probability. Most text packages, such as Python's *NLTK* or *tm* in R, have prebuilt list of stopwords in English as well as other languages. In some cases, it may be useful to add additional stopwords to the standard lists that are relevant to the specific context. For example, we drop instances of the word "business" and "company" because they are so commonly used in the transcripts as to be rendered uninformative for our purposes. Another way to achieve this purpose is to drop words that appear in too few or too many documents – for example dropping all terms that appear in less than five percent or more than 95 percent of documents. This will serve to eliminate terms that are highly specific to one document (e.g. proper nouns like the name of a company) or that are common enough to have little specific meaning (e.g. "business" in the context of this sample).

Stemming. Finally, *stemming* of words to their root form – an algorithmically-assisted process by which "run", "runner", and "running" would all be reduced to the stem "run" (Lovins 1968) – is often helpful in achieving coherent topics. This step is common but not required for achieving a good model. Packages for stemming terms often use simple rules that fail to reduce all terms to the proper root (for example, "better" would not be reduced to its root form "good" because they do not have the same stem). A related but more difficult task, known as *lemmatization*, uses more complex algorithms that take the full context of a phrase into account to reduce terms to their base forms rather than just simple rules for identifying root words. However, lemmatization algorithms typically run far slower than stemming algorithms, and therefore stemming – which we employ – is more commonly used for purposes such as topic modeling.

Output and Interpretation. Choosing the optimal number of topics for a topic model to produce over a set of documents is often characterized as more of an art than a science. Measures of a model's fit to the corpus, such as perplexity and log likelihood, can provide some guidance. It is worth noting that these measures do not always line up exactly with human judgments of semantic coherence, and human judgment remains the most popular way of selecting a final model (Chang, Boyd-Graber et al. 2009). Coherence is typically best determined by examining the most likely terms for each topic: a good model should allow an observer to intuitively assign a title to each of the topics with a quick glance at the most probable terms.

After estimating models across ranging from 10 to 100 topics, we determined the most coherent set of topics to emerge from a model with 40 topics. We found that within this range, the log-likelihood values of the models were maximized in the range of 40 to 50 topics. We then examined the models associated with this smaller range of topics and found that in models with greater than 45 topics, some of the topics appeared "repetitive" in semantic meaning, in that in some cases there were multiple topics that seemed to apply to the same subject matter without a clear distinction. Finally, we settled on the model with 40 topics as having the set of topics that had the most obvious and intuitive labels. The most probable words for each topic in this model are displayed in Figure 3.

INSERT FIGURE 3 ABOUT HERE

In interpreting the model, the researchers assigned names to each topic based on its most probable terms. We noted that some of the topics appear to be industry-specific, while others are more general. Topic 8, for example, clearly seems to refer to marketing and branding ("brand", "market", "product", "brands", "consumer") while Topic 33 seemed related to retail ("customers", "quality", "price", "product", "buy"). The more general topics appear to span both work-related subjects (Topic 32, which we named Corporate Boards, for example: "board", "members", "shares", "chairman", "holding"), as well as personal subjects (Topic 28, Family History: "family", "person", "generation", "house", "grandfather"). This breadth reflects the variety of subjects encountered in the freewheeling interview format, and the manner in which it provides a unique view into the thoughts of each CEO.

To generate document-level covariates from the topic model, we calculate the proportion of words belonging to each topic in each segment. As our corpus structure consists of long documents split into segments, we collapse each topic proportion back to the original document – i.e., interview transcript – by weighting by the length of each segment. The resulting covariates each have a value between zero and one, and the proportions of the forty topics will sum to one for each interview.

For example, Topic 2: Hospitality, has a mean value of 0.019, meaning that the average interview has less than two percent of its content estimated to belong to this topic. The interview transcript of P.R.S. Oberoi, however, has a value of 0.253 for Topic 2, indicating that approximately a quarter of the interview content concerns hospitality – an intuitive result, as Oberoi is the chairman of Oberoi Hotels and Resorts. Other topics are more universal and evenly distributed, such as Topic 14: Childhood, but even these show substantial variance between interviews. Mallika Sarabhai, the director of the Darpana Academy of Performing Arts, for example, spends a good portion of the interview discussing her childhood and family upbringing, which is reflected in a value of 0.12 for Topic 14. Other CEOs, such as Andre Esteves of BTG Pactual, spend little to no time at all on personal topics – in Esteves's case, this is reflected in a Topic 14 value of 0.004.

We highlight three groups of topics that are interesting to consider in our analysis, because they are potentially relevant to all of the CEOs in the interview sample: family, government, and society. There are several topics that pertain to each of these issues. Three topics – Topic 7: Family Difficulties, Topic 14: Childhood, and Topic 28: Family History – each relate to different aspects of the CEOs' personal histories and family lives. We sum these three to create a measure of *Family Topics*. Similarly, there are four *Government Topics* that pertain to the public sector and the state: Topic 4: Politics, Topic 10: Law, Topic 25: Taxes, and Topic 38: Public Sector. Finally, four *Society Topics* relate to the broader community and society at large: Topic 16: Development, Topic 21: Community, Topic 24: Society, and Topic 39: Women's Empowerment. We return to these topics in our analysis of how communication styles are related to the content of a CEO's speech.

Text Sentiment Analysis

Separate from the topic model, sentiment analysis is a valuable way to get a sense of the emotional valence of a document. These methods are usually dictionary-based. The sentiment measures in this paper are calculated using the *syuzhet* R package (Jockers 2015), which employs crowd-sourced lexicons developed by Saif Mohammad at the National Resource Council of Canada (Mohammad & Turney 2013). In future work, it may be useful to develop custom dictionaries specific to the purposes of strategy researchers. The NRC lexicons used here correspond to two sentiment categories, positive and negative. For each sentiment, the terms in the lexicon have a binary value for association; for example, the word "abandon" is assigned negative sentiment value, while the word "ability" is given positive sentiment. This approach is somewhat crude, as it does not consider the context or word order of a phrase, but on balance it typically performs nearly as well as more complex approaches (Mohammad et al 2013). We sum the terms associated with each of the sentiments at the sentence level, and then calculate the proportion of each document dedicated to each sentiment, so that the values sum to one.

As an example of how this process works, consider the following passage from Anu Aga's (then Chairperson of Thermax Social Initiative Foundation in India) transcript:

> *I don't think joining HR was difficult but what was difficult was getting back to work after a gap of many years. I wondered how I could be away the whole day and come home late, leaving the children without me. I kept thinking what if my children or my mother-in-law got sick and needed me. I was a bit anxious about how the other professionals in HR who had studied HR would accept me. But I must say, we had a wonderful team.*

This segment would be scored with five negative words ("difficult", "gap", "late", "sick", "anxious") and two positive words ("mother", "wonderful"). If this short section were the entire interview, these sentiment values would then be converted to proportions, with a value of 0.71 for *negativ*e and 0.29 for *positive*. The higher negative value reflects that the segment dwells mainly on negative sentiment (concern and anxiety about returning to work) punctuated with some positive sentiment (warm thoughts about the team). Across the entire interview, the sentiment values provide a picture of the extent to which the CEO prefers to reflect on negative emotions, or adopts a more positive tone – a key component of style.

Facial Image Recognition Analysis

The third analytical tool employed in this paper uses supervised ML technology that takes a static facial image as input and generates as output, weights along eight facial expressions, i.e. *Anger*, *Contempt*, *Disgust*, *Fear*, *Happiness*, *Neutral*, *Sadness*, and *Surprise*. This tool – the Microsoft Azure

Computer Vision REST API – was developed by Microsoft and builds on research by Yu and Zhang (2015).[5] We first describe the algorithm underlying this tool and then explain in detail, the use of the technology.

The Convolutional Neural Network Algorithm. The Microsoft API utilizes a version of a class of algorithms known as convolutional neural networks (Yu and Zhang 2015). Although the technical details of the algorithm are vital for researchers of artificial intelligence, for strategy researchers, we instead summarize the conceptual ideas.

A supervised neural network algorithm is implemented in three steps. In the first step, researchers employ a "training set" (frame-by-frame snapshots of a video in our case) that are labeled according to their facial emotions. In the second step, the actual input image is transformed using a neural network, into a field of "weighted pixels". These pixel weights are used to generate values for parameters such as 'how open is the mouth', 'curvature of lips', 'dimples on the cheek', etc. These parameters are then used to generate "output values" for facial expressions of the input image. In the third step, the weights (on the pixels) are optimized based on minimizing a loss function/error function, where the error is coded based on the difference of the "output values" of facial expressions coded in the prior step, and the "target value" of the same facial expressions. The target values of facial expressions are generated based on the same parameters used to generate the output values; however, unlike the "output values", the "target values" are based on data from the training set.

To explain how the neural network algorithm works, we build on the rich literature in the field of computer science of employing neural net (NN) methods. As Duffner (2008) says, NN algorithms are inspired by the human brain and its capacity to perform complex tasks by means of inter-connected neurons, each performing a relatively simple operation. Similar to the human brain, a NN is a trainable structure consisting of a set of inter-connected "neurons", each implementing a very simple function. Collectively, the NN performs a complex classification function or approximation task.

In the case of facial image recognition algorithms, each "neuron" corresponds to a pixel in the image data. The task of the algorithm is to "read" the input image file and generate a set of "weights" to be assigned to pixels to code the parameters of interest such as, 'skin color' or 'how open is the mouth'. As an example, skin color might indicate the existence of eyes or hair (versus rest of the face). An open mouth with dimple on the cheeks might indicate the facial expression of 'Happiness'.

[5] Available at https://azure.microsoft.com/en-us/services/cognitive-services/emotion/

In a simple, brute-force approach, the NN algorithm could consider each pixel of the input image data, and assign weights to every pixel, to compute the parameters of interest and minimize the loss function (as described earlier). However, this would be a case of "over-specification" and would be computationally intractable for most image datasets. Instead, the NN algorithm conducts localized optimization, where pixels in a "neighborhood" are assigned weights to successively generate higher level weights. A convolutional NN algorithm builds on this principle, by converting an input image into a multi-layer hierarchical structure, where the first layer relates to the input image, the next few layers relate to "shallow" collections of pixels, where pixels are grouped based on their neighborhood (e.g. neighborhoods comprising the edges of the image, part of the nose, part of the eye, etc.) and the subsequent layers relate to "deep" explorations of distant neighborhoods, covering the entire face. To summarize, the weights are iteratively chosen to minimize the loss function described earlier. Once the final weights are chosen, the algorithm generates scores for the facial expression emotions.

Capturing Static Frames from Video File. We use the Microsoft API which outputs facial expression scores for a set of images with the NN algorithm described above. Before using the Microsoft tool, researchers must prepare the facial image data. If the facial image data is available as part of a video file (as in our case), this entails capturing individual static image frames from the video file. This task can be achieved by using media player applications. We use a cross-platform, open-source media player called the 'VLC media player', which allows for the capture and export of static image frames from video data by using its "scene video filter" option. Settings within the VLC filter preferences can be used to adjust the number of frames extracted, their associated filenames, and filetype. We captured one static image frame per second of video footage and most importantly, only used the static image frames that related to the face of the CEO. In other words, we dropped from the sample, all static image frames related to the face of the interviewer and static image frame without any facial images (e.g. title frame).

It is important to note, however, that the algorithm implemented by the Microsoft API is also able to 'recognize' a face amid other objects in a static image. Although seemingly an obvious innovation, this ability represents a major breakthrough in artificial intelligence image recognition technology, which makes the use of facial expression tools much more accessible across scientific fields. In effect, the ability to recognize faces removes the major barrier of having to manually crop images so that the faces contained therein are made apparent.

Using the Microsoft Azure Computer Vision REST tool. Once the static images are ready for use, researchers can employ the Microsoft tool to generate facial image recognition scores. To do so, they must first apply for an Application Program Interface (API) key from Microsoft Cognitive Services for permission to use the Face API. A free trial and set of API keys for the Face API is available to researchers through the Microsoft Cognitive Services website.[6] Signing up for the Face API grants a single user a key which permits processing up to 30,000 static images, at a rate of twenty images per minute. The API returns emotion scores for the five facial emotions, where each emotion receives a score between zero and one, according to the algorithm developed by Microsoft. That data is reported back in a JSON file. We used SAS to collate the facial emotion scores and the frame number from the collection of JSON files.

Facial Expression Data Output. The data returned from the Microsoft API assigns scores between 0 and 1 for each of eight different emotions -- *Anger, Contempt, Disgust, Fear, Happiness, Neutral, Sadness,* and *Surprise* – for each image. The sum of the eight scores (for the eight emotions) for a given image is equal to 2. Therefore, a score for a given emotion can be interpreted as an indicator of the intensity of the emotion expressed relative to the other emotions that could be expressed. Because a set of images for a given interviewee represents one-second snapshots of the interviewee's video, taking the average score of *fear*, for example, for the entire video gives a summary of the extent to which the individual on camera expressed fear. Figure 4 displays examples of static frames with high scores for each of the eight emotions recognized by the algorithm.

INSERT FIGURE 4 ABOUT HERE

Discovering "Styles" through Factor Analysis

In the next step, we use the facial emotion and text sentiment scores to discover clusters of communication styles among the CEOs in our sample (Section D of Figure 1). The reasoning behind this step is that the facial cues and sentiment terms employed by the speaker are likely to reveal some information about the speaker's preferred mode of communicating. Much in the same way we might expect that bystanders could watch a CEO conduct a meeting and come to general agreement about whether that CEO is serious and buttoned-up, loose and informal, or expressive and excitable, we are assuming in this analysis that the sentiments expressed in the speaker's words and facial expressions will help to approximate these "styles" in a way that the content of the interview cannot. We use factor analysis to discover the styles in our sample although any form of dimensionality reduction could serve a similar purpose.

[6] https://azure.microsoft.com/en-us/services/cognitive-services/face/

Factor Analysis Results. In our factor analysis, we include ten variables: the two text sentiment measures (*Positive* and *Negative*) and the eight facial emotion measures (*Anger*, *Contempt*, *Disgust*, *Fear*, *Happiness*, *Neutral*, *Sadness*, and *Surprise*). Figure 5 displays a scree plot of the eigenvalues for the top ten factors in the resulting analysis. We select the first four factors as the most prominent styles present in the sample, as these four cumulatively represent 75 percent of variance explained, and the value of the eigenvalues diminish substantially from the fifth factor.

INSERT FIGURE 5 ABOUT HERE

We termed these four styles *Expressive*, *Stern*, *Dour* and *Contented* after examining the factor loadings, which are displayed in Table 1. The first factor, *Expressive,* is characterized by strong positive language and dramatic facial expressions, particularly surprise and fear. Those strong in this factor display significantly fewer neutral facial expressions, which is notable as for most of the interviews neutral is the most dominant facial emotion. The second factor, which we named *Stern*, is characterized by more angry and disgusted facial expressions and less facial happiness; however, this factor is also associated with more neutral faces and a modest association with positive language, leading us to interpret it as a stern, no-nonsense style.

The third factor, on the other hand, we term *Dour* as its strongest associations are with negative language and angry facial expressions. Finally, we name the fourth factor *Contented*, as it is most characterized by facial happiness, with a lack of surprise, fear, or emotional language. Those strongest in this style are likely those CEOs whose resting face is a smile, rather than a neutral expression, as this factor loads negatively on neutral facial expressions.

INSERT TABLE 1 ABOUT HERE

<u>*Relationships between Styles and Topics*</u>

Correlations. As previously mentioned, we highlighted three topic groups of interest to our analysis: Family Topics, Government Topics and Society Topics. We chose these broad subjects because, unlike industry- or region-specific matters, they are likely to be relevant to all of the CEOs in our sample. The extent to which the CEOs choose to dwell on these broad subjects in an unstructured interview is therefore likely to be influenced by their communication style. A CEO with an intimate, personal style of communicating – such as our *Contented* style, for example – might be more likely to share stories about their family life, while a CEO who communicates with a forward-looking and inspirational style – such as our *Expressive* category – might spend more time talking about the future of society.

INSERT FIGURE 6 ABOUT HERE

Figure 6 visualizes a heatmap of correlations between the four communication styles identified in our factor analysis and the topic groupings. We observe that CEOs strong in the *Expressive* style are much more likely to talk about society topics (correlation coefficient = 0.33) and much less likely to dwell on government topics (-0.41). CEOs who are more *Dour,* however, spend more time on both society and government (coefficients of 0.21 and 0.23, respectively). CEOs who rate higher in the *Contented* style are much more likely to devote the interview to speaking about their family (0.25).

Regression Analysis. Table 2 tests the strength of these relationships in a regression analysis. As communication styles are associated with cultural and gender differences – which could also influence the content covered in an interview – we control for gender and region of origin in our models. The full model (results shown in Columns 3, 6, and 9) estimates an OLS regression with the following specification:

$$Topic\_Group_i = \beta_0 + \beta_1 Expressive_i + \beta_2 Stern_i + \beta_3 Dour_i + \beta_4 Contented_i + \beta \boldsymbol{X}_i + \varepsilon_i$$

The covariate vector $\boldsymbol{X}_i$ includes gender and region indicators for Asia, Africa, and Latin America (the omitted region being the Middle East).

INSERT TABLE 2 ABOUT HERE

As the dependent variables – the topic groups – represent a proportion between zero and one, we can interpret the coefficients as follows: a one standard deviation increase in the *Contented* style score is associated with approximately one percent more of the interview devoted to family subjects (Column 1, p=0.008). This association is small but meaningful; despite the fact that female and Asian CEOs are also more likely to spend time on family topics, the relationship holds when these factors are accounted for (Columns 2-3). Likewise, when region and gender are included in the model, a standard deviation increase in the *Dour* style is associated with a one percent increase in government topics (Column 6, p=0.012), while an equivalent increase in the *Expressive* score is associated with four percent less time on government topics (p=0.001). These relationships, too, hold when accounting for the fact that African CEOs are more likely to discuss government topics.

Finally, an increase of one standard deviation in the *Dour* and *Expressive* scores are each associated with an approximately one percent increase in society topics (Column 9, p=0.019 and p=0.049, respectively). These style associations hold when controlling for the fact that both women CEOs, and CEOs from Asia, Africa and Latin America (relative to those from the Middle East, which is the omitted region in the model) are more likely to dwell on topics related to society.

DISCUSSION

In this paper, we outline a novel synthesis of *three methodologies – topic modeling (using unsupervised ML), sentiment analysis of text, and facial image expression recognition (using supervised ML)* – with an application to CEO oral communication. Exposition of these methodologies allows us to respond to the call made by Helfat and Peteraf (2015) to study verbal language and non-verbal communication, important inputs to managerial cognitive capabilities. Specifically, we collected and processed verbal and non-verbal data from a set of video-recorded interviews with CEOs and founders from emerging markets to show that both components are informative when identifying different *communication styles*.

As a demonstration of a potential application of the novel methods documented in this paper, our analysis indicates that just knowing the CEO's communication style allows us to predict what topics are important to the CEO. The topics that are important to a CEO are indications of the CEO's attention. A wealth of studies have focused on how CEO attention is a predictor of behavior (e.g. D'Aveni and MacMillan, 1990; Yadav et al., 2007; Bandiera et al, 2018) but it is difficult to predict what will attract a CEO's attention. Our analysis offers a glimpse of a potentially important predictor, i.e. CEO communication style. Because our video data come from recorded sessions of unstructured interviews, in which CEOs engaging in free association with minimal prompting from an interviewer, our setting provides a unique perspective on how CEOs view what is important to them. Recent work has taken an interest in measuring how CEOs allocate their attention as a key input into understanding how they make decisions about firm strategy. In particular, researchers gather data by collecting information on how CEOs spend their days through detailed diaries (Bandiera et al., 2018). Our approach sheds light on two promising ways to further the literature on CEO attention and decision-making by (1) introducing new type of data and measurement approach – topic modeling of transcripts of unstructured interviews and analysis of facial expressions based on image data – for gathering information on how CEOs prioritize certain matters and topics, and (2) revealing how a new factor – communication style – might be related to CEO attention.

In the context of research on CEO communication, the set of methodologies developed in this paper could be used in the literature of cognitive frames (Kaplan, 2008), interpretation (Barr 1998) and how CEOs spend time (Bandiera et al, 2011, 2018). More broadly, these new methods to code verbal and non-verbal oral communication could be particularly instrumental in research that uses analyses of language. As Suddaby and Greenwood (2005) outline, drawing from Buke's (1969) notion of language as "symbolic action", several streams of research related to strategy employ the analysis of language. Important sub-fields of related research include semiotics (Barley, 1983), hermeneutics (Phillips and Brown, 1993), discursive analysis (Kilduff, 1993), narrative analyses (Boje, 1995) and rhetorical analysis (Freedman and Medway, 1994). Scholars in each of these sub-fields could benefit from using the methodologies outlined in this paper.

More broadly for strategy research in the age of Twitter, Instagram and YouTube, these tools could be used by strategy scholars to code text, static image and video data in a wide variety of settings. Arguably a new set of methodologies to work with qualitative data such as text, static image and video images provides an empirical breakthrough. In fact, as a recent SMJ editorial persuasively argues, studies using qualitative empirical methods have been instrumental in advancing the field of strategic management (Bettis et al., 2015). The article outlines several qualitative methods that have been used in strategy research including QCA (Ragin, 2014), first and second order analysis (Gioia, 2014), the case study method (Eisenhardt, 2014) and rhetorical analysis (Suddaby, 2014).

The exposition of our novel set of methodologies to utilize oral history data adds to the relatively thin literature on the use of historical data in strategy research. In particular, Jones and Khanna (2006) outline two dimensions of historical data that makes it difficult for use in broad strategy research – such data is often "qualitative" and often "small sample". The authors then suggest methods that strategy scholars could use to analyze historical data and list methods related to Boolean algebra (Ragin, 1987), string analyses (Abbott, 2001) and computational models (O'Rourke and Williamson, 1999). Oral history data – especially those accompanied by images or video – arguably an underutilized data source for strategy research, often shares the qualitative and small sample properties outlined by Jones and Khanna (2006) and our novel set of methodologies provides strategy scholars yet another empirical tool to use to further historical analysis in strategy research. In effect, we show how even with a small sample of interviews (n = 59), our approach through segmenting each interview transcript allows for a meaningful and replicable quantitative analysis through topic modeling, sentiment analysis and facial image recognition analysis.[7]

[7] In the appendix, we list selected oral history archives, mostly housed in university libraries that contain a diverse array of interviews with business leaders, covering a wide range of industries, regions, and topics. One notable resource is Columbia University's Oral History Archive, which has been widely acknowledged as the largest searchable database of oral history records in the world, giving access not only to audio and video records of interviews with business executives, but also their accompanying transcripts as well.

Our study has several limitations. First, because our data are limited to interviews with CEOs of firms in emerging markets, we cannot generalize our results about CEOs' emotion and topical attention to those in other settings, such as developed markets. We encourage researchers to adopt our methods to future projects that might examine such a comparison. In addition, in terms of data limitations, as Kaplan (2008) states, the study of oral interview data suffers from the risk of retrospective bias as managers would likely adapt their memories of their views in prior years to subsequent outcomes. As for other technical limitations, we also can only account for differences in the region-of-origin for our CEO interviewees and the firms they represent. However, as a feature of the interview data collection, the CEOs' regions are also associated with whether the interviews themselves were conducted in English. For instance, most CEOs from South American countries were interviewed in their native Spanish, which meant that our analysis could only incorporate the English translations of their interview transcripts.

While the current study represents our efforts to advance our understanding of how ML methods could be gainfully used in strategy research, future efforts could augment our current study in several ways. Firstly, while we focus coding on text and facial expressions of oral communication, it is possible to additionally use the intonation of the voice to code yet one more dimension of oral communication. Secondly, future research might augment our ML based topic modeling analysis in several ways: it would be interesting to investigate the sensitivity of topic model results to translation effects. Additionally, although our approach utilized unsupervised LDA to estimate topic models, it is possible that a supervised approach could produce additional insights on topic estimates (Ramage, et al 2009). A supervised approach would require researchers to read through a sample of transcripts and to associate certain words with pre-determined topics, giving the topic model a fixed prior for structuring the relationship between estimated topics. A supervised approach is encouraged when the language used in a corpus of documents has excessive jargon, such that relevant experts would be able to identify which specific and salient words should cohere together as a topic. The language in our interviews do not arguably reflect the excessive use of jargon, but it is possible that other oral business histories exhibit higher proportions of industry-specific terminology. Additionally, while we use 'bag of words' methods to construct the topics, it might also be interesting to study how the order of words correlate with sentiments expressed. Thirdly, future research might augment our ML based facial expression analysis by subjecting facial images to an unsupervised ML facial recognition algorithm. Though the computer science literature has favored supervised approaches to code facial expression (given that the number of unique human facial expressions are much lower than the number and variation in topics within text), it might be interesting to attempt an unsupervised coding of facial expressions, for robustness sake. Lastly, future research might augment our textual sentiment analysis by creating and using a lexicon of words, curated from papers published in the field of strategy, to code sentiments expressed in the words spoken or written by CEOs.

In conclusion, from the perspective of strategy research, we document three replicable methodologies, combining the analyses of text and facial image data using machine learning methods. We exploit an underutilized type of data for strategy research, i.e. oral history text and video data, and describe in detail three replicable methods based on topic modeling of text, sentiment analysis of text and facial image emotion recognition. We also develop a proof of concept of using our methodology and provide evidence suggestive of how communication style correlates with the content of CEO communication. This result speaks to the importance of studying *both* verbal and non-verbal language, highlighted by Helfat and Peteraf (2015). Most importantly, our set of methodologies opens the door for strategy scholars to use easily available, yet underutilized text, image and video data in a wide variety of settings.

References

Athey, S., 2018. The impact of machine learning on economics. In Economics of Artificial Intelligence. University of Chicago Press.

Antweiler, W. and Frank, M.Z., 2004. Is all that talk just noise? The information content of internet stock message boards. The Journal of finance, 59(3), pp.1259-1294.

Bandiera, O; L Guiso, A Prat, R Sadun, "What do CEOs do?", Review of Financial Studies, 2017, Forthcoming

Bandiera, O., Lemos, R., Prat, A. and Sadun, R., 2013. Managing the family firm: evidence from CEOs at work (No. w19722). National Bureau of Economic Research.

Barley, S.R., 1983. Semiotics and the study of occupational and organizational cultures. Administrative Science Quarterly, pp.393-413.

Barley, S. R. 1990, "Images of imaging: Notes on doing longitudinal field work." Organization Science. 1:220-247.

Barr, P.S., 1998. Adapting to unfamiliar environmental events: A look at the evolution of interpretation and its role in strategic change. Organization Science, 9(6), pp.644-669.

Baum, J.R., Locke, E.A. and Kirkpatrick, S.A., 1998. A longitudinal study of the relation of vision and vision communication to venture growth in entrepreneurial firms. Journal of applied psychology, 83(1), p.43.

Bettis, R.A., Gambardella, A., Helfat, C. and Mitchell, W., 2015. Qualitative empirical research in strategic management. Strategic Management Journal, 36(5), pp.637-639.

Blei, D. M., et al. (2003). "Latent Dirichlet Allocation." Journal of Machine Learning Research 3.

Bromiley, P. and A. Marcus, "The Deterrent to Dubious Corporate Behavior: Profitability, Probability and Safety Recalls," Strategic Management Journal, 10 (1989), 233-250.

Calori, R., Johnson, G. and Sarnin, P., 1994. CEOs' cognitive maps and the scope of the organization. Strategic Management Journal, 15(6), pp.437-457.

Chang, J., et al. (2009). "Reading Tea Leaves: How Humans Interpret Topic Models." Neural Information Processing Systems.

Chatterjee, A. and Hambrick, D.C., 2011. Executive personality, capability cues, and risk taking: How narcissistic CEOs react to their successes and stumbles. Administrative Science Quarterly, 56(2), pp.202-237.

D'Aveni, R.A. and MacMillan, I.C., 1990. Crisis and the content of managerial communications: A study of the focus of attention of top managers in surviving and failing firms. Administrative science Quarterly, pp.634-657.

Daft, R.L., Sormunen, J. and Parks, D., 1988. Chief executive scanning, environmental characteristics, and company performance: An empirical study. Strategic Management Journal, 9(2), pp.123-139.

Delgado-García, J.B., La Fuente-Sabaté, D. and Manuel, J., 2010. How do CEO emotions matter? Impact of CEO affective traits on strategic and performance conformity in the Spanish banking industry. Strategic Management Journal, 31(5), pp.562-574.

Duffner, S., 2008. Face image analysis with convolutional neural networks.

Duncan, R.B., 1972. Characteristics of organizational environments and perceived environmental uncertainty. Administrative science Quarterly, pp.313-327.

Dunning, J.H., 1998. American investment in British manufacturing industry. Taylor & Francis US.

Eisenhardt E. 2014. Theory building from cases. Available at: http://smj.strategicmanagement.net/.

Freedman, A. and Medway, P., 1994. Locating genre studies: Antecedents and prospects. Genre and the new rhetoric, pp.1-20.

Gamache, D.L., McNamara, G., Mannor, M.J. and Johnson, R.E., 2015. Motivated to acquire? The impact of CEO regulatory focus on firm acquisitions. Academy of Management Journal, 58(4), pp.1261-1282.

Gao, C., Zuzul, T., Jones, G. and Khanna, T., 2017. Overcoming Institutional Voids: A Reputation-Based View of Long-Run Survival. Strategic Management Journal.

Gioia D. 2014. A 1st-order/2nd-order qualitative approach to understanding strategic management. Available at: http://smj.strategicmanagement.net/.

Glaeser, E.L., Kominers, S.D., Luca, M. and Naik, N., 2018. Big data and big cities: The promises and limitations of improved measures of urban life. Economic Inquiry, 56(1), pp.114-137.

Griffiths, T. L. and M. Steyvers (2004). "Finding scientific topics." PNAS 101.

Hambrick, D.C. and Mason, P.A., 1984. Upper echelons: The organization as a reflection of its top managers. Academy of Management Review, 9(2), pp.193-206.

Hambrick, D.C. and Macmillan, I.C., 1985. Efficiency of product R&D in business units: The role of strategic context. Academy of Management Journal, 28(3), pp.527-547.

Helfat, C.E. and Peteraf, M.A., 2015. Managerial cognitive capabilities and the microfoundations of dynamic capabilities. Strategic Management Journal, 36(6), pp.831-850.

Hendricks, K. B. and V. R. Singhal, "Quality Awards and the Market Value of the Firm: An Empirical Investigation," Management Science, 42 (1996), 415-436.

Herrmann, P. and Nadkarni, S., 2014. Managing strategic change: The duality of CEO personality. Strategic Management Journal, 35(9), pp.1318-1342.

Hill, R.C. and Levenhagen, M., 1995. Metaphors and mental models: Sensemaking and sensegiving in innovative and entrepreneurial activities. Journal of Management, 21(6), pp.1057-1074.

Hiller, N.J. and Hambrick, D.C., 2005. Conceptualizing executive hubris: the role of (hyper-) core self-evaluations in strategic decision-making. Strategic Management Journal, 26(4), pp.297-319.

Hornik, Kurt, and Bettina Grün. "topicmodels: An R package for fitting topic models." Journal of Statistical Software 40.13 (2011): 1-30.

Huang, A. Lehavy, R., Zang, A., and Zheng, R. 2017. “Analyst Information Discovery and Interpretation Roles: A Topic Modeling Approach.” Management Science, Forthcoming.

Jacobsen, R., 1988. The persistence of abnormal returns. Strategic management journal, 9(5), pp.415-430.

Jockers, M., 2017. Package ‘syuzhet’. URL: https://cran. r-project. org/web/packages/syuzhet.

Jones, G. and Khanna, T., 2006. Bringing history (back) into international business. Journal of International Business Studies, 37(4), pp.453-468.

Jones, G., 2005. Multinationals and global capitalism: From the nineteenth to the twenty first century. Oxford University Press on Demand.

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

Kilduff, M., 1993. Deconstructing organizations. Academy of Management Review, 18(1), pp.13-31.

Khanna, Tarun and Krishna G Palepu (with Richard Bullock), Winning in emerging markets, Harvard business press, 2010.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J. and Mullainathan, S., 2017. Human decisions and machine predictions. The quarterly journal of economics, 133(1), pp.237-293.

Kogut, B.M. ed., 1993. Country competitiveness: Technology and the organizing of work. Oxford University Press on Demand.

Larcker, D.F. and Zakolyukina, A.A., 2012. Detecting deceptive discussions in conference calls. Journal of Accounting Research, 50(2), pp.495-540.

Lefebvre, L.A., Mason, R. and Lefebvre, E., 1997. The influence prism in SMEs: The power of CEOs' perceptions on technology policy and its organizational impacts. Management Science, 43(6), pp.856-878.

Lehavy R, Li F, Merkley K (2011) The effect of annual report readability on analyst following and the properties of their earnings forecasts. Accounting Rev. 86(3):1087–1115

Lovins, J. B. (1968). "Development of a Stemming Algorithm." Mechanical Translation and Computational Linguistics 11.

Loughran T, McDonald B (2016) Textual analysis in accounting and finance: A survey. J. Accounting Res. 54(4):1187–1230

Menon, Anoop R., Clarence Lee, and Haris Tabakovic. 2018. “Using Machine Learning to Predict High-Impact General Technologies”, Working Paper.

Mintzberg, H., 1987. Crafting strategy (pp. 66-75). Boston, MA, USA: Harvard Business School Press.

Mullainathan, S. and Spiess, J., 2017. Machine learning: an applied econometric approach. Journal of Economic Perspectives, 31(2), pp.87-106.

Norton, R.W., 1978. Foundation of a communicator style construct. *Human Communication Research*, *4*(2), pp.99-112.

Perloff, R.M., 2013. *Political communication: Politics, press, and public in America*. Routledge.

Phillips, N. and Brown, J.L., 1993. Analyzing communication in and around organizations: A critical hermeneutic approach. Academy of Management journal, 36(6), pp.1547-1576.

Ragin, C.C., 2014. The comparative method: Moving beyond qualitative and quantitative strategies. Univ of California Press.

Ramage, D., Hall, D., Nallapati, R. and Manning, C.D., 2009. "Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora." Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing. (pp. 248-256). Association for Computational Linguistics.

Raudenbush, S.W. and A.S. Bryk. 2002. Hierarchical linear models: Applications and data analysis methods (Vol. 1). Thousand Oaks, CA: Sage.

Riedl, M. and C. Biemann (2012). "Text segmentation with topic models." Journal for Language Technology and Computational Linguistics 27(1).

Salancik, G.R. and Meindl, J.R., 1984. Corporate attributions as strategic illusions of management control. Administrative science quarterly, pp.238-254.

Suddaby, R. and Greenwood, R., 2005. Rhetorical strategies of legitimacy. Administrative science quarterly, 50(1), pp.35-67.

Tasker, S.C., 1998. Bridging the information gap: Quarterly conference calls as a medium for voluntary disclosure. Review of Accounting Studies, 3(1), pp.137-167.

van Dyk, D. A. and T. Park (2008). "Partially Collapsed Gibbs Samplers." Journal of the American Statistical Association 103(482): 790-796.

Watzlawick, Paul, and J. H. Beavin. "B., & Jackson, DD (1967)." Pragmatics of human communication (1967).

Westley, F. and Mintzberg, H., 1989. Visionary leadership and strategic management. Strategic management journal, 10(S1), pp.17-32.

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

Yao, L., et al. (2009). "Efficient methods for topic model inference on streaming document collections." Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM.

Yu, Z. and Zhang, C., 2015, November. Image based static facial expression recognition with multiple deep network learning. In Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (pp. 435-442). ACM.

Figures and Tables

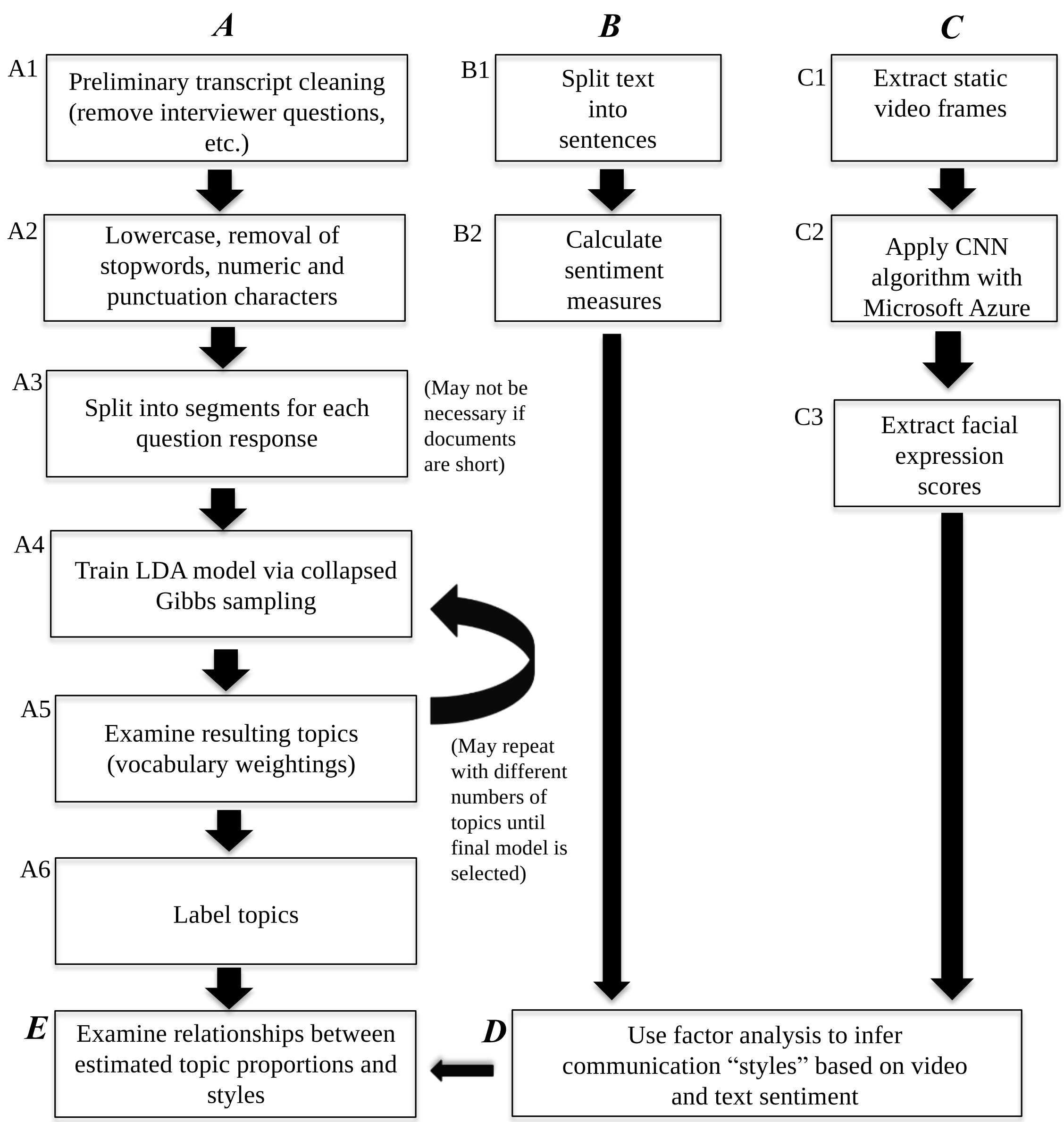

Figure 1: Methods Overview

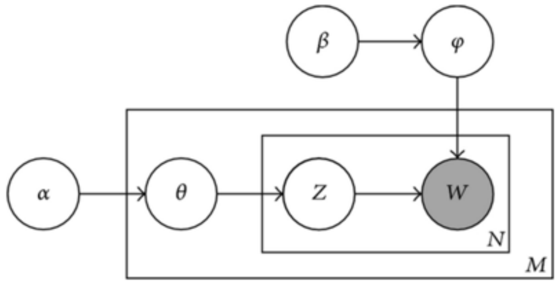

Figure 2: Plate Diagram for LDA

[1]: Personal Growth	[2]: Hospitality	[3]: Management	[4]: Politics	[5]: Opportunity	[6]: Professional Relationships	[7]: Family Difficulties	[8]: Branding	[9]: Economic Growth	[10]: Law
experience	people	management	president	able	people	difficult	brand	growth	law
helped	training	culture	minister	start	understand	decided	market	believe	court
understand	hotel	manager	finance	environment	live	daughter	products	economy	issue
learn	service	team	prime	looking	respect	tough	brands	future	firm
learned	staff	level	speak	opportunity	talk	extremely	consumer	growing	laws
friends	hotels	managers	meet	south	jobs	ready	joint	institutions	corruption
hard	job	organization	happened	opportunities	fighting	course	marketing	grow	deal
easy	course	top	relationship	join	living	husband	retail	result	act
run	trained	corporate	meeting	grow	discussion	decision	venture	steel	field
manage	train	key	party	coming	value	step	share	leaders	control

[11]: Education	[12]: Investment	[13]: Organized Labor	[14]: Childhood	[15]: Communication	[16]: Development	[17]: Banking	[18]: Media	[19]: Infrastructure	[20]: Change
school	million	industry	father	call	development	bank	film	water	change
education	market	labor	life	idea	terms	banks	media	city	comes
children	sold	period	war	created	economic	banking	happening	real	changed
schools	buy	happened	sit	research	process	crisis	television	build	hand
university	sell	union	age	mobile	corporate	capital	advertising	built	grew
students	bought	situation	child	center	leadership	credit	agency	land	taking
college	billion	plant	friend	form	example	insurance	person	significant	instance
foundation	license	exactly	outside	phone	human	central	people	looked	changes
girls	half	demand	morning	issues	resources	financial	art	building	changing
program	selling	industrial	children	willing	talking	money	agencies	brought	matter

[21]: Community	[22]: Manufacturing	[23]: Energy	[24]: Society	[25]: Taxes	[26]: Narratives	[27]: Money	[28]: Family History	[29]: Problems	[30]: Interest
create	project	example	society	government	days	money	family	problem	percent
community	production	gas	feel	power	story	pay	person	question	trust
challenge	factory	energy	social	tax	name	cost	generation	talk	balance
support	successful	sense	impact	course	managing	whatever	house	problems	rate
local	supply	means	ago	lost	kept	month	grandfather	car	third
challenges	basically	natural	issues	gone	months	dollars	brother	happens	obviously
possible	major	distribution	organization	governments	mind	spend	brought	bad	funds
value	close	common	giving	resources	director	job	uncle	talking	start
communities	market	outside	wrong	taxes	happened	profit	head	love	rates
taking	huge	experience	responsibility	national	book	costs	happen	suddenly	term

[31]: Fair Trade	[32]: Corporate Boards	[33]: Retail	[34]: Capital Markets	[35]: Regional Development	[36]: Globalization	[37]: Healthcare	[38]: Public Sector	[39]: Women's Empowerment	[40]: FX
tea	board	customers	markets	global	world	happy	sector	women	foreign
example	members	quality	investment	east	technology	care	public	home	exchange
poor	shares	price	capital	build	especially	heart	private	woman	economy
times	chairman	product	emerging	middle	continue	thank	system	men	international
success	holding	buy	infrastructure	partner	western	hospital	financial	village	abroad
trade	professional	sell	projects	local	developed	operation	political	mother	export
consumers	management	shop	model	region	developing	healthcare	governance	involved	course
poverty	shareholders	customer	opportunity	standards	largest	vision	view	access	import
fair	decisions	sales	investors	office	remember	health	policy	workers	exports
low	joined	competition	continent	eventually	strong	believe	improve	happened	equipment

Figure 3: Top Terms for Estimated LDA Model

Figure 4: Examples of Static Frames Representing the Eight Facial Expressions (Score in bold text)

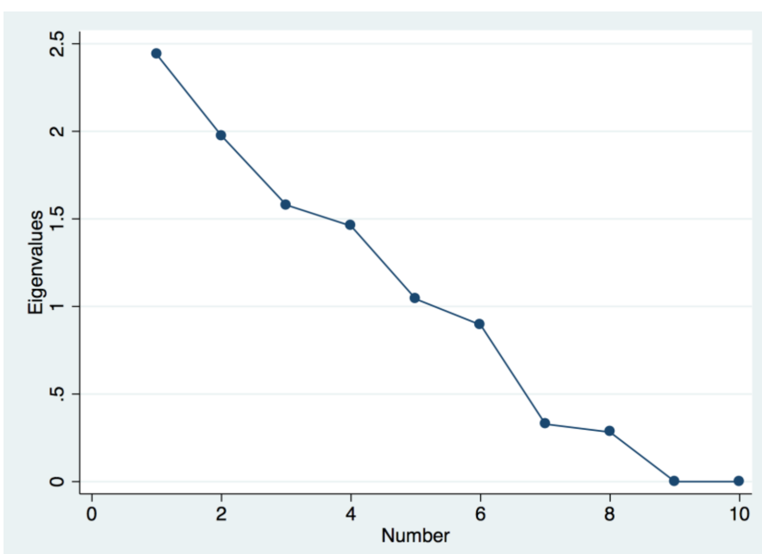

Figure 5: Scree Plot of Factor Analysis

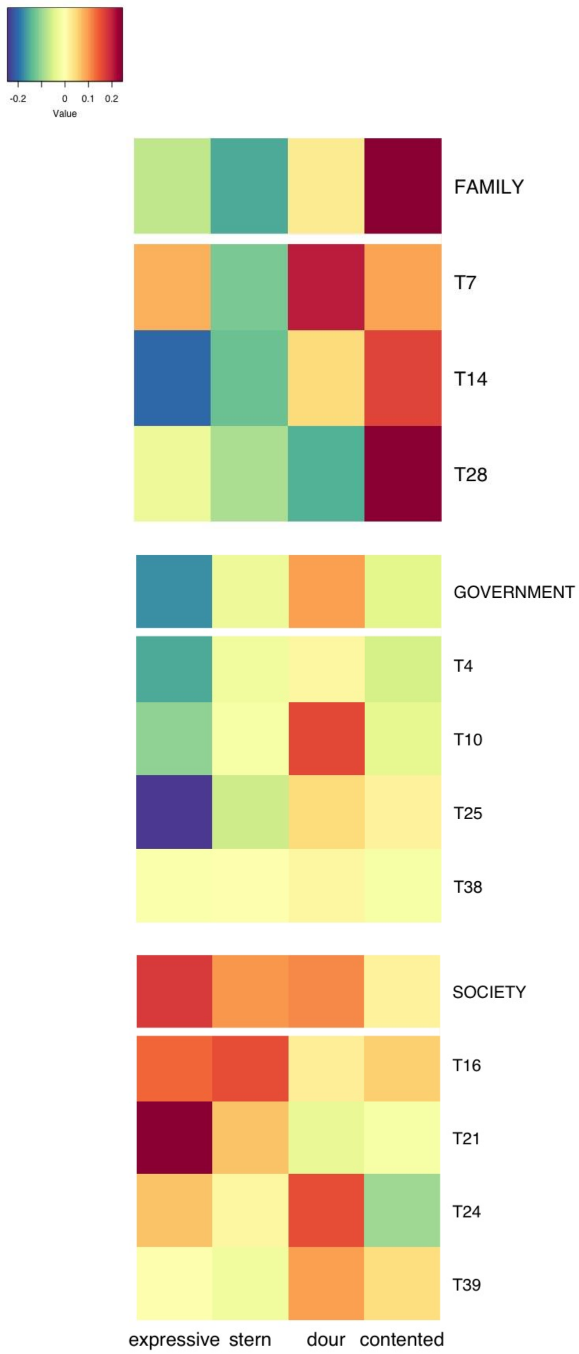

Figure 6. Heatmap of Correlations Between Communication Style and Selected Topics.

Note: Each square signals the value of the correlation of a given communication style with a given topic. For example, the "expressive" style is correlated with "Society" Topics at $r = .33$ and with "Government" Topics at $r = -.41$. Topics are grouped by "Family", "Government", "Society." The first row of each topic group represents correlations of each communication style with the topic group as a whole.

Table 1: Factor Loadings for Communication Styles

		Style		
Variable	**Expressive**	**Stern**	**Dour**	**Contented**
Text: Positive	**0.77**	0.28	**-0.56**	-0.03
Text: Negative	**-0.77**	-0.28	**0.56**	0.03
Video: Anger	0.17	**0.59**	**0.6**	0.16
Video: Contempt	0.16	0.26	0.05	0.38
Video: Disgust	0.37	**0.61**	0.4	0.37
Video: Fear	**0.59**	-0.29	0.45	-0.44
Video: Happiness	0.04	**-0.65**	-0.21	**0.6**
Video: Neutral	**-0.58**	**0.58**	-0.29	-0.47
Video: Sadness	0.28	-0.25	0.13	0.3
Video: Surprise	**0.55**	-0.37	0.27	**-0.54**

Table 2: How Styles Predict Topics

	Dependent Variable:								
	Family Topics			Government Topics			Society Topics		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Expressive	-0.003 (0.004)	-0.01 (0.004)	-0.002 (0.005)	-0.02 (0.01)	-0.02 (0.01)	-0.03 (0.01)	0.02 (0.01)	0.02 (0.01)	0.01 (0.01)
	p = 0.38	p = 0.18	p = 0.65	p = 0.003	p = 0.004	p = 0.001	p = 0.005	p = 0.03	p = 0.05
Stern	-0.01 (0.01)	-0.01 (0.005)	-0.01 (0.01)	-0.004 (0.01)	-0.004 (0.01)	-0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
	p = 0.16	p = 0.22	p = 0.31	p = 0.49	p = 0.48	p = 0.29	p = 0.15	p = 0.10	p = 0.12
Dour	0.001 (0.01)	0.002 (0.01)	0.001 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
	p = 0.83	p = 0.74	p = 0.90	p = 0.03	p = 0.03	p = 0.01	p = 0.03	p = 0.02	p = 0.02
Contented	0.01 (0.004)	0.01 (0.004)	0.01 (0.004)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	0.002 (0.01)	).0001 (0.01)	- 0.004 (0.01)
	p = 0.01	p = 0.03	p = 0.06	p = 0.31	p = 0.35	p = 0.33	p = 0.73	p = 0.99	p = 0.54
Female		0.03 (0.02)	0.02 (0.02)		-0.004 (0.01)	-0.004 (0.02)		0.04 (0.02)	0.04 (0.02)
		p = 0.10	p = 0.18		p = 0.77	p = 0.79		p = 0.07	p = 0.05
Asia			0.02 (0.01)			0.0004 (0.02)			0.04 (0.01)
			p = 0.03			p = 0.99			p = 0.005
Africa			0.003 (0.02)			0.04 (0.02)			0.08 (0.02)
			p = 0.89			p = 0.06			p = 0.0005
Lat.Am.			0.003 (0.01)			0.01 (0.02)			0.06 (0.02)
			p = 0.71			p = 0.63			p = 0.001
Observations	59	59	59	59	59	59	59	59	59
R-squared	0.09	0.15	0.18	0.24	0.24	0.3	0.19	0.26	0.42

Note: OLS estimates with robust standard errors in parentheses.

APPENDIX: Oral History Archives

Archive Name	Source
Harvard University Creating Emerging Markets Initiative	http://www.hbs.edu/creating-emerging-markets/interviews/Pages/default.aspx
UCLA Center for Oral History	http://oralhistory.library.ucla.edu/Browse.do?coreDescCvPk=27901&Subject=Business
Columbia University Oral History Archives	http://library.columbia.edu/locations/ccoh.html
World Bank Oral history archive	http://oralhistory.worldbank.org/
Indiana University Center for the Study of History and Memory	http://www.indiana.edu/~cshm/alphalist.html
University of California Berkeley Oral History Collection	http://www.lib.berkeley.edu/libraries/bancroft-library/oral-history-center/search-oral-histories
University of Connecticut Oral History	http://www.oralhistory.uconn.edu/catalog.html
University of Kentucky Louie B. Nunn Center for Oral History	https://kentuckyoralhistory.org/
The British Library	https://www.bl.uk/collection-guides/oral-histories-of-business-and-finance
The History Factory	http://www.historyfactory.com/
History Associates	https://www.historyassociates.com/who-we-serve/our-clients/
University of Florida Oral History Collections	http://ufdc.ufl.edu/ohfbl