Measuring Corporate Culture Using Machine Learning*

Kai Li
Sauder School of Business
University of British Columbia
2053 Main Mall, Vancouver, BC V6T 1Z2
Canada
kai.li@sauder.ubc.ca

Feng Mai
School of Business
Stevens Institute of Technology
1 Castle Point Terrace
Hoboken, NJ 07030, USA
feng.mai@stevens.edu

Rui Shen
Nanyang Business School
Nanyang Technological University
50 Nanyang Avenue
Singapore 639798
shenrui@ntu.edu.sg

Xinyan Yan
School of Business Administration
University of Dayton
300 College Park
Dayton, OH 45469-2251
xyan02@udayton.edu

This version: December 2018

Abstract

Using one of the latest machine learning techniques—the word embedding model (Mikolov et al. 2013)—and 217,387 earnings call transcripts, we obtain five corporate cultural values: innovation, integrity, quality, respect, and teamwork (Guiso, Sapienza, and Zingales 2015) for 77,541 firm-year observations over the period 2001–2018. We conduct a large number of tests to validate our measures and demonstrate the advantages of our approach over several alternative methods. We show that corporate culture plays an important role in deal incidence and merger pairing, and that post-merger, acquirers' cultural values are positively associated with their target firms' cultural values pre-merger.

Keywords: machine learning; word embedding; unsupervised learning; corporate culture; cultural fit; acculturation; mergers and acquisitions

JEL classification: C45; G34

* We are grateful for helpful comments from Serdar Aldatmaz, Ling Cen, Sandeep Dahiya, Sudipto Dasgupta, D.J. Fairhurst, Ning Gao (discussant), Jon Garfinkel, Randall Heron, Gilles Hilary, Harrison Hong (discussant), Seokwoo Lee, Jiasun Li, Erik Lie, Xing Liu, Guangli Lu, Vojislav Maksimovic (discussant), David McLean, Vladimir Mukharlyamov, Amrita Nain, Andy Puckett, Yiming Qian, John Ries, Geoff Tate (discussant), Rajesh Vijayaraghavan, Aazam Virani (discussant), Kate Waldock, Xiaofei Zhao, seminar participants at Chinese University of Hong Kong, George Mason University, Georgetown University, University of Iowa, and Washington State University, and conference participants at the Drexel Corporate Governance Conference early ideas session, the ECGC International Workshop, University of Tennessee Smokey Mountain Finance Conference, the Rising Stars Conference, the China International Conference in Finance, the Northern Finance Association Meetings, the Financial Management Association Meetings, and the Inaugural Conference on Intelligent Information Retrieval in Accounting and Finance. We also thank our research assistants Xing Liu, Guangli Lu, and Iris Wang. We acknowledge financial support from the OSU Risk Institute, the Social Sciences and Humanities Research Council of Canada (Grant Number: 435-2018-0037), and Singapore Ministry of Education Academic Research Fund Tier 1 (Grant Number: RG92/14). All errors are our own.

Measuring Corporate Culture Using Machine Learning

Abstract

Using one of the latest machine learning techniques—the word embedding model (Mikolov et al. 2013)—and 217,387 earnings call transcripts, we obtain five corporate cultural values: innovation, integrity, quality, respect, and teamwork (Guiso, Sapienza, and Zingales 2015) for 77,541 firm-year observations over the period 2001–2018. We conduct a large number of tests to validate our measures and demonstrate the advantages of our approach over several alternative methods. We show that corporate culture plays an important role in deal incidence and merger pairing, and that post-merger, acquirers' cultural values are positively associated with their target firms' cultural values pre-merger.

Keywords: machine learning; word embedding; unsupervised learning; corporate culture; cultural fit; acculturation; mergers and acquisitions

JEL classification: C45; G34

1. Introduction

In a recent survey of North American Chief Executive Officers (CEOs) and Chief Financial Officers (CFOs), over half of senior executives view corporate culture as one of the top three factors that affect their firm's value, and over 90% of them believe that improving corporate culture would increase firm value. Cultural fit in mergers and acquisitions (M&As) is so important that about half of executives would walk away from a culturally misaligned target (Graham, Grennan, Harvey, and Rajgopal 2017).

What is corporate culture? According to O'Reilly and Chatman (1996, p. 160), corporate culture is "a system of shared values (that define what is important) and norms that define appropriate attitudes and behaviors for organizational members (how to feel and behave)." Why does corporate culture matter? Culture matters because employees will inevitably face choices that cannot be properly regulated ex ante (O'Reilly 1989; Kreps 1990). Unlike deeply-held national cultural values, corporate culture is defined by a set of operational practices and is path-dependent (Weber, Shenkar, and Raveh 1996; Guiso, Sapienza, and Zingales 2015; Graham, Grennan, Harvey, and Rajgopal 2016). Extant literature has limited large sample evidence on the role of corporate culture in firm policy and performance (see our literature review in Section 2.1), possibly because the notion of corporate culture is somewhat nebulous, and thus raises numerous measurement issues in empirical research (see the review by Zingales 2015, and interview evidence in Graham et al. 2016).

How do we measure corporate culture? Our starting point is the often-mentioned values by the S&P 500 firms on their corporate websites (Guiso, Sapienza, and Zingales 2015): innovation, integrity, quality, respect, and teamwork, each of which serves as a 'value word'—i.e., it expresses a core corporate value. We then use one of the latest machine learning techniques—the word embedding model (Mikolov et al. 2013)—and earnings call transcripts to measure those values.

We make an important methodological contribution by introducing a novel machine learning method to quantify text—the word embedding model (Mikolov et al. 2013, *aka* word2vec), a neural network model that measures associations between words using the context in which they appear. Using this machine learning method, we construct what we term a 'culture dictionary' of words and phrases culled from earnings call transcripts that most frequently appear in close association with each cultural value. For example, the method identifies words such as *creativity, agility,* and *nimbleness* and phrases such as *technological advancement* and *the proof is in the pudding* are frequently used by corporate executives to promote the cultural value of *innovation*. The cultural value of *innovation* is then based on a frequency count of those words and phrases in earnings call transcripts.

The word embedding model is based on a simple, time-tested concept in linguistics: Words tend to co-occur with neighboring words with similar meanings (Harris 1954); the model thus identifies synonyms from neighboring words. To illustrate, suppose we want to examine the relationship between three words: *collective, partnership*, and *governance*. We can start by counting how many times any neighboring words appear near these three specific words in a collection of documents (in our particular setting that would be the entire collection of earnings calls). We find that *share, fruitful,* and *joint* tend to appear most often near *collective* and *partnership*; and *oversight* and *proper* tend to appear most often near *governance*. We record the number of times those five words—*share, fruitful, joint, oversight,* and *proper*—appear in a vector for each of these three words. In this case, we can use a vector [4, 5, 5, 0, 1] to represent *collective* where 4 is the number of times the word *share* appears close to the word *collective*, and 5 is the number of times the word *fruitful* appears close to the word *collective*, etc. Similarly, we can use a vector [3, 6, 7, 0, 0] to represent *partnership*, and a vector [0, 0, 1, 10, 9] to represent *governance*. Such vector representation of a word allows us to compute the association between any pair of words using the cosine similarity of their underlying vectors. The cosine similarity between *collective* and *partnership* is 0.97 and the cosine similarity between *collective* and *governance* is 0.13. We conclude that *collective* and *partnership* are semantically closer to each other than *collective* and *governance* based on the textual context in which neighboring words are found. Put differently, *partnership* is a closer synonym than *governance* to *collective*. In practice, the algorithm uses the entire collection of earnings calls and transforms any word to a vector of fixed dimension, which allows us to compute the cosine similarity between any two words.

To measure corporate culture, we start with seed words defining each cultural value and then expand them into our culture dictionary comprising hundreds of words and phrases. As an example, Guiso, Sapienza, and Zingales (2015) list *collaboration* and *cooperation* as the seed words for the cultural value of *teamwork*. We can compute the cosine similarity between each unique word (or phrase) in earnings calls with *collaboration* and *cooperation*. For each seed word, we select the top *n* closest synonyms. We expand this set of synonyms one more time by finding each synonym's top *n* most similar words. Out of the final set of at most $2n^2$ synonyms, we select the top words (or phrases) with the closest associations to the average vector representing *collaboration* and *cooperation*. This snowball sampling procedure provides a culture dictionary that captures the cultural value of *teamwork*. At the firm-year level, we obtain the cultural value of *teamwork* by counting the frequency of those words (or phrases) in the dictionary defining the value.

Using 217,387 earnings calls from the Thomson Reuters' StreetEvents (SE) database and Factiva over the period 2001-2018, we first train the word embedding model and then obtain corporate cultural values for 8,427 unique firms (77,541 firm-year observations). We validate our corporate culture measures using well-established markers for best practices in corporate innovation, integrity, product quality, respect, and teamwork, and show that our measures are positively and significantly associated with those markers. We also compare our main measures based on the questions and answers (QA) section of earnings calls with alternative measures based on: i) the entire call, including the management presentation section and QA section; ii) a simple count of Guiso, Sapienza, and Zingales' seed words including the value word (e.g., *innovation*) in the QA section; and iii) by applying the word embedding model to the Management's Discussion and Analysis (MD&A) section of annual reports (10-Ks). We show that applying the word embedding model to the QA section of earnings calls represents a significant improvement to existing approaches to measuring corporate culture based on validation tests.

For an application of our measures, we examine the role of corporate culture in M&As using a sample of close to 8,000 deals over the period 2003–2017. We first show that firms that score high on the cultural value of innovation are more likely to be acquirers, whereas firms that score high on the cultural values of integrity, quality, and respect are less likely to be acquirers. In terms of merger pairing, we find that firms closer in cultural values are more likely to do a deal together. In terms of post-merger outcome, we find limited evidence that cultural fit matters in post-merger integration, retention, and deal performance. These findings might not be surprising given that, as we have shown, culturally misaligned pairs do not initiate a deal in the first place. Finally, we show that post-merger, acquirers' cultural values are positively associated with their target firms' cultural values pre-merger, suggesting acculturation. We conclude that corporate culture plays a significant role in deal incidence and merger pairing, and that importantly, corporate culture itself is also shaped by M&As.

Our paper differs from prior work and thus contributes to the literature in a number of ways. First, to the best of our knowledge, our paper is one of the first in finance to employ deep learning—a new machine learning paradigm that uses layers of artificial neural networks to learn from unstructured data—to automatically process earnings calls and score cultural values. We show that compared to known alternative approaches, our machine learning method yields a high-quality culture dictionary useful for measuring corporate culture, and is also scalable to a large collection of textual data. Second, our paper provides new insight into the role of corporate culture in M&As by employing a richer set of measures for cultural congruence beyond a simple distance measure as commonly used in prior work, and by highlighting the tension between external-facing corporate culture (innovation) and internal-facing corporate culture (quality) and examining its role in post-merger integration. Third and finally, we also examine whether and how M&As shape corporate culture, contributing to our understanding of the broad question of how corporate culture evolves slowly over time. We are one of the first in finance to provide large sample evidence on acculturation whereby post-merger, the culture of acquirers exhibits traces of that of their target firms.

2. Measuring Corporate Culture Using Machine Learning

2.1. Prior literature on corporate culture

A number of recent papers explore the relationship between corporate culture and firm policy using proxies for the former. Cronqvist, Low, and Nilsson (2009) find that a broad range of spinoffs' financing and investment policies appear to be more similar to the policies of their parents than to those of similar-sized industry peers, even in cases in which the spinoffs are run by outside CEOs. The authors measure corporate culture with firm fixed effects and indices on employee relations and diversity from the KLD Research & Analytic. Using the annual rankings of the *Best Companies to Work for in America* by the Great Place to Work Institute (GPWI) as a proxy for firms with a strong culture of trust (SCT), Bargeron, Smith, and Lehn (2015) find that the size of acquisitions announced by SCT firms is significantly smaller than the size of acquisitions announced by other firms. Furthermore, when SCT firms make large acquisitions, their returns are lower, and they are more likely to suffer a loss in their GPWI ranking compared to other SCT firms. Using the incidence of options backdating as a proxy for an unethical corporate culture, Biggerstaff, Cicero, and Puckett (2015) show that these firms are more likely to commit financial fraud to overstate earnings. Using corporate executives' personal traits such as reckless behavior or frugality as a proxy for the corporate culture of the firms these executives manage, Davidson, Dey, and Smith (2015) find that firms whose CEOs and CFOs have a legal record are more likely to commit fraud, and firms with extravagant CEOs are associated with a loose control environment characterized by more fraud and unintentional material reporting errors. Using ties to multinationals as a proxy for a corporate culture of transparency, Braguinsky and Mityakov (2015) find that private Russian firms with closer ties to multinationals are associated with improved transparency of wage reporting and reduced accounting fraud. Using the GPWI's survey of employees' perception of top management as a proxy for the corporate culture of integrity, Guiso, Sapienza, and Zingales (2015) find that integrity is strongly associated with firm value. Using the similarity in firms' corporate social responsibility characteristics as a proxy for cultural similarity, Bereskin, Byun, Officer, and Oh (2017) find that culturally similar firms are more likely to merge. Moreover, these mergers are associated with greater synergies, superior long-run operating performance, and fewer write-offs of goodwill. Using the last names of a firm's leaders to capture corporate risk culture, Pan, Siegel, and Wang (2017) examine its effect on firm policy.

The exceptions are Grennan (2013) and Fiordelisi and Ricci (2014).[1] Using the seven cultural attributes of O'Reilly, Chatman, and Caldwell (1991) and O'Reilly, Caldwell, Chatman, and Doerr (2014), and their synonyms from WordNet, Grennan (2013) scores cultural values by counting the frequency of those synonyms in employee reviews. She finds that corporate culture is an important channel through which corporate governance affects firm value. Using the four cultural attributes of the Competing Value Framework (Cameron, De Graff, Quinn, and Thakor 2006)—collaborate, compete, control, and create—and synonyms provided by those authors, Fiordelisi and Ricci (2014) identify additional synonyms from the Harvard IV-4 Psychosocial Dictionary, and score cultural values by counting the frequency of those synonyms in 10-Ks. They find that CEO turnover-to-performance sensitivity is strengthened in firms with an internal focus and is weakened in firms with an external focus, and that firms with an external focus are less likely to have an insider CEO successor.

One positive feature of using an external dictionary to identify synonyms is that its composition is beyond the control of a researcher; i.e., a researcher cannot pick and choose which words are semantically close to the word "*collaborate*." However, external dictionaries have limitations for our purpose. WordNet is a lexical database for the English language, but does not contain certain words commonly used in finance and accounting. For example, words frequently shown up in earnings calls such as *bylaw*, *verticalization,* and *standardization* have no synonyms in WordNet. In addition, associations among words in WordNet are derived from general English usage and may not be finance or accounting context-specific. For example, the only synonym for *offshore* in WordNet is *seaward*, whereas the word embedding method that we employ identifies *outsourcing, regionalization,* and *globalization* as

[1] It is worth noting that our five values largely overlap with alternative corporate cultural paradigms, including the seven value system of O'Reilly, Chatman, and Caldwell (1991) and O'Reilly, Caldwell, Chatman, and Doerr (2014): adaptability, collaboration, customer-orientation, detail-orientation, integrity, results-orientation, and transparency; the four value system of the Competing Value Framework (Cameron, De Graff, Quinn, and Thakor 2006): collaborate, compete, control, and create; and the seven cultural values of Graham et al. (2016): adaptability, collaboration, community, customer-oriented, detail-oriented, integrity, and results-oriented.

synonyms. The latter set is arguably more specific to the context of earnings calls, during which relocating business overseas is more likely to be the topic than the literal meaning of "towards the ocean." Relatedly, English words have many meanings, and a word categorization scheme derived for psychology and sociology such as the Harvard IV-4 Psychosocial Dictionary may not translate well into the realm of business (Loughran and McDonald 2011). Moreover, 10-Ks are legal documents prepared and/or vetted by corporate lawyers and investor relations, hindering their usefulness for measuring corporate culture.

2.2. An introduction to machine learning

There are two main types of machine learning algorithms: supervised learning and unsupervised learning. Under supervised learning, the goal is to make predictions by learning patterns from labeled data. To capture a specific concept (such as a cultural value) from text using the supervised learning approach, one needs a considerable number of documents as training samples that are labeled in terms of their relatedness to the concept. This requirement is typically not an issue for some applications in finance because the document labels are directly observable in the form of firm outcomes. For example, Loughran and McDonald (2011) use 10-K filing day returns to help identify negative words in finance, Bodnaruk, Loughran, and McDonald (2015) use dividend omissions or underfunded pensions to construct lexicons from 10-Ks for financial constraints, and Hoberg and Maksimovic (2015) search 10-K texts for any statement indicating that a firm may have to delay its investments due to financial liquidity issues. Such labeled documents, however, are not available for identifying cultural values.

Under unsupervised learning, the goal is to model the underlying structure or distribution in the data in order to learn more about the data. Typical tasks are clustering, dimension reduction, and feature extraction. We employ unsupervised machine learning of earnings calls to measure corporate culture; i.e., our approach can extract context-specific words and phrases from calls without relying on human-labeled documents.

2.3. Our approach to measuring corporate culture

When asked, “Which of the following have been most influential in setting your firm’s current culture?” From a list of possibilities including current CEO, owners, founders, our reputation or image in the market place, internal policies and procedures, and hard times experienced, more than half of the top executives surveyed by Graham et al. (2017) identify the current CEO to be the most important driver of a firm’s current culture. Consistent with the survey evidence, prior studies such as Biggerstaff, Cicero, and Puckett (2015), Davidson, Dey, and Smith (2015), and Guiso, Sapienza, and Zingales (2015) use CEO attributes and behaviors to proxy for corporate culture. We thus expect earnings calls, as a commonly-employed external corporate communication channel involving mostly CEOs and sometimes other top executives speaking to analysts, to reveal the set of values that are important to those corporate leaders and their company; Graham et al. (2016) also recommend earnings calls as the primary avenue for capturing corporate culture.

Our starting point is the five most often-mentioned values by the S&P 500 firms on their corporate websites (Guiso, Sapienza, and Zingales 2015): *innovation* (80% of the time), *integrity* (70%), *quality* (60%), *respect* (70%), and *teamwork* (50%). Guiso, Sapienza, and Zingales (2015) also provide units of meaning (i.e., seed words) for each value after checking all other words that are clustered with a value by each firm and their frequency across firms.[2] We then use machine learning to develop an expanded, context-specific dictionary for measuring cultural values. To achieve this, we make an important methodological contribution to quantify text by using the word embedding model (Mikolov et al. 2013), a neural network method that measures associations between words using the context in which they appear.

Two words are considered neighbors if they are no farther apart than five words in a document (Mikolov et al. 2013), which is also the default setting in the word embedding computer algorithm. For each word, the algorithm uses a neural network to summarize its common neighbors in earnings calls. The

[2] For example, to find the seed words for *integrity*, the authors check all other words clustered with *integrity* by each company and their frequency across companies. They then take words most commonly associated with *integrity*. The word *ethics* is shown to be associated with *integrity* in about 34% of companies and is added on the seed word list for *integrity*.

information about common neighbors is condensed into a vector of fixed dimension (we use 100-dimensional vectors in our study). That is, word embedding converts a word to a 100-dimensional vector that represents the meaning of that word (see Appendix A for more details).[3] We can then quantify the association between two word-vectors $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ using the cosine similarity:

$$\text{cosine}(\boldsymbol{w}_1, \boldsymbol{w}_2) = \frac{\boldsymbol{w}_1 \cdot \boldsymbol{w}_2}{|\boldsymbol{w}_1||\boldsymbol{w}_2|} = \frac{\sum_{i=1}^{100} w_{1,i} w_{2,i}}{\sqrt{\sum_{i=1}^{100} {w_{1,i}}^2}\sqrt{\sum_{i=1}^{100} {w_{2,i}}^2}}, \tag{1}$$

where $w_{1,i}$ is the ith element in vector $\boldsymbol{w}_1$. A high degree of similarity between two vectors indicates the two words are semantically close.

Once we have the cosine similarity between any two words, we construct the culture dictionary by iteratively associating a set of words gleaned from earnings calls to the seed words defining each cultural value.[4] Such a procedure, known as bootstrapping, is common in information retrieval literature for learning new semantic lexicons (Riloff and Jones 1999). As an example, Guiso, Sapienza, and Zingales (2015) list *collaboration* and *cooperation* as seed words for the cultural value of *teamwork*. We can compute the cosine similarity between each unique word in earnings calls with *collaboration* (*cooperation*). For each seed word, we select the top 20 closest synonyms.[5] We expand this set of synonyms once again by finding each synonym's top 20 most similar words.[6] Out of the final set of at most 2×20^2 synonyms, we select the top 500 words that are closest to the average of the vectors of the

---

[3] The neural network can be viewed as performing dimension reduction (e.g., principal component analysis) on the matrix of neighboring word frequencies, see Table A1 in Appendix A. For each term, the original dimensions (columns) in Table A1 are the frequencies of observing neighboring words: *share*, *fruitful*, *joint*…. The frequencies in each row vector capture the semantic meaning of the term. The word embedding model reduces dimensions in Table A1 using a neural network and creates new dimensions that are, roughly speaking, linear combinations of the original dimensions in Table A1. For example, after word embedding, dimension (column) 1 is no longer how often we observe the word *share* near a term (let *n_share* denote the count), but may be 0.3**n_share* + 0.2**n_fruitful* + 0.5**n_joint* + ..., a new composite variable constructed from the frequencies of the original neighboring words.

[4] Although it is possible to generate the dictionary using only the value words (e.g., *innovation*), several cultural values identified by Guiso, Sapienza, and Zingales (2015) encompass meanings beyond their value words. For example, the cultural value of r*espect* includes the meaning of respecting diversity as well as empowering employees. Using the set of seed words as listed in Guiso, Sapienza, and Zingales (2015) allows us to capture the broader meanings of different cultural values.

[5] We also use top 10 and 50 closest synonyms per iteration, and find that those choices do not change our main findings, while using the top 20 closest synonyms leads to the most sensible dictionary.

[6] We stop after two iterations because we find that more than two iterations bring in too many irrelevant words to the dictionary.

two seed words *collaboration* and *cooperation*. Let the vector representations for the seed words be $V_{\{collaboration\}} = [x_1, x_2, \dots, x_d]$, and $V_{\{cooperation\}} = [y_1, y_2, \dots, y_d]$; we select synonyms that have the highest cosine similarity to $[0.5(x_1 + y_1), 0.5(x_2 + y_2), \dots, 0.5(x_d + y_d)]$.

We make a number of adjustments to the basic procedure described above to improve the quality of our dictionary for corporate culture. First, since some idiomatic phrases have meanings that are not captured by simply summing up their individual words, we use the method recommended by Mikolov et al. (2013) to find common phrases in the text and treat them as single words.[7] Second, for our dictionary to be generalizable to all firms and industries, we only include words that are among the top 20% of the most frequently used words in the entire earnings call dataset (and that are sufficiently close to one of the seed words with a cosine similarity > 0.4). Third, we exclude named entities such as persons, organizations, and monetary values using a natural language processing library (see section 3.1 for more details).[8] Fourth, if a word is a synonym for multiple cultural values, we only include it in the dictionary for the value with the highest cosine similarity between the word-vector of that word and the averaged word-vector of the seed words for that value. Finally, to make sure that our dictionary has high specificity, we exclude the most frequent 2,000 words in the entire earnings call dataset. We also manually inspect the dictionary and remove about 15% of the words and phrases (340 out of 2,244) that are industry specific (e.g. *aerodynamic*), too broad (e.g. *managerial*), or typos (e.g. *focuss*).

At the firm-year level, we measure each cultural value by first counting the frequency of our dictionary words defining that cultural value in a call and then normalizing the frequency by the length of the call (i.e., the total word count); there will be five cultural values for each firm at any point in time.

---

[7] For example, "socially responsible" may have a specific meaning that is different from socially + responsible. We treat "socially responsible" as a single word because the frequency with which it appears in calls divided by the multiplication of their component frequencies is greater than 10, as recommended by Mikolov et al. (2013).

[8] In unsupervised machine learning, different goals can lead to different word lists. Our goal is to produce a word list general enough that other researchers can use it to score corporate culture in different contexts. Had our goal been to more effectively score individual companies, keeping named entities such as persons, organizations, and monetary values may have improved the list's power. For example, if a company mentions a particular patent or a particular new product, we know those words are associated with innovation. The same goes for industry-specific technologies. In this paper, we manually filter out words such as radiographic.

2.4. Prior studies employing machine learning

Our approach is related to a growing strand of literature that applies machine learning to financial data.[9] Li (2010) uses naïve Bayes to classify the tone of forward-looking statements and shows its association with firm characteristics. Bao and Datta (2014) use a variant of Latent Dirichlet Allocation (LDA) to find topics in 10-Ks related to risks. LDA is an advanced textual analysis technique that uncovers underlying topics (via clusters of words) in a large set of documents, a powerful unsupervised learning method. Huang, Lehavy, Zang, and Zheng (2018) employ LDA to quantify the content of analyst reports and show that analysts play both information discovery and interpretation roles. Purda and Skillicorn (2015) and Perols et al. (2017) develop machine learning models to detect fraud from financial statements. In contemporaneous work, Erel, Stern, Tan, and Weisbach (2018) develop machine learning algorithms to select directors based on performance. They find that, when compared with a pool of potential candidates, directors predicted to do poorly by their algorithms indeed perform much worse than directors predicted to do well. They conclude that machine learning has the potential to help improve corporate governance practices. Routledge, Sacchetto, and Smith (2018) predict mergers using the MD&A section of 10-Ks and identify key phrases related to M&A decisions and performance.

Compared to these studies, the main strength of our approach is that it combines machine learning with theory testing. We populate a culture dictionary starting from specific dimensions of interest without being distracted by noise in the textual data such as tones or other factors unrelated to culture. Such a theory-guided approach is necessary. As powerful as machine learning is, it can only learn from historical data or rules generated by humans. For example, in Erel et al. (2018), the algorithm learns from the ways in which board members were chosen in the past. In our case, the word embedding model learns the meaning of a word from how it has been used in earnings calls. Machine learning cannot, however,

[9] Textual analysis has been employed by a growing number of finance and accounting papers to examine the tone, sentiment, and readability of corporate 10-K filings, newspaper articles, press releases, and investor message boards (see, for example, Antweiler and Frank 2004; Tetlock 2007; Li 2008; Tetlock, Saar-Tsechansky, and Mackassy 2008; Loughran and McDonald 2011 2014; Jegadeesh and Wu 2013; and a recent survey by Loughran and McDonald 2016).

automatically identify cultural values without human guidance on the meaning of culture. Therefore, we need to feed the algorithm the findings from Guiso, Sapienza, and Zingales (2015) so the algorithm learns what corporate culture is, and the specific cultural values we are seeking.

Combining machine learning with theory testing also allows us to focus on the role of corporate culture in M&As. Many extant machine learning studies focus on predictive analysis, meaning that the algorithms use the input (e.g., a financial document or the tone in the document) to predict the outcome (e.g., firm performance). Although many algorithms deliver powerful predictive performance, subtle but important features in the data such as corporate cultural values may get lost. As a case in point, using 10-Ks to predict deal incidence, Routledge, Sacchetto, and Smith (2018) show that phrases associated with deal incidence also describe recent performance (e.g., 'net loss' or 'material effect'), or financial constraints (e.g., 'credit facility' or 'accounts receivables'). Their predictive model highlights the importance of the Q-theory of mergers (see, for example, Manne 1965; Jovanovic and Rousseau 2002) and financial synergies, but does not allow us to examine whether corporate culture plays any role in M&As. In contrast, we employ unsupervised machine learning with a clear goal of capturing corporate culture and then investigating its role in M&As.

In summary, our approach leverages machine learning to extract interpretable features (i.e., cultural values) from a large volume of textual data. Our dictionary can be applied to other empirical studies involving corporate culture.

3. Data, Measurement, and Validation

3.1. Cleaning, parsing, and matching earnings call data to CRSP/Compustat

With the implementation of Regulation Fair Disclosure (Reg FD) in 2000, firms have been required to make their earnings calls publicly available. Our primary data on earnings calls come from the Thomson Reuters' StreetEvents (SE) database over the period from January 1, 2001 to May 25, 2018 (with limited coverage in 2001 and 2018).

The SE database provides call transcripts in the XML (i.e., extensible markup language) format. We use a Python parser to extract the text and meta-data (description of the data, including company names, tickers, dates, and types of calls) from the XML files. Based on the meta-data of the files, Table IA1 Panel A in the Internet Appendix presents the distribution of the type of calls in the SE database (eventTypeId). For our purpose of scoring corporate culture that is not affected by major corporate events (such as M&As) or industry trends (such as industry conferences), we will employ earnings calls (i.e., "Earning Conference Call/Presentations"). There are 270,879 earnings calls, representing about 70% of all calls.

Before scoring corporate culture, we remove foreign firms based on their exchange suffix codes.[10] There are 187,319 calls made by domestic firms, representing about 70% of all earnings calls.

To train the word2vec model, we prepare the entire collection of earnings calls (i.e., the corpus) using a series of pre-processing steps. First, we use spaCy, a python natural language processing library, to parse the text. The spaCy library helps us lemmatize the words by removing their inflectional endings (e.g., statements → statement). The library also conducts Named-Entity Recognition (NER), which identifies persons by name, locations, and organization names within the text. We do not consider the named entities when constructing the culture dictionary. Second, we convert all the lemmatized words to lower case. Third, we strip the line breaks, punctuation marks, numeric (0-9), stop words (i.e., high-frequency functional words, such as "a," "of," "the," "this," and "it,") and words with fewer than three characters from the corpus. We use the stop words dictionary in Stone, Dennis, and Kwantes (2011). Lemmatization and removing the stop words and short words can reduce the parameter space of the word embedding model. Table A2 in Appendix A lists the seed words from Guiso, Sapienza, and Zingales

[10] The Thomson One Exchange List is available at https://www.usherbrooke.ca/salledemarche/fileadmin/sites/salledemarche/documents/Thomson-One/Guides/Thomson_One_Exchange_List.pdf. We only keep domestic firms without exchange codes or with codes indicating that the firm is traded on the following U.S. exchanges: A for Amex, D for Nasdaq ADF, O for Nasdaq, N for NYSE, and P for NYSE Arca.

(2015) and a portion of the expanded dictionary for each cultural value. Table IA2 in the Internet Appendix provides the full dictionary.

To score corporate culture at the firm-year level, we employ fuzzy matching and manual checking to link calls from SE and Factiva to firms in CRSP/Compustat Merged File with GVKEY. Table IA1 Panel C provides an overview of our sample. Roughly, about 85% firm-years come from the SE database; the remainder are from Factiva. Table IA1 Panel D compares firm-years from the SE database with those from Factiva. We show that Factiva firms are larger and have higher leverage than SE firms. On the other hand, SE firms have better operating and stock performance, faster sales growth, and higher institutional ownership than Factiva firms.

3.2. Measuring corporate culture

An earnings call typically consists of a management presentation section and a QA section. Matsumoto, Pronk, and Roelofsen (2011) find that the QA section is more informative than the presentation section, and its greater information content is positively associated with the number of analysts following. Frankel, Mayew, and Sun (2010) show no difference in tone between the two sections, and Larcker and Zakolyukina (2012) find no differences in identifying deceptive statements from both sections. Given that the presentation section is more likely to be scripted and/or vetted by corporate lawyers and investor relations than the QA section, whereas the latter is more spontaneous and harder to engage in window dressing than the former, our main measures will be based on the QA section.[11]

The raw score for each cultural value is the frequency of dictionary words defining the value normalized by the total number of words in the QA section of a call. We exclude short QA sections with

[11] A call is often led by CEOs and assisted by other top executives. Prior work identifies a number of CEO attributes such as hubris (Roll 1986), or overconfidence (Malmendier and Tate 2005) that might show up in calls. Hollander, Pronk, and Roelofsen (2010) and Cohen, Lou, and Malloy (2017) further note that managers regularly leave participants on calls in the dark by not answering their questions, or "cast" their conference calls by disproportionately calling on bullish analysts. Given that our corporate culture measures are based on a culture dictionary that is not constructed based on personal attributes or positive (negative) words (Loughran and McDonald 2011), we do not expect these features of earnings calls to bias our measures in any significant way. The large number of validation tests further assuages this concern.

fewer than 50 words (about 3.4% of the calls). When a firm makes multiple calls in a year, we sum the word frequencies and the document lengths of all calls before normalizing (i.e., treating these calls together as a single call). Our firm-year measures of cultural values are based on three-year moving averages of annual values. Our final sample consists of cultural values for 8,427 firms and 77,541 firm-year observations.

Table 1 provides an overview of our sample. Panel A presents the summary statistics for corporate cultural value measures and some basic firm characteristics. Consistent with Guiso, Sapienza, and Zingales (2015), *innovation* is the most frequently mentioned cultural value, whereas *teamwork* is the least frequently mentioned cultural value, based on earnings calls.[12]

Panel B presents the autocorrelation of corporate culture value measures. We calculate autocorrelation for firms with more than 15 observations over the sample period. We show that the mean correlation between year *t* and year *t - 1* cultural values ranges from 0.718 for *integrity* and 0.815 for *innovation*, and the mean correlation between year *t* and year *t - 2* cultural values ranges from 0.404 for *respect* and 0.540 for *innovation*. By the fifth lagged correlation, the mean values are close to zero, suggesting that corporate culture evolves slowly over time.

Panel C presents the correlations of corporate culture measures and firm characteristics. We show that among the five cultural values, the correlation between *innovation* and *quality* is the highest, at 0.447, and the correlation between *respect* and *teamwork* is the second highest, at 0.352, while the correlation between *integrity* and *innovation* is the lowest, at 0.074, and the correlation between *quality* and *integrity* is the second lowest, at 0.080. We further show that firm size is positively and significantly associated with *innovation*, and negatively and significantly associated with *quality*, *respect*, and *teamwork*. Leverage is negatively and significantly associated with all five cultural values. Operating performance is positively and significantly associated with *innovation*, and negatively and significantly associated with

[12] One way to benchmark our summary statistics for cultural values is to compare our summary statistics with the summary statistics in Loughran and McDonald (2011, Table 2) regarding the two positive/negative sentiments. Based on textual analysis of 10-Ks, they show that the mean/median for Fin-Neg (negative) is 1.39%/1.36%, and for Fin-Pos (positive) is 0.75%/0.74%.

four other values. Sales growth has little association with any of the five cultural values (in terms of economic significance). Past return has a small negative association with four out of five values (with *innovation* being the exception). Top5 institutional ownership is positively and significantly associated with *innovation*, and negatively and significantly associated with four other values.

Table 2 lists top-ranked S&P 500 firms in different corporate cultural values over three subperiods. We first show that a firm's strong culture can change over time. For example, Alphabet Inc. scores high in *innovation* during the subperiod 2001-2006, and it drops out of being the top firms in *innovation* over the two subperiods 2007-2012 and 2013-2018. Moreover, we show that a firm can excel in multiple cultural values. Over the subperiod 2007-2012, both Salesforce.com Inc. and eBay Inc. score high in *innovation* and *quality*, and Alphabet Inc. scores high in both *quality* and *teamwork*. Finally, we also see some stability in corporate culture, at least for those top ranked firms. For example, Cognizant Tech Solutions and Procter & Gamble Co. score high in *innovation*, Cincinnati Financial Corp. scores high in *integrity*, Aetna Inc. and Apollo Education Group Inc. score high in *respect*, and Citrix Systems Inc. scores high in *teamwork* during the entire sample period.

Figure 1 plots the five cultural values across 12 Fama-French industries over the sample period. We see some interesting patterns. Over time, most industries become more innovative and score higher in *innovation*. The healthcare industry stands out by scoring the highest in *respect* and *teamwork* over the sample period.

In summary, Table 2 and Figure 1 show that corporate culture evolves slowly over time.

3.3. Validating our measures of corporate culture

Given the general concern that commonly advertised values (e.g., on corporate websites) do not capture how values are perceived and upheld by employees (Guiso, Sapienza, and Zingales 2015), it is important to validate our measures using well-established markers for best practices in the corporate world. To that end, we employ a large number of markers for the five cultural values.

To validate the cultural value of *innovation*, we use LnPatent, R&D spending, and innovation strength. LnPatent is the natural logarithm of one plus the number of patents filed and eventually granted in a year. The data are from Kogan, Papanikolaou, Seru, and Stoffman (2017). R&D spending is R&D expenditures normalized by total assets. Innovation strength is an indicator variable that takes the value of one if a firm is considered to have strengths in innovation and R&D, and zero otherwise. KLD defines strength in innovation as "The company is a leader in its industry for research and development (R&D), particularly by bringing notably innovative products to market." The data are from KLD.

To validate the cultural value of *integrity*, we use malfeasance in accounting and backdating executives' option grants (Biggerstaff, Cicero, and Puckett 2015). Restatement is an indicator variable that takes the value of one if a firm later restates its (annual or quarterly) financial statements, and zero otherwise. The data are from Audit Analytics. Backdating is an indicator variable that takes the value of one if option grants to a firm's CEO are backdated, and zero otherwise. To identify backdating, we follow Heron and Lie (2009), whose estimation methodology is based on the assumption that, in the absence of backdating or other types of grant date manipulation, the distributions of stock returns during the month before and after grant dates should be roughly the same.[13] The data on CEOs' option grants are from the Thomson Reuters' Insider Filing database.

To validate the cultural value of *quality*, we use product quality, product safety, and top brand. Product quality is an indicator variable that takes the value of one if a firm is considered to have strengths in product quality, and zero otherwise. KLD defines strengths in product quality as "The company has a long-term, well-developed, company-wide quality program, or it has a quality program recognized as exceptional in U.S. industry." Product safety is an indicator variable that takes the value of one if a firm is not considered to have concerns about product safety, and zero otherwise. KLD defines concerns in product safety as "The company has recently paid substantial fines or civil penalties, or is involved in major recent controversies or regulatory actions, relating to the safety of its products and services." The

[13] We thank Randy Heron for providing us with the SAS codes used in Heron and Lie (2009) to identify option grant backdating.

data for both variables are from KLD. Top brand is an indicator variable that takes the value of one if a firm is included in the top 500 list of Brand Finance rankings, and zero otherwise. The list is constructed by Brand Finance (http://brandirectory.com/) and is available from 2007 to 2017.

To validate the cultural value of *respect*, we use diversity, which is the number of diversity strengths minus the number of diversity concerns. The data are from KLD.

To validate the cultural value of *teamwork*, we use employee involvement and best employer. The former is an indicator variable that takes the value of one if a firm is considered to have strengths in employee involvement, and zero otherwise. KLD defines employee involvement as "The company strongly encourages worker involvement and/or ownership through stock options available to a majority of its employees; gain sharing, stock ownership, sharing of financial information, or participation in management decision making." The data are from KLD. The latter is an indicator variable that takes the value of one if a firm is included in the "100 Best Companies to Work for in America" list, and zero otherwise. Training, employee voice, and work design are the main criteria Fortune uses to create its "100 Best Companies to Work for in America" list (see Edmans (2011) and our Appendix B for details). Edmans (2011) shows that firms on Fortune's list have greater employee satisfaction. The list covers data up to 2016.

Table 3 presents the results of validation tests using cultural values based on the QA section of calls—our main measure. In Panel A, we show that the cultural value of *innovation* is positively and significantly associated with all three measures of corporate innovation activities. This positive association remains after controlling for industry and year fixed effects as well as firm size and operating performance. In Panel B, we show that the cultural value of *integrity* is negatively and significantly associated with accounting malfeasance: restatement. Moreover, we show that the cultural value of integrity is negatively and significantly associated with backdating executives' option grants. In Panel C, we further show that the cultural value of *quality* is positively and significantly associated with two out of three measures of product quality: product safety and top brand. In Panel D, we show that in two out of three specifications, the cultural value of *respect* is positively and significantly associated with the diversity score reported by KLD. Finally, in Panel E, we show that the cultural value of *teamwork* is positively and significantly associated with both employee involvement from KLD and the best employer ranking by Fortune magazine (with one exception).

As a much higher hurdle of validating our measures as well as to illustrate the positive correlations among all five measures (see Table 1 Panel C), we introduce an encompassing specification where we put all five value measures on the right-hand side while the dependent variables are different proxies for each of the five cultural values. Table IA3 in the Internet Appendix presents the results.

Panel A shows that with only one exception (column (4)), there remains a positive and significant association between the cultural value of *innovation* and any of the three measures for corporate innovation activities, after controlling for all four other cultural values. Panel B shows that with only one exception (column (3), albeit with the right sign), there remains a negative and significant association between the cultural value of *integrity* and any of the two measures of unethical behaviour in a company, after controlling for all four other cultural values. Panel C shows that with the exception of product quality (columns (1)-(3)), there remains a positive and significant association between the cultural value of *quality* and any of the two other measures of product quality—product safety and top brand—after controlling for all four other cultural values. It is worth noting that the cultural value of *innovation* is positively and significantly associated with product quality (with one exception, column (3)), which is not surprising given that the correlation between *innovation* and *quality* is the highest among the five values (at 0.447, see Table 1 Panel C). Panel D shows that after controlling for all four other cultural values, the cultural value of *respect* is no longer positively and significantly associated with diversity. Curiously, we find that three cultural values—*innovation*, *integrity* (with one exception, column (2)), and *teamwork* (with one exception, column (1))—are all positively and significantly associated with diversity. Panel E shows that with only one exception (column (4)), there remains a positive and significant association between the cultural value of *teamwork* and any of the two measures of employee engagement and satisfaction, after controlling for all four other cultural values. Moreover, we show that the cultural value of *quality* is positively and significantly associated with employee involvement.

In summary, the validation tests in Table 3 and Table IA3 reassure us that our measures of corporate culture are correlated with shared values and practices by employees at large and have performed as expected.

3.4. Other ways of measuring corporate culture

Our main measures of corporate culture are obtained by applying our culture dictionary to the QA section of calls. Applying our culture dictionary to the full transcript of calls, we generate an alternative set of corporate cultural measures, and label them with a suffix _full.

Given that we are among the first to apply the word embedding model to quantify culture, the question inevitably arises: How is our approach performing compared to a simple alternative using the list of seed words provided by Guiso, Sapienza, and Zingales (2015) and the specific value word (e.g., *innovation*)? We employ a simple count of the seed words (plus the value word) in the QA section of calls to generate a new alternative set of corporate cultural measures, and label them with a suffix _seed.

So far, we employ earnings calls to score corporate culture following the recommendation of Graham et al. (2016). An alternative would be the MD&A section of 10-Ks as employed by Fiordelisi and Ricci (2014). Applying the word embedding model to the MD&A section of 10-Ks over the fiscal year 1993-2017, we generate another alternative set of corporate cultural measures, and label them with a suffix _10k.

Table 4 presents the summary statistics of all alternative measures and their correlations with our main measures based on the QA section of calls. We show that the correlations between the alternative measures based on the full call and our main measures are the highest, ranging from 0.813 for *respect* to 0.871 for *teamwork*. The correlations between the alternative measures based on a simple count of seed words and our main measures are the second highest, ranging from 0.213 for *teamwork* to 0.732 for *quality*. The correlations between the alternative measures based on 10-Ks and our main measures are the lowest.

Tables 5-7 run a horse race between our main measures and each of the three alternative sets of measures using validation tests as in Table 3, except that we have both our main measure and an alternative measure of the same value in the regression specification.[14] In Table 5, we show that with the exception of *respect,* our main measures of corporate culture dominate the alternative measures based on the full call. When we put both measures in the same specification, most of the time, our main measures are significantly associated with the measures they are intended to capture. In Tables 6-7, our main measures always dominate alternative measures based on a simple count of seed words or applying the word embedding model to 10-Ks.

Finally, we also consider two other possibilities to score corporate culture. The first is to apply the word embedding model to employee reviews such as Glassdoor.com. Although employee reviews are a sensible source for learning corporate culture (Grennan 2013; Graham et al. 2016), the data are not publicly available. In addition, data from employee review sites have limited temporal coverage and many firms have very few reviews.[15] The second alternative is to apply topic modeling tools like LDA to earnings calls. However, there is no guarantee that those topics will be related to corporate culture. Huang et al. (2018) find that most topics extracted from earnings calls are either industry-specific or performance-related. We apply LDA to the QA section of earnings calls, and Table IA5 in the Internet Appendix lists different top topics from this exercise. Consistent with Huang et al. (2018), we show that all these topics are not significantly related to corporate cultural values.

In summary, both validation tests and horse races between our main measures of corporate culture and a large number of alternatives suggest that the word embedding model generates a high-quality culture dictionary useful for scoring corporate cultural values.

3.5. Words with multiple senses

---

[14] Table IA4 in the Internet Appendix employs an encompassing specification where we put all four cultural value measures on the right-hand side. With the exception of cultural value *respect*, we show that our main measures based on applying the word embedding model to the QA section outperform all three alternative measures.

[15] We note that the median number of reviews for a public firm in a year is only five based on Glassdoor.com, which limits our ability to obtain firm-year observations of cultural values.

One limitation of the word embedding model is that multiple senses (meanings) of a word are combined into a single vector. Because our corpus is from a very specific domain—earnings call transcripts, the meaning of a word derived from such corpus is less likely to be ambiguous compared to a more general corpus like Wikipedia (Magnini et al. 2002). Nonetheless, we conduct a robustness check by examining the correlation between culture values used in our main analysis and culture values measured using a dictionary in which words with multiple senses are removed.

To do so, we use an algorithm developed by Pelevina et al. (2016) to learn word senses from embedding vectors. The intuition of the algorithm is that for a focal word, we can find its most similar synonyms given its word vector. Then a clustering of the focal word's synonyms is performed based on their semantic similarity.[16] Multiple clusters among the synonyms would indicate that the focal word has multiple senses.

For example, the algorithm finds that the word *measurable*, which is in our culture dictionary for the cultural value of *quality*, has two senses in the transcript corpus. The first sense is captured by a cluster of synonyms including *definite*, *visible*, and *profound*. The second sense is captured by another cluster of synonyms including *quantifiable*, *numeric*, and *root cause*. These two senses correspond to the two definitions of the word *measurable* in the New Oxford American Dictionary. The first sense relates to the definition "large enough to be measured; noticeable; definite," whereas the second sense reflects a more literal definition "able to be measured." As another example, the word *arsenal* has two senses. One is close to *augment*, *offering*, and *comprehensive suite*. Another is close to *mortar*, *torpedo*, and *stealth*, again showing the literal-figurative dichotomy of the senses.

We find that only a small fraction of the words (217 or 11%) in our dictionary has more than one sense. We compute the culture values at the firm-year level using the same method on the QA section but

[16] Specifically, the algorithm performs a graph (network) clustering on a focal word's synonyms. Using its default parameters, the algorithm first produces the top 200 synonyms of a focal word as nodes. Out of a total of ((200 × 199) / 2) pairs of nodes, the algorithm then adds 200 edges between the most similar synonym pairs. The focal word itself is not part of the graph. Finally, the algorithm performs a graph clustering to divide the graph into one or more clusters (representing one or more senses of the focal word). The algorithm discards clusters with fewer than five nodes.

remove these multi-sensed words from the dictionary. The correlation between the measures with and without the multi-sensed words are high, ranging from 0.91 (integrity) to 0.97 (quality) among the dimensions, suggesting that words with multiple senses are of small significance in our setting. Given the high correlations and the specific domain of our corpus, we opt to keep the multi-sensed words in the dictionary.

4. Corporate Culture and M&As

4.1. Hypothesis development

M&As are a setting in which employees of the merging firms with possibly conflicting values and preferences must work together to achieve synergies. If they do not have similar beliefs about the best way of doing things, impediments such as mismatched corporate goals, mistrust, poor morale, and high employee stress and turnover could reduce teamwork and coordination, make post-merger integration difficult, and lower productivity. For example, in firms with an innovation-dominant culture, creating future opportunities in the marketplace through innovation is the ultimate goal, while in firms with a quality-dominant culture, creating value through internal improvements in efficiency and the implementation of better processes and quality enhancements is the long-range goal. Anticipating that the costs of integrating two culturally distant firms will erode or even overwhelm potential synergistic gains, we expect to see fewer deals between firms with conflicting corporate cultures and lower value creation in deals involving culturally distant pairs.[17] In contrast, firms with more congruent corporate cultures are less likely to run into post-merger integration problems and hence their combination is more likely to be well received by the market and is associated with better post-merger long-run performance. The cultural fit hypothesis thus suggests that differences in corporate culture of firm-pairs are a key determinant of deal incidence, acquirer-target firm pairing, and post-merger deal performance.

---

[17] Van den Steen (2010) posit that organization members with homogeneous beliefs and values tend to have weaker incentives to collect information and experiment.

On the other hand, corporate culture may play a limited role in M&As for a number of reasons. First, unlike deeply-held national cultural values, corporate culture is path-dependent and potentially can be shaped by major corporate events (Weber, Shenkar, and Raveh 1996). Nahavandi and Malekzadeh (1988) and Cartwright and Cooper (1993) highlight the process of cultural adaptation and acculturation in M&As whereby post-merger integration leads to some degree of change in merging firms' cultures and practices. Grennan (2013) shows that ownership change can lead to cultural change. Bargeron, Lehn, and Smith (2015) find that acquirers buying large targets are more likely to lose their culture of trust. Guiso, Sapienza, and Zingales (2015) show that the cultural value of integrity among newly public firms changes rapidly over time. Graham et al. (2017) note that about a tenth of executives describe their culture as currently changing. Second, a shorter cultural distance between firm pairs does not necessarily imply cultural congruence, as congruence can also be achieved by complementarity, and not always by achieving similarity; compatible culture does not mean similar culture (Weber, Shenkar, and Raveh 1996; Krishnan, Miller, and Judge 1997). Moreover, merging firms with different cultures might develop a jointly determined culture; there is no such thing as a cultural clash *a priori*. Finally, according to the Q-theory of mergers, well-run firms buy underperforming firms and, by managing them better, achieve gains—the market for corporate control is in essence a contest between management teams competing to run businesses. Based on this neo-classical view of mergers, contracts, economic incentives, and takeovers might have fully resolved any organizational differences, leaving no role for corporate culture in M&As. The acculturation hypothesis thus predicts that merging firms with different cultures will develop a jointly determined culture.

4.2. Measures of cultural fit/conflict

Given that our measures for corporate culture are multi-dimensional (i.e., *innovation*, *integrity*, *quality*, *respect*, and *teamwork*), it is important to develop measures of cultural fit/conflict that capture the richness of our corporate culture measures. To this end, we first introduce two commonly used summary measures of cultural distance.

Cultural similarity is the cosine similarity between two five-by-one vectors capturing the cultural dimensions of a firm-pair. The bigger the value of this summary measure, the closer corporate culture is between a firm-pair. *Cultural distance* is the square root of the sum of squared differences between a firm-pair across all five cultural values (i.e., the Euclidean distance). The smaller the value of this summary measure, the closer corporate culture is between a firm-pair.

To identify the dominant cultural value(s) for each firm-year, we introduce a set of indicator variables: innovation dominant, integrity dominant, quality dominant, respect dominant, and teamwork dominant. A dominant cultural value at a point in time is one where a firm's industry (based on a two-digit SIC code) median-adjusted culture value ranks in the top quartile in the Compustat universe. The ranking is done on an annual basis to take into account the fact that corporate culture might evolve slowly over time.

We also introduce two cultural conflict measures and one cultural fit measure based on the guiding principle behind each cultural value. As discussed earlier, *innovation* is external facing while *quality* is internal facing, and hence these two cultural values are inherently each other's opposite. *Ext-Int conflict* is an indicator variable that takes the value of one when a firm-pair has the opposite dominant culture of *innovation* vis-à-vis *quality*, and zero otherwise. *Ext-Int conflict2* embodies the same idea, except that we expand the internal facing cultures to include *respect*. *Ext-Int conflict2* is an indicator variable that takes the value of one when a firm-pair has the opposite dominant culture of *innovation* vis-à-vis *quality* or *respect*, and zero otherwise. Among the five corporate cultural values, *integrity*, *respect*, and *teamwork* clearly share the commonality of dealing with people within an organization. *People focus* is an indicator variable that takes the value of one when a firm-pair shares the dominant culture of *integrity*, *respect*, or *teamwork*, and zero otherwise.

4.3. Deal incidence and merger pairing

Our sample comprises all U.S. deals completed from January 1, 2003 to December 31, 2017 and reported in the Thomson Reuters' SDC Platinum Database on Mergers and Acquisitions. Table IA6 in the

Internet Appendix presents an overview of the acquirer sample and the pair sample used in the deal incidence and merger pairing analysis, respectively. Table IA7 Panels A and B present the summary statistics for the acquirer sample and the pair sample. Panel C presents the correlations between corporate culture measures and firm and characteristics for the acquirer sample. Panel D presents the correlations between cultural similarity measures and other measures of proximity between an acquirer and its target including same state, HP similarity (Hoberg and Phillips 2010), and same industry (based on two-digit SIC codes).

Table 8 Panel A presents coefficient estimates from a linear probability model (LPM) and a conditional logit model to predict acquirers. Column (1) presents the results from the LPM when the entire Compustat population of firm-years with cultural values is used (without matching); columns (2) and (3) present the results when industry- and size-matched controls are used;[18] columns (4) and (5) present the results when industry-, size-, and B/M-matched controls are used.

Across different specifications, we find that firms scoring high on the cultural value of *innovation* are more likely to be acquirers, whereas firms scoring high on the cultural values of *integrity*, *quality*, *respect* are less likely to be acquirers, supporting our conjecture that firms with an externally-oriented culture like innovation are more likely to do deals. In terms of economic significance, using the specification in column (1), we find that when the cultural value of *innovation* increases by one standard deviation, the likelihood of a firm becoming an acquirer increases by 2.01%, whereas when the cultural value of *integrity/quality/respect* increases by one standard deviation, the likelihood of a firm becoming an acquirer decreases by 0.81%/2.82%/0.77%. In contrast, when the value of leverage/past return

[18] First, to form the industry- and size-matched control firms, for each acquirer of a deal announced in year t, we find up to five matching acquirers by industry—where the industry definitions are based on the narrowest SIC grouping that includes at least five firms—and by size from Compustat/CRSP in year $t - 1$ for firms that were neither an acquirer nor a target firm in the three-year period prior to the deal. We further require that control firms' size be within [0.5, 1.5] times that of the event firm. In the end, 52% (17%) acquirers are matched at the four-digit (three-digit) SIC industry level, 56% (16%) target firms are matched at the four-digit (three-digit) SIC industry level, and the remainder are at the two-digit SIC industry level.

increases by one standard deviation, the likelihood of a firm becoming an acquirer decreases/increases by 2.91%/2.72%. The effect of cultural values is clearly economically significant.

Other findings not directly related to corporate culture are nonetheless consistent with prior work in M&As (see, for example, Moeller, Schlingemann, and Stulz 2004; Gaspar, Massa, and Matos 2005; Bena and Li 2014). In particular, we show that larger firms, and firms with faster sales growth, stronger prior year returns, and higher institutional ownership are more likely, whereas firms with high leverage are less likely, to be acquirers.

Table 8 Panel B presents coefficient estimates from a conditional logit model to predict merger pairs.[19] Columns (1) and (2) present the results when industry- and size-matched controls are used, and columns (3) and (4) present the results when industry-, size-, and B/M-matched controls are used. Columns (1) and (3) present the regression results using cultural similarity to measure cultural congruency, and columns (2) and (4) present the regression results using cultural distance to measure cultural congruency.

We find that firms closer in cultural values are more likely to do a deal together, whereas firms farther apart in cultural values are less likely to do so, which supports the cultural fit hypothesis. We further find that firms headquartered in the same state or sharing similar product descriptions in 10-K filings (*HP similarity* as defined in Hoberg and Phillips 2010) are more likely to do deals together. In terms of economic significance, using the specifications in column (1)/(2), we find that when the measure of *cultural similarity/distance* increases by one standard deviation, the likelihood of a firm-pair becoming an acquirer-target increases/decreases by 2.49%/3.16%. In contrast, when the two firms have their headquarters in the same state instead of different states, the likelihood of a firm-pair becoming an acquirer-target increases by 10.17%; and when the measure of product description similarity increases by

[19] Results using the LPM are largely similar to those reported using the conditional logit model.

one standard deviation, the likelihood of a firm-pair becoming an acquirer-target increases by 13.11%.[20] The effect of cultural similarity is clearly economically significant.

As discussed earlier, cultural similarity (distance) is a summary measure of cultural congruence between two firms, but does not offer insight into how one firm's culture is congruent to that of another. Table IA8 presents the regression results using five interaction terms capturing the congruency in dominant cultures between acquirers and targets. We find that firm-pairs sharing the dominant cultural values of *integrity, quality*, or *teamwork* are more likely to become acquirer-target pairs.

Overall, Tables 8 and IA8 provide strong evidence in support of the cultural fit hypothesis that firms sharing similar corporate culture are more likely to do deals together.

4.4. Cultural fit and deal outcome

Under the cultural fit hypothesis, we would expect that better deal outcomes would be achieved among firms with more congruent corporate culture. Our short-run performance measure is the three-day combined stock price reaction of acquirers and targets (*Combined CAR(-1, 1)*). Our long-run performance measure is the acquirer one-year buy-and-hold abnormal return (*BHAR1)* and the change in ROA from the year of deal announcement to the year after deal completion (*ΔROA1*) following Chen, Harford, and Li (2007) and Li, Qiu, and Shen (2018). Our measures of ex post integration/retention problems are based on a textual analysis of 10-Ks, in particular in the MD&A and/or risk factors sections, following Hoberg and Phillip (2017). The measure, *integration* (*retention*), is an indicator variable that takes the value of one if an acquirer reports integration (retention) problems within one year after deal completion, and zero otherwise. Detailed variable definitions are provided in Appendix B.

Table IA9 presents the summary statistics of the pair sample used for the ex post deal outcome analysis. The sample consists of 810 deals for which we have available data to compute the price reaction

[20] The conditional logit model does not allow us to calculate the marginal effects. For deal probability, we estimate an equivalent (unconditional) logit model with deal fixed effects and compute the economic magnitude using the average marginal effect of the independent variable multiplied by the standard deviation of the variable (if continuous) or by one (if binary).

for both the acquirer and its target firm. For the deal outcome analysis, we further require that acquirers do not engage in any other significant deals for one year after the focal deal's completion.

Table 9 presents the results on the relationship between the acquirer-target cultural fit, and the short- and long-run deal performance and post-merger integration/retention problems. In Panel A where the dependent variable is *Combined CAR(-1, 1),* we show that firm-pairs with more sizeable cultural distance or opposite dominant culture of *innovation* vis-à-vis *quality* (or *respect*) are associated with worse short-run deal performance. In Panel B where the dependent variable is *BHAR1*, we fail to find any significant association between measures of cultural distance and acquirers' one-year buy-and-hold abnormal returns. In Panel C where the dependent variable is *ΔROA1*, we find that firm-pairs sharing the dominant people-oriented culture of *teamwork* (or *integrity*, or *respect*) are associated with a bigger increase in ROA one year after deal completion. In Panel D where the dependent variable is post-merger integration, we show that firm-pairs with opposite dominant culture of *innovation* vis-à-vis *quality* (or *respect*) are associated with significantly more integration challenges compared to their counterparts that are not in opposing cultures. In contrast, firm-pairs sharing the dominant people-oriented culture of *teamwork* (or *integrity*, or *respect*) are less likely to experience integration challenges. We find no significant association between measures of cultural fit/conflict and post-merger retention problems (Panel E).

Overall, Tables 9 provides somewhat limited evidence in support of the cultural fit hypothesis that cultural congruence is associated with better deal outcomes. This evidence might not be surprising given that we show in Table 8 that culturally misaligned firm pairs do not initiate a deal in the first place.

4.5. Post-merger acculturation

In the field of anthropology and cross-cultural psychology, acculturation is generally defined as "changes induced in (two cultural) systems as a result of the diffusion of cultural elements in both directions" (Berry 1980, p. 215). We conjecture that a successful merger will also involve members of the acquirer and the target firm to adapt to each other and resolve emergent conflicts; thus, the merger itself could also shape corporate culture.

Table 10 provides suggestive evidence. The sample consists of 605 (415) deals one year (three years) after deal completion. We further require that acquirers do not engage in any other significant deals for one year (three years) after the focal deal's completion. We show that within either the one-year or three-year period after deal completion, the acquirer's cultural values are significantly related to both the acquirer's and the target's values pre-merger (with the exception of *innovation*), suggesting that mergers might help acquirers to create a new jointly-determined culture, consistent with the acculturation hypothesis.

5. Conclusions

Using one of the latest machine learning techniques—the word embedding model (Mikolov et al. 2013)—and 217,387 earnings call transcripts, we obtain five corporate cultural values—*innovation*, *integrity*, *quality*, *respect*, and *teamwork* (Guiso, Sapienza, and Zingales 2015) for 77,541 firm-year observations over the period 2001–2018. We conduct a large number of tests to validate our measures and demonstrate the advantages of our approach over several alternative methods. We show that corporate culture plays an important role in deal incidence and merger pairing, and that post-merger, acquirers' cultural values are positively associated with their target firms' cultural values pre-merger. We conclude that machine learning is useful for scoring corporate culture and holds promise for more applications in the field of finance.

Appendix A
Introduction to word embedding

The word embedding model is among the hallmarks of the recent surge of deep learning (LeCun, Bengio, and Hinton 2015). The goal is to represent each word as a vector of about 20-500 dimensions, based on the textual context in which the word is found. The vectorized representation of words allows us to compute the similarity between words and cluster them based on the underlying concept. In our application of measuring corporate culture, we employ word embedding to learn how words in earnings calls relate to each other. Given the learned relationship, we identify a broad set of words and phrases that describes cultural values and can be used to score firms accordingly.

Our approach is based on a simple, time-tested concept in linguistics: Words tend to co-occur with neighboring words with similar meanings (Harris 1954). The following is a simple example to illustrate the gist of the approach. Suppose we want to examine the relationship between three words: *collective, partnership*, and *governance*. We start by counting how many times any neighboring word appears near these three words in a collection of documents. The following table summarizes the counts.

Table A1. Terms and neighboring word counts

	Neighboring Word Counts				
Terms	*share*	*fruitful*	*joint*	*oversight*	*proper*
collective	4	5	5	0	1
partnership	3	6	7	0	0
governance	0	0	1	10	9

We see that *share, fruitful,* and *joint* tend to appear most often near *collective* and *partnership*, and *oversight* and *proper* tend to appear most often near *governance*. We can use a vector [4, 5, 5, 0, 1] to represent *collective*, a vector [3, 6, 7, 0, 0] to represent *partnership*, and a vector [0, 0, 1, 10, 9] to represent *governance*. Given any two vectors, we can use their cosine similarity to measure their association. The cosine similarity between *collective* and *partnership* is

$$cos(\boldsymbol{x}, \boldsymbol{y}) = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}} = \frac{4 \times 3 + 5 \times 6 + \cdots + 1 \times 0}{\sqrt{4^2 + 5^2 + \cdots + 1^2}\sqrt{3^2 + 6^2 + \cdots + 0^2}} = 0.97$$

Similarly, the cosine similarity between *collective* and *governance* is 0.13. We conclude that *collective* and *partnership* are semantically closer to each other than *collective* and *governance*.

Such crude vector representation based on co-occurring neighboring words has clear drawbacks. For one, the vectors are only five components long because we list only five neighboring words. In reality, the dimension of the vectors can be hundreds of thousands if we include all neighboring words. For another, if our goal is to discover new words that are related to cultural values, the above count-based method cannot scale up—we would need to maintain a table that has $|V|$ rows and $|V|$ columns, where $|V|$ is the number of unique words in the vocabulary.

Word embedding solves the problem by using a large neural network model to learn high-quality vector representations that are manageable (a fixed dimension *d* that is usually between 20-500), but preserve as many properties of the original co-occurrence relationship as possible.[21] Levy and Goldberg (2014) prove that the word embedding algorithm is a transformation of the singular value decomposition (a dimension reduction technique) of Table A1. The algorithm first initializes each word's vector randomly. The vector representations are parameters in a neural network that can be trained using data. The neural network reads through the collection of documents several times and trains itself to perform a prediction task: Given any word in the document, predict its neighboring words. The training is complete when the parameters fit the data well. As it happens, the trained parameters, i.e., each word's vector representation, can be used to summarize the word's semantic information in the document. That is, we can use a vector to represent a word's meaning.

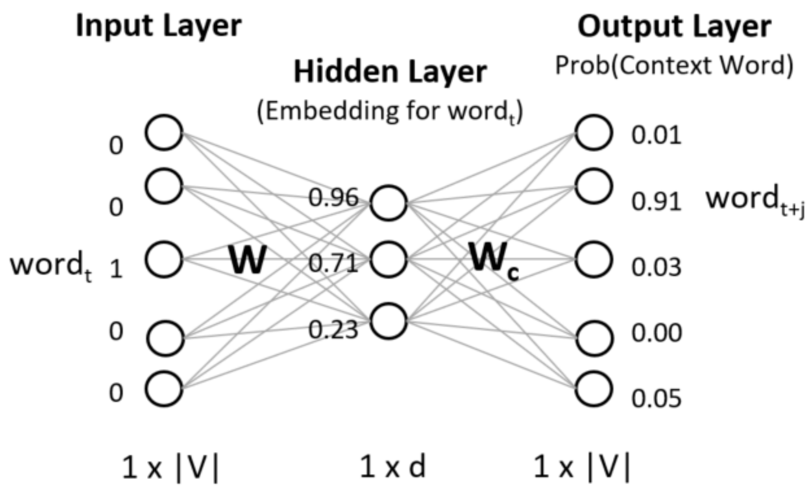


Figure A1. Illustration of the neural network for word embedding

We now describe the neural network in detail. We adopt the skip-gram model (Mikolov et al. 2013) to calculate the word embedding vectors. Figure A1 provides an illustration of the model. The skip-gram model is a feed-forward neural network—given an input word, the neural network outputs neighboring words. Predicting each word's surrounding words is equivalent to maximizing the log probability:

$$\frac{1}{|V|}\sum_{t=1}^{|V|}\sum_{-k\leq j\leq k, j\neq 0}\log p(w_{t+j}\,|w_t),$$

where k is the "window size" of the context (5 words in our case), w_t is a word at location t, and $|V|$ is the size of the vocabulary. Note that each word can be naturally represented using a $|V|$ dimensional one-hot row vector.[22] A single-hidden-layer neural network, parameterized by a $|V| \times d$ weight matrix W, first projects[23] a word w to a vector v_w in $\mathbb{R}^d$, where v_w is simply the corresponding row in W. The network's output softmax layer, parameterized by a second $d \times |V|$ weight matrix W_c, uses the v_w as the

[21] Occurrence refers to the similarity of two words, say *collective* and *partnership*, based on how often we see the same neighboring words around them. In the simple example, we often see the same set of words {e.g., *share*, *fruitful*, *joint*, etc. accompanying both *collective* and *partnership*.

[22] A one-hot vector is a vector with a single 1 and the others 0. Since there are $|V|$ unique words, each word can be represented using a one-hot vector with a unique entry being 1. For example, a is [1, 0, 0, 0, …], ability is [0, 1, 0, 0, 0, …], able is [0, 0, 1, 0, …] , zoo is [0, 0, 0, …,0, 1].

[23] A one-hot row vector with the *w*th entry being 1 multiplying *W* outputs the *w*th row of *W*.

input to predict the probability of observing each context word *c* in the context of *w*. The corresponding column in W_c is denoted as v_c. That is:

$$p(c|w) = \frac{\exp(v_c^{\mathrm{T}} v_w)}{\sum_{c' \in C} \exp(v_{c'}^{\mathrm{T}} v_w)}.$$

Putting it together, the log-likelihood of the entire model is computed by summing over all (w, c) combinations:

$$\arg\max_{W, W'} \prod_{w \in V} \prod_{c \in c(w)} p(c|w;\ W,\ W_c)$$

where $c \in c(w)$ is the set of all context words for word w.

The above neural network can be viewed as two layers of regressions concatenated, with the first layer of the regressions' output becoming another layer's input. The first layer contains *d* linear regressions, each taking the same input, the one-hot vector (|V| dimensional) of a word [0, 0, 0, 1, …, 0], and outputs a single number. Together the output of the first layer is the vector $[x_1, x_2, \ldots, x_d]$. The output of the first layer is then used as the input of a multinomial logistic regression, and the final output is the probabilities of neighboring words.

The learning of word vectors v_w's is achieved when the log-likelihood is maximized, i.e., the neural network is trained using data (i.e., a collection of documents). For such a feed-forward neural network, the W and W_c can be initialized randomly. As the neural network passes through the text word by word, it keeps predicting the surrounding words given the current word. The neural network will make mistakes, and a back-propagation algorithm or other approximation training algorithm can adjust W and W_c by learning from those mistakes. After 5-10 passes of the entire text collection, the neural network becomes adept at the task. The training is now complete, and the v_w's (rows of W) are our final d-dimension vector representation of each word.

We use an open source Python package Gensim to train the word embedding model.[24] Other deep learning packages such as TensorFlow and Keras can also be used for training.

References:

LeCun, Y., Y. Bengio, and G. Hinton, 2015. Deep learning, *Nature* 521, 436-444.

Levy, O., and Y. Goldberg, 2014. Neural word embedding as implicit matrix factorization, *Advances in Neural Information Processing Systems,* 2177-2185.

Mikolov, T., I. Sutskever, K. Chen, G.S. Corrado, and J. Dean, 2013. Distributed representations of words and phrases and their compositionality, *Advances in Neural Information Processing Systems,* 3111-3119.

[24] The word embedding model is considered a mature technique in computational linguistic (like LDA), and hence no further fine tuning is called for.

Appendix B
Matching SE data to Compustat/CRSP

In the vast majority of the calls, we can easily extract the company name and fiscal year or quarter from the event title toward the end of a call (see, for example, "Q4 2012 Venoco, Inc. Earnings Conference Call"). For the remainder (less than 2% of the calls), we use various heuristics rules to infer the company name and fiscal year.[25] We note that once a target firm has been acquired, Thomson Reuters backfills company names and tickers in event titles to reflect the name of the acquirer, which inevitably complicates identifying target firms' calls. Table IA1 Panel B provides an example of backfilling: The company name given via backfilling is EnergySolutions Inc. (the acquirer's name), rather than the target firm's pre-merger name, Duratek Inc.

To maximize matching between SE's company name to GVKEY or PERMNO, we employ a multipronged approach. We use SAS' SPEDIS function to match the first 25 characters of a company's name extracted from the body of a call to a company's name from CSRP with a distance score less than or equal to 50. We repeat the same process using company names from Compustat. We also use SAS' COMPGED function to match a company's full name extracted from the body of a call to a company's name from CRSP/Compustat Merged File with a distance score less than or equal to 200.[26] A perfectly matched pair would be the case with exactly the same company name from SE and CRSP in order to get PERMNO (or Compustat to get GVKEY, or CRSP/Compustat Merged File to get GVKEY), and the distance score from SAS would be zero.

For less than perfect matching cases based on company names (i.e., the distance score is greater than zero), if the company's name in the event title (subject to backfilling) is the same as the name extracted from the body of the same call (without backfilling), i.e., the company name and ticker symbol are accurate (not subject to backfilling), we use both the ticker and fiscal year to match with CRSP in order to get PERMNO.

For the remaining less than perfect matching cases after fuzzy name matching and ticker matching, we ask our research assistants to search CRSP and Compustat's Code Lookup functions to ensure correct matching, paying particular attention to variations in names and abbreviations (e.g., HLDGS, INTL).[27] We also use PERMNO (from CRSP) to match with GVKEY (from Compustat and CRSP/Compustat Merged File).

In the end, after fuzzy matching and manual checking, we are able to match about 80% of the calls from SE to firms in CRSP/Compustat Merged File with GVKEY.

Given that the company name and ticker are subject to backfilling and that we are able to match about 80% of calls in the SE database, to further increase our matching rate we resort to another popular source for earnings calls – Factiva, over the period from January 1, 2002 to December 31, 2017. We apply both fuzzy name matching and manual checking, and end up with a matching rate of about 70% of all calls. We then merge the calls from SE and those from Factiva based on GVKEY and fiscal year.

---

[25] For example, we use regular expressions to extract years 2012/2005 and company names from the event titles "AT &T's 4Q12 Earnings Conference Call" and "PMC-Sierra Third Quarter 2005 Conference Call".

[26] The reason we use company names provided by CRSP, Compustat, and CRSP/Compustat Merged File for matching is because occasional small variations in company names occur across these three databases. Relying on multiple sources helps us capture as many matches as possible before manual checking.

[27] Both CRSP and Compustat have Code Lookup functions, whereas CRSP/Compustat Merged File does not. Therefore, we use only the Code Lookup functions of CRSP and Compustat in manual checking.

Table A2
Cultural values, seed words, and dictionary

This table presents a subset of our culture dictionary.[28] The full dictionary is provided in Table IA1 of the Internet Appendix.

Cultural values	Seed words	The dictionary (a subset)
Innovation	creativity, excellence, improvement, passion, pride, leadership, growth, performance, efficiency, results	execution, innovation, excellence, efficiency, productivity, creativity, fanatical, focus, effectiveness, differentiation, agility, relentlessly, technological_advancement, passion, speed_agility, powerfully, responsiveness, relentless, innovate, competitiveness, virtuous_cycle, improvement, tenacity, adaptability…
Integrity	ethics, accountability, honesty, fairness, responsibility, transparency	accountability, transparency, ethic, oversight, transparency_accountability, integrity, responsibility_accountability, governance, rigor, utmost, responsibility, continuity, zero_tolerance, seriousness, credibility, sense_urgency, consistency, alignment, moral, humility, assure, autonomy, accountable, thoroughness, hold_accountable, ethical…
Quality	customer, commitment, dedication, value, expectations	quality, value, commitment, customer, reputation, loyal, choice, satisfaction, client, committed, commitment, tireless, loyalty, relationship, perseverance, dedication, longevity, stickiness, proposition, willingness, solve_problem, reliable, capable, affordability, dedication_commitment…
Respect	diversity, inclusion, development, talent, employees, dignity, empowerment	talent, skill, employee, talented, talented_workforce, empowerment, highly_skilled, highly_motivate, talented_dedicated, employer_choice, empowered, talent_pool, competency, culturally, incredibly_talented, alumnus, attract_talent , expertise, skillset, student_faculty…
Teamwork	collaboration, cooperation	collaborative, cooperative, collaborate, cooperation, teamwork, collegial, working, collaboratively, coordinate, coordination, cooperatively, collaborating, coordinated_effort, collaboration, teaming, unite, collective, engage, involvement, partner, comarketing, joint, cordial, coordinated, symbiotic, jointly, consultative, engaged, professionalism, supportive, involved, multi_disciplinary…

[28] It is worth noting that our seed word list for integrity does not include "trust" and "ownership". When trust and ownership are used in mission statements as in Guiso, Sapienza, and Zingales (2015), they may have culture related meanings. But when they are used in earnings call transcripts, they often mean "trust company" or equity ownership, which does not have much to do with corporate culture. To avoid the biases, we have removed those two words from the seed word list for integrity.

Appendix B
Variable definitions

All continuous variables are winsorized at the 1st and 99th percentiles. All dollar values are in 2016 dollars.

Variable	Definition
Culture Variables	
Innovation	Percentage of innovation-related words in the QA section of earnings calls averaged over a three-year window.
Integrity	Percentage of integrity-related words in the QA section of earnings calls averaged over a three-year window.
Quality	Percentage of quality-related words in the QA section of earnings calls averaged over a three-year window.
Respect	Percentage of respect-related words in the QA section of earnings calls averaged over a three-year window.
Teamwork	Percentage of teamwork-related words in the QA section of earnings calls averaged over a three-year window.
Innovation_full	Percentage of innovation-related words in earnings calls averaged over a three-year window.
Integrity_full	Percentage of integrity-related words in earnings calls averaged over a three-year window.
Quality_full	Percentage of quality-related words in earnings calls averaged over a three-year window.
Respect_full	Percentage of respect-related words in earnings calls averaged over a three-year window.
Teamwork_full	Percentage of teamwork-related words in earnings calls averaged over a three-year window.
Innovation_seed	Percentage of innovation-related seed words and the value word innovation based on a simple count in the QA section of earnings calls averaged over a three-year window.
Integrity_seed	Percentage of integrity-related seed words and the value word integrity based on a simple count in the QA section of earnings calls averaged over a three-year window.
Quality_seed	Percentage of quality-related seed words and the value word quality based on a simple count in the QA section of earnings calls averaged over a three-year window.
Respect_seed	Percentage of respect-related seed words and the value word respect based on a simple count in the QA section of earnings calls averaged over a three-year window.
Teamwork_seed	Percentage of teamwork-related seed words and the value word teamwork based on a simple count in the QA section of earnings calls averaged over a three-year window.
Innovation_10k	Percentage of innovation-related words in the MD&A section of 10-K averaged over a three-year window.
Integrity_10k	Percentage of integrity-related words in the MD&A section of 10-K averaged over a three-year window.
Quality_10k	Percentage of quality-related words in the MD&A section of 10-K averaged over a three-year window.
Respect_10k	Percentage of respect-related words in the MD&A section of 10-K averaged over a three-year window.
Teamwork_10k	Percentage of teamwork-related words in the MD&A section of 10-K averaged over a three-year window.
Cultural similarity	Cosine similarity between firm *a* and firm *b*'s culture vectors [$innovation_a$, $integrity_a$, $quality_a$, $respect_a$, $teamwork_a$] and [$innovation_b$, $integrity_b$, $quality_b$, $respect_b$, $teamwork_b$]. A higher value indicates similar cultures.

Cultural distance	Euclidean distance between firm *a* and firm *b*'s culture vectors [$innovation_a$, $integrity_a$, $quality_a$, $respect_a$, $teamwork_a$] and [$innovation_b$, $integrity_b$, $quality_b$, $respect_b$, $teamwork_b$]. The culture scores are first standardized by subtracting the mean and dividing by the standard deviation of each year. A lower value indicates similar cultures.
Innovation-dominant	An indicator variable that takes the value of one if a firm's industry (based on two-digit SIC code) median-adjusted culture score of innovation ranks in the top quartile in the Compustat universe in a year, and zero otherwise.
Integrity-dominant	An indicator variable that takes the value of one if a firm's industry (based on two-digit SIC code) median-adjusted culture score of integrity ranks in the top quartile in the Compustat universe in a year, and zero otherwise.
Quality-dominant	An indicator variable that takes the value of one if a firm's industry (based on two-digit SIC code) median-adjusted culture score of quality ranks in the top quartile in the Compustat universe in a year, and zero otherwise.
Respect-dominant	An indicator variable that takes the value of one if a firm's industry (based on two-digit SIC code) median-adjusted culture score of respect ranks in the top quartile in the Compustat universe in a year, and zero otherwise.
Teamwork-dominant	An indicator variable that takes the value of one if a firm's industry (based on two-digit SIC code) median-adjusted culture score of teamwork ranks in the top quartile in the Compustat universe in a year, and zero otherwise.
Ext-Int conflict	An indicator variable that takes the value of one if an acquirer's dominant cultural value is innovation (quality) and its target's dominant cultural value is quality (innovation), and zero otherwise.
Ext-Int conflict2	An indicator variable that takes the value of one if an acquirer's dominant cultural value is innovation (either quality or respect) and its target's sole dominant cultural value is either quality or respect (innovation), and zero otherwise.
People focus	An indicator variable that takes the value of one if an acquirer's dominant cultural value is integrity, respect, or teamwork, and its target's sole dominant cultural value is integrity, respect, or teamwork, and zero otherwise.
Validation Variables	
Backdating	An indicator variable that takes the value of one if option grants to a firm's CEO are backdated, and zero otherwise. To identify backdating, we follow Heron and Lie (2009) whose estimation methodology rests on the assumption that, in the absence of backdating or other types of grant date manipulation, the distributions of stock returns during the month before and after grant dates should be roughly the same, implying that the distribution of return differences should be centered on zero. The data on option grants to CEOs are from the Thomson Reuters' Insider Filing database.
Best employer	An indicator variable that takes the value of one if a firm is included in "100 Best Companies to Work for in America" list, and zero otherwise. Fortune compiles the ranking based on the following methodology (Edmans 2011). Two-thirds of the score comes from employee responses to a 57-question survey created by the Great Place to Work Institute in San Francisco, which covers topics such as attitudes toward management, job satisfaction, fairness, and camaraderie. The remaining one-third of the score comes from the Institute's evaluation of factors such as a firm's demographic makeup, pay and benefits programs, and culture. The final score covers four areas: credibility (communication to employees), respect (opportunities and benefits), fairness (compensation, diversity), and pride/camaraderie (teamwork, philanthropy, celebrations). The list is available until 2016.
Diversity	The number of diversity strengths minus the number of diversity concerns as reported by KLD database. The data are available until 2016.

Employee involvement	An indicator variable that takes the value of one if a firm is considered to have strengths in employee involvement, and zero otherwise. KLD defines employee involvement as "The company strongly encourages worker involvement and/or ownership through stock options available to a majority of its employees; gain sharing, stock ownership, sharing of financial information, or participation in management decision making." The data are available until 2016.
Innovation strength	An indicator variable that takes the value of one if a firm is considered to have strengths in innovation and R&D, and zero otherwise. KLD defines strength in innovation as "The company is a leader in its industry for research and development (R&D), particularly by bringing notably innovative products to market." The data are available until 2016.
LnPatent	Natural logarithm of one plus the number of patents filed and eventually granted in a year. The data are from Kogan, Papanikolaou, Seru, and Stoffman (2017) and available at https://iu.app.box.com/v/patents until 2010.
Product quality	An indicator variable that takes the value of one if a firm is considered to have strengths in product quality, and zero otherwise. KLD defines strength in product quality as "The company has a long-term, well-developed, company-wide quality program, or it has a quality program recognized as exceptional in U.S. industry." The data are available until 2016.
Product safety	An indicator variable that takes the value of one if a firm is not considered to have concerns in product safety, and zero otherwise. KLD defines concerns in product safety as "The company has recently paid substantial fines or civil penalties, or is involved in major recent controversies or regulatory actions, relating to the safety of its products and services." The data are available until 2016.
R&D spending	R&D expenses scaled by total assets.
Restatement	An indicator variable that takes the value of one if a firm later restates its (annual or quarterly) financial statements, and zero otherwise. The data are from Audit Analytics.
Top brand	An indicator variable that takes the value of one if a firm is included in the top 500 list of Brand Finance rankings, and zero otherwise. The list is constructed by Brand Finance (http://brandirectory.com/) and is available from 2007 to 2017.
Outcome Variables	
Combined CAR(-1, 1)	Weighted average cumulative abnormal return (in percentage points) of the acquirer and the target from one day before to one day after the deal announcement date. Abnormal return is calculated by subtracting the CRSP value-weighted market return from the weighted average stock return of the acquirer and the target.
BHAR1	One-year buy-and-hold abnormal stock return of the acquirer after deal completion constructed following Chen, Harford and Li (2007) and Li, Qiu, and Shen (2018). Specifically, we first sort the NYSE/NASDAQ/AMEX firms each month into NYSE size deciles, and then further partition the bottom decile into quintiles, producing 14 total size groups. We simultaneously sort firms into book-to-market (B/M) deciles. After determining which of the 140 (14 size × 10 B/M) groups the acquirer is in at the month-end prior to deal completion, we choose from that group the control firm that is the closest match on prior year stock return and is not involved in any significant acquisition activity in the prior year (three years). One-year (three-year) buy-and-hold return (starting from the month after deal completion) is then calculated for the acquirer and the control firm. Finally, the one-year (three-year) buy-and-hold abnormal return is the difference between the acquirer return and the corresponding contemporaneous control firm return. To compute the variable, the acquirer cannot complete any confounding deal with a transaction value greater than 1% of the acquirer's total assets within the year after deal completion.

ΔROA	The difference (in percentage points) between an acquirer's ROA in the year after deal completion and ROA in the year when the deal is announced. To compute the variable, the acquirer cannot complete any confounding deal with a transaction value greater than 1% of the acquirer's total assets within the year after deal completion.
Integration	An indicator variable that takes the value of one if an acquirer makes statements about integration challenges in its 10-K filing one-year after deal completion, and zero otherwise. Specifically, merger-related keyword list is merger, mergers, merged, acquisition, acquisitions, and acquired. Integration-related keyword list 1 includes integration, integrate, integrating, and other synonyms. Integration-related keyword list 2 includes challenge, challenging, difficulties, difficulty, inability, failure, unsuccessful, and other synonyms. We require at least one word from the merger list and from both integration lists showing up in the same paragraph for the integration indicator variable to take the value of one.
Retention	An indicator variable that takes the value of one if an acquirer makes statements about employee retention issues in its 10-K filing one-year after deal completion, and zero otherwise. Specifically, merger-related keyword list is merger, mergers, merged, acquisition, acquisitions, and acquired. Retention-related keyword list 1 contains employee, employees, personnel and other synonyms. Retention-related keyword list 2 contains departure, departures, retention and other synonyms. We require at least one word from the merger list and from both retention lists showing up in the same paragraph for the retention indicator variable to take the value of one.
Firm Characteristics	
Firm size	Natural logarithm of total assets.
Book-to-market (B/M)	Book value of equity divided by market value of equity.
Leverage	Book value of debt divided by the sum of book value of debt and market value of equity.
ROA	Income before extraordinary items scaled by total assets.
Sales growth	The growth rate of sales, calculated as (sales in year *t* – sales in year *t - 1*)/ sales in year *t - 1*.
Past return	Buy-and-hold stock return in the year prior to deal announcement.
Top5 institutions	The fraction of shares outstanding held by the five largest institutional investors prior to deal announcement.
Deal Characteristics	
All cash	An indicator variable that takes the value of one if a bid involves only cash payment to the target shareholders, and zero otherwise.
All stock	An indicator variable that takes the value of one if a bid involves only stock swap with the target shareholders, and zero otherwise.
Same industry	An indicator variable that takes the value of one if an acquirer is from the same two-digit SIC industry as its target firm, and zero otherwise.
Tender offer	An indicator variable that takes the value of one if a bid is a tender offer made to target shareholders, and zero otherwise.
Relative size	The ratio of deal transaction value to an acquirer's total assets.
Same state	An indicator variable that takes the value of one if an acquirer's and its target's headquarters are in the same state, and zero otherwise.
HP similarity	Acquirer-target pairwise similarity scores based Hoberg and Phillips (2016).

References:

Antweiler, W., and M. Z. Frank, 2004. Is all that talk just noise? The information content of Internet stock message boards, *Journal of Finance* 59, 1259–1293.

Bargeron, L., J. Smith, and K. Lehn, 2015. Employee-management trust and M&A activity, *Journal of Corporate Finance* 35, 389–406.

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

Bena, J., and K. Li, 2014. Corporate innovations and mergers and acquisitions, *Journal of Finance* 69, 1923–1960.

Bereskin, F., S. K. Byun, M. S. Officer, and J.-M. Oh, 2017. The effect of cultural similarity on mergers and acquisitions: Evidence from corporate social responsibility, *Journal of Financial and Quantitative Analysis* forthcoming.

Berry, J. W., 1980. Social and cultural change, in: Triandis, H. C., and Brislin, R. W. (eds.), *Handbook of Cross-Cultural Psychology*, Vol. 5, Boston: Allyn & Bacon, 211–279.

Biggerstaff, L., D. C. Cicero, and A. Puckett, 2015. Suspect CEOs, unethical culture, and corporate misbehavior, *Journal of Financial Economics* 117, 98–121.

Bodnaruk, A., T. Loughran, and B. McDonald, 2015. Using 10-K text to gauge financial constraints, *Journal of Financial and Quantitative Analysis* 50*,* 623–646.

Braguinsky, B., and S. Mityakov, 2015. Foreign corporations and the culture of transparency: Evidence from Russian administrative data, *Journal of Financial Economics* 117, 139–164.

Cameron, K. S., J. De Graff, R. E. Quinn, and A. Thakor, 2006. *Competing Values Leadership: Creating Value in Organisations*, Cheltenham, UK: Edward Elgar.

Cartwright, S., and C. L. Cooper, 1993. The role of culture compatibility in successful organizational marriage, *Academy of Management Executive* 7, 57–70.

Chen, X., J. Harford, and K. Li, 2007. Monitoring: Which institutions matter? *Journal of Financial Economics* 86, 279–305.

Cohen, L., D. Lou, and C. Malloy, 2017. Playing favorites: How firms prevent the revelation of bad news, Harvard Business School working paper.

Cronqvist, H., A. Low, and M. Nilsson, 2009. Persistence in firm policies, firm origin, and corporate culture: Evidence from corporate spin-offs, University of Miami working paper.

Davidson, R., A. Dey, and A. Smith, 2015. Executives' "off-the-job" behavior, corporate culture, and financial reporting risk, *Journal of Financial Economics* 117, 5–28.

Edmans, A., 2011. Does the stock market fully value intangibles? Employee satisfaction and equity prices, *Journal of Financial Economics* 101, 621–640.

Erel, I., C. Tan, L. H. Stern, and M. S. Weisbach, 2018. Selecting directors using machine learning, Ohio State University working paper.

Fiordelisi, F., and O. Ricci, 2014. Corporate culture and CEO turnover, *Journal of Corporate Finance* 28, 66–82.

Frankel, R., W. J. Mayew, and Y. Sun, 2010. Do pennies matter? Investor relations consequences of small negative earnings surprises, *Review of Accounting Studies* 15, 220–242.

Gaspar, J.-M., M. Massa, and P. Matos, 2005. Shareholder investment horizons and the market for corporate control, *Journal of Financial Economics* 76, 135–165.

Graham, J. R., J. Grennan, C. R. Harvey, and S. Rajgopal, 2016. Corporate culture: The interview evidence, Duke University working paper.

Graham, J. R., J. Grennan, C. R. Harvey, and S. Rajgopal, 2017. Corporate culture: Evidence from the field, Duke University working paper.

Grennan, J., 2013. A corporate culture channel: How increased shareholder governance reduces firm value? Duke University working paper.

Guiso, L., P. Sapienza, and L. Zingales, 2015. The value of corporate culture, *Journal of Financial Economics* 117, 60–76.

Harris, Z. S., 1954. Distributional structure, *Word* 10, 146–162.

Heron, R. A., and E. Lie, 2009. What fraction of stock option grants to top executives have been backdated or manipulated? *Management Science* 55, 513–525.

Hoberg, G., and V. Maksimovic, 2015. Redefining financial constraints: A text-based analysis, *Review of Financial Studies* 28, 1312–1352.

Hoberg, G., and G. Phillips, 2010. Product market synergies and competition in mergers and acquisitions: A text-based analysis, *Review of Financial Studies* 23, 3773–3811.

Hoberg, G., and G. Phillips, 2017. Product integration and merger success, USC working paper.

Hollander, S., M. Pronk, and E. Roelofsen, 2010. Does silence speak? An empirical analysis of disclosure choices during conference calls, *Journal of Accounting Research* 48, 531–563.

Huang, A. H., R. Lehavy, A. Y. Zang, and R. Zheng, 2018. Analyst information discovery and interpretation roles: A topic modeling approach, *Management Science* 64, 2473–2972.

Jegadeesh, N., and D. Wu, 2013. Word power: A new approach for content analysis, *Journal of Financial Economics* 110, 712–729.

Jovanovic, B., and P. Rousseau, 2002. The Q-theory of mergers, *American Economic Review* 92, 198–204.

Kogan, L., D. Papanikolaou, A. Seru, and N. Stoffman, 2017. Technological innovation, resource allocation, and growth, *Quarterly Journal of Economics* 132, 665–712.

Kreps, D., 1990. Corporate culture and economic theory, in: Alt, J. E., Shepsle, K. A. (eds.), *Perspectives on Positive Political Economy*, Cambridge University Press, Cambridge, 93–104.

Krishnan, H. A., A. Miller, and W. Q. Judge, 1997. Diversification and top management team complementarity: Is performance improved by merging similar or dissimilar teams? *Strategic Management Journal* 18, 361–374.

Larcker, D. F., and A. Zakolyukina, 2012. Detecting deceptive discussions in conference calls, *Journal of Accounting Research* 50, 495–540.

Li, F., 2008, Annual report readability, current earnings, and earnings persistence, *Journal of Accounting and Economics* 45, 221–247.

Li, F., 2010. The information content of forward-looking statements in corporate filings—A naïve Bayesian machine learning approach, *Journal of Accounting Research* 48, 1049–1102.

Li, K., B. Qiu, and R. Shen, 2018. Organization capital and mergers and acquisitions, *Journal of Financial and Quantitative Analysis* in press.

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

Loughran, T., and B. McDonald, 2014. Measuring readability in financial disclosures, *Journal of Finance* 69, 1643–1671.

Loughran, T., and B. McDonald, 2016. Textual analysis in accounting and finance: A survey, *Journal of Accounting Research* 54, 1187–1230.

Magnini, B., C. Strapparava, G. Pezzulo, and A. Gliozzo, 2002. The role of domain information in word sense disambiguation, *Natural Language Engineering* 8, 359–373.

Malmendier, U., and G. Tate, 2008. Who makes acquisitions? CEO overconfidence and the market's reaction, *Journal of Financial Economics* 89, 20–43.

Manne, H., 1965. Mergers and the market for corporate control, *Journal of Political Economy* 73, 110–120.

Matsumoto, D., M. Pronk, and E. Roelofsen, 2011. What makes conference calls useful? The information content of managers' presentations and analysts' discussion sessions, *Accounting Review* 86, 1383–1414.

Mikolov, T., I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, 2013. Distributed representations of words and phrases and their compositionality, *Advances in Neural Information Processing Systems*, 3111–3119.

Moeller, S. B., F. P. Schlingemann, and R. M. Stulz, 2004. Firm size and the gains from acquisitions, *Journal of Financial Economics* 73, 201–228.

Nahavandi, A., and A. R. Malekzadeh, 1988. Acculturation in mergers and acquisitions, *Academy of Management Review* 13, 79–90.

O'Reilly, C., 1989. Corporations, culture, and commitment: Motivation and social control in organizations, *California Management Review* 31, 9–25.

O'Reilly, C., and J.A. Chatman, 1996. Culture as social control: Corporations, cults, and commitment, in: Staw, B. M., and L. L. Cummings (eds.), *Research in Organizational Behavior*, Vol. 18, Greenwich, CT: JAI Press, Inc., 157–200

O'Reilly, C., J. A. Chatman, and D. Caldwell, 1991. People and organizational culture: A profile comparison approach to assessing person-organization fit, *Academy of Management Journal* 34, 487–516.

O'Reilly, C., J. A. Chatman, D. Caldwell, D., and B. Doerr, 2014. The promise and problems of organizational culture: CEO personality, culture, and firm performance, *Group & Organization Management* 39, 595–625.

Pan, Y., S. Siegel, and T. Y. Wang, 2017. Corporate risk culture, *Journal of Financial and Quantitative Analysis* 52, 2327–2367.

Pelevina, M., N. Arefiev, C. Biemann, and A. Panchenko, 2016. Making sense of word embeddings, *Proceedings of the 1st Workshop on Representation Learning for NLP*, 174–183.

Perols, J. L., R. M. Bowen, C. Zimmermann, and B. Samba, 2017. Finding needles in a haystack: Using data analytics to improve fraud prediction, *Accounting Review* 92, 221–245.

Purda, L., and D. Skillicorn, 2015. Accounting variables, deception, and a bag of words: Assessing the tools of fraud detection, *Contemporary Accounting Research* 32, 1193–1223.

Riloff, E., and R. Jones, 1999. Learning dictionaries for information extraction by multi-level bootstrapping, *AAAI/IAAI*, 474–479.

Roll, R., 1986. The hubris hypothesis of corporate takeovers, *Journal of Business* 59, 197–216.

Routledge, B. R., S. Sacchetto, and N. A. Smith, 2018. Predicting merger targets and acquirers from text, Cargenie Mellon University working paper.

Stone, B., S. Dennis, and P. J. Kwantes, 2011. Comparing methods for single paragraph similarity analysis, *Topics in Cognitive Science* 3, 92–122.

Tetlock, P. C., 2007. Giving content to investor sentiment: The role of media in the stock market, *Journal of Finance* 62, 1139–1168.

Tetlock, P. C., M. Saar-Tsechansky, and S. Mackassy, 2008. More than words: Quantifying language to measure firms' fundamentals, *Journal of Finance* 63, 1437–1467.

Van den Steen, E., 2010. Culture clash: The costs and benefits of homogeneity, *Management Science* 56, 1718–1738.

Weber, Y., O. Shenkar, and A. Raveh, 1996. National and corporate cultural fit in mergers/acquisitions: An exploratory study, *Management Science* 42, 1215–1227.

Zingales, L., 2015. The "cultural revolution" in finance, *Journal of Financial Economics* 117, 1–4.

Figure 1
Cultural values across 12 Fama-French industries over time

This figure plots the five cultural values across 12 Fama-French industries over time. The y axis is the average percentage of words in earnings calls based on our culture dictionary for each corporate cultural value.

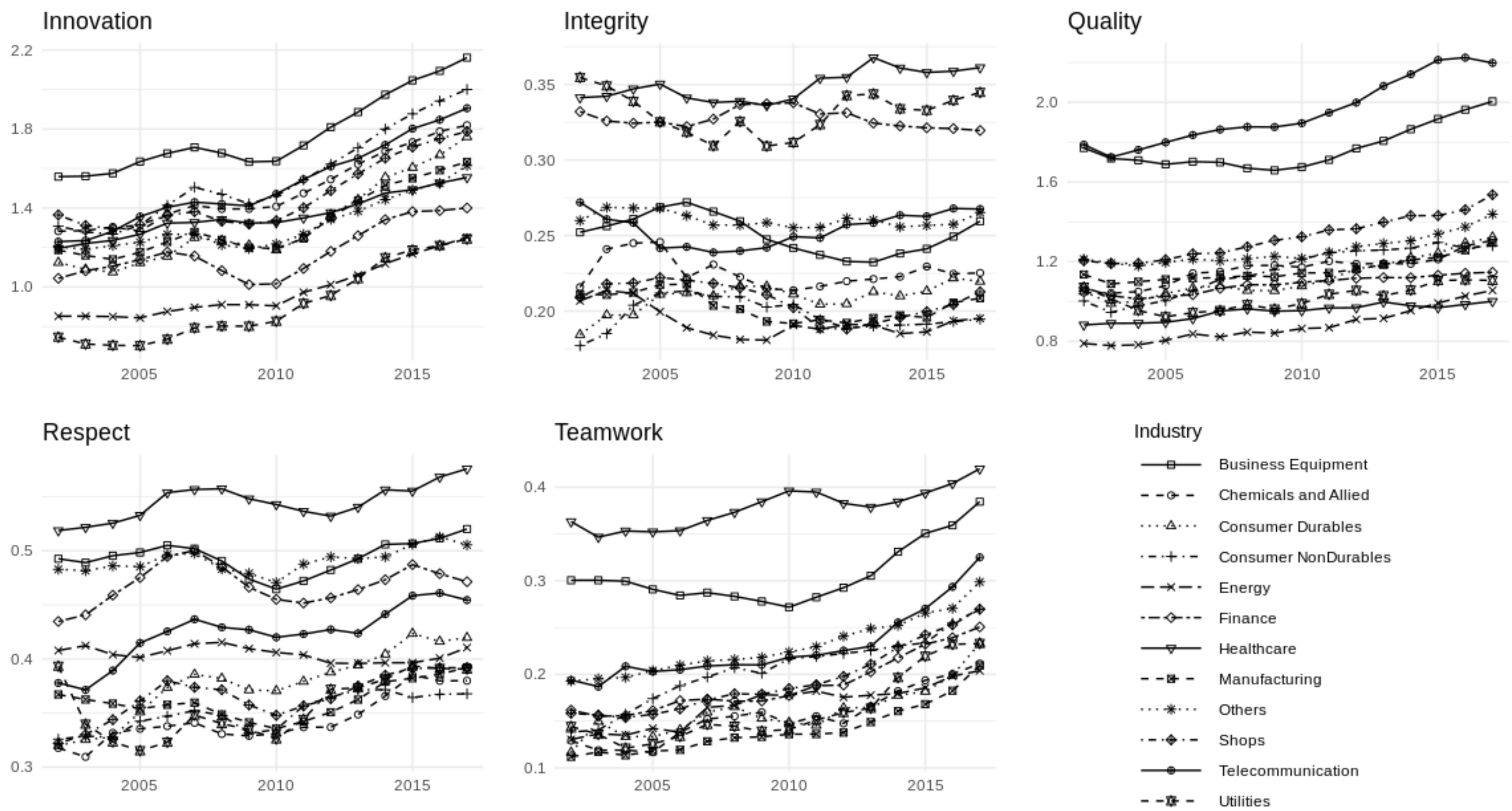


Table 1
Summary statistics of corporate cultural values

The sample consists of 77,541 firm-year observations (8,427 firms) with earnings calls over the period 2001-2018. Panel A provides the summary statistics. Panel B presents the autocorrelations of corporate culture value measures. We calculate the autocorrelation for each firm with more than 15 observations over the sample period. We report the mean, median (in brackets), and standard deviation (in parentheses) of autocorrelations across firms. Panel C presents the correlations between corporate cultural value measures and firm characteristics. Definitions of the variables are provided in Appendix B. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

Panel A: Summary statistics for corporate cultural value measures and firm characteristics

Variable	Obs.	Mean	10th Percentile	Median	90th Percentile	SD
Innovation	77,541	1.420	0.671	1.317	2.313	0.656
Integrity	77,541	0.265	0.094	0.224	0.492	0.176
Quality	77,541	1.294	0.646	1.152	2.157	0.621
Respect	77,541	0.456	0.207	0.407	0.765	0.247
Teamwork	77,541	0.243	0.052	0.181	0.519	0.213
Total assets	59,902	9,522.1	62.734	1,009.0	17,297	32,734
Leverage	58,299	0.243	0.000	0.177	0.613	0.244
ROA	59,856	-0.034	-0.234	0.024	0.112	0.234
Sales growth	58,797	0.038	-0.206	0.065	0.310	0.319
Past return	52,031	0.135	-0.471	0.070	0.732	0.565
Top5 institutions	59,902	0.202	0.000	0.220	0.391	0.159

Panel B: The autocorrelations of corporate cultural value measures

Variable in year t	Obs.	Year t-1	Year t-2	Year t-3	Year t-4	Year t-5
		0.815	0.540	0.219	0.100	0.049
Innovation	1,879	[0.853]	[0.589]	[0.241]	[0.083]	[0.037]
		(0.141)	(0.292)	(0.439)	(0.476)	(0.508)
		0.718	0.391	-0.004	-0.080	-0.118
Integrity	1,879	[0.750]	[0.411]	[-0.034]	[-0.149]	[-0.171]
		(0.170)	(0.288)	(0.398)	(0.411)	(0.441)
		0.774	0.482	0.135	0.012	-0.006
Quality	1,879	[0.813]	[0.524]	[0.136]	[0.008]	[-0.037]
		(0.162)	(0.298)	(0.430)	(0.452)	(0.479)
		0.728	0.404	0.014	-0.069	-0.111
Respect	1,879	[0.764]	[0.422]	[-0.018]	[-0.127]	[-0.173]
		(0.169)	(0.286)	(0.398)	(0.415)	(0.452)
		0.765	0.464	0.120	0.029	-0.023
Teamwork	1,879	[0.800]	[0.501]	[0.102]	[-0.019]	[-0.069]
		(0.158)	(0.299)	(0.431)	(0.455)	(0.479)

Panel C: The correlation matrix

	Innovation	Integrity	Quality	Respect	Teamwork	Firm size	Leverage	ROA	Sales growth	Past return
Innovation	1.000									
Integrity	0.074***	1.000								
Quality	0.447***	0.080***	1.000							
Respect	0.234***	0.248***	0.164***	1.000						
Teamwork	0.274***	0.232***	0.239***	0.352***	1.000					
Firm size	0.035***	0.005	-0.092***	-0.106***	-0.268***	1.000				
Leverage	-0.216***	-0.027***	-0.139***	-0.136***	-0.204***	0.368***	1.000			
ROA	0.015***	-0.124***	-0.027***	-0.118***	-0.286***	0.414***	-0.014***	1.000		
Sales growth	0.028***	-0.007*	-0.002	0.005	-0.009**	0.052***	-0.080***	0.220***	1.000	
Past return	0.004	-0.022***	-0.014***	-0.014***	-0.026***	0.027***	-0.115***	0.177***	0.207***	1.000
Top5 institutions	0.019***	-0.052***	-0.030***	-0.015***	-0.067***	0.043***	-0.073***	0.159***	0.043***	0.027***

Table 2
Top-ranked S&P 500 firms by corporate cultural values

This table presents a snapshot of top-ranked S&P 500 firms by corporate cultural values. Panel A presents the top-ranked S&P 500 firms over the period 2001-2006. Panel B presents the top-ranked S&P 500 firms over the period 2007-2012. Panel C presents the top-ranked S&P 500 firms over the period 2013-2018.

Panel A: Top-ranked S&P 500 firms over the period 2001-2006

Innovation	**Integrity**	**Quality**	**Respect**	**Teamwork**
Ikon Office Solutions	St. Paul Cos.	Alphabet Inc	Apollo Education Group Inc	Citrix Systems Inc
Cognizant Tech Solutions	Allstate Corp	Juniper Networks Inc	Cognizant Tech Solutions	Quintiles Transnational Corp
Linear Technology Corp	Safeco Corp	NetApp Inc	Aetna Inc	DXC Technology Company
Alphabet Inc	Travelers Cos Inc	Sprint PCS Group	Quintiles Transnational Corp	Alphabet Inc
Intl Game Technology	Fannie Mae	Univision Communications Inc	Boston Properties Inc	Novell Inc
Procter & Gamble Co	Encompass Health Corp	Cognizant Tech Solutions	Kimco Realty Corp	Unisys Corp
Home Depot Inc	Genworth Financial Inc	Comcast Corp	Convergys Corp	IMS Health Holdings Inc
Coca-Cola Co	CNO Financial Group Inc	Sprint Corp	S&P Global Inc	Enterasys Networks Inc
Medco Health Solutions Inc	Cincinnati Financial Corp	PTC Inc	Sherwin-Williams Co	First Data Corp
Amazon.com Inc	Progressive Corp	McKesson Corp	Cigna Corp	Symbol Technologies

Panel B: Top-ranked S&P 500 firms over the period 2007-2012

Innovation	**Integrity**	**Quality**	**Respect**	**Teamwork**
Ikon Office Solutions	Cincinnati Financial Corp	Akamai Technologies Inc	Adtalem Global Education Inc	Scripps Networks Interactive
Intl Business Machines Corp	Safeco Corp	Netflix Inc	Apollo Education Group Inc	Novell Inc
Cognizant Tech Solutions	Progressive Corp	Discovery Communications Inc	VF Corp	IMS Health Holdings Inc
Procter & Gamble Co	Genworth Financial Inc	Sprint Corp	Motorola Mobility Hldgs Inc	Citrix Systems Inc
eBay Inc	Ambac Financial Group Inc	Salesforce.com Inc	Aetna Inc	Gap Inc
Adobe Systems Inc	MGIC Investment Corp	Alphabet Inc	Cigna Corp	McAfee Inc
Linear Technology Corp	Tribune Media Co	Juniper Networks Inc	DDR Corp	3com Corp
Medco Health Solutions Inc	Louisiana-Pacific Corp	Amazon.com Inc	Cognizant Tech Solutions	Alphabet Inc
Salesforce.com Inc	National City Corp	Scripps Networks Interactive	Avalonbay Communities Inc	Unisys Corp
Intl Game Technology	Unum Group	eBay Inc	Prologis Inc	Kate Spade & Co

Panel C: Top-ranked S&P 500 firms over the period 2013-2018

Innovation	**Integrity**	**Quality**	**Respect**	**Teamwork**
Tiffany & Co	Cincinnati Financial Corp	Facebook Inc	Adtalem Global Education Inc	Scripps Networks Interactive
Procter & Gamble Co	Genworth Financial Inc	Netflix Inc	Apollo Education Group Inc	Salesforce.com Inc
Cognizant Tech Solutions	MGIC Investment Corp	Paypal Holdings Inc	Cigna Corp	Electronic Arts Inc
Salesforce.com Inc	Ambac Financial Group Inc	Dish Network Corp	Gartner Inc	BroadVision Inc
Adobe Systems Inc	Blackrock Inc	Time Warner Inc	Invesco Ltd	Synchrony Financial
Hanesbrands Inc	Baxalta Inc	Amazon.com Inc	Cognizant Tech Solutions	Walgreens Boots Alliance Inc
Gartner Inc	Allstate Corp	Viacom Inc	Aetna Inc	Citrix Systems Inc
Intl Business Machines Corp	PG&E Corp	Scripps Networks Interactive	Affiliated Managers Grp Inc	Under Armour Inc
Newell Brands Inc	Schein (Henry) Inc	Akamai Technologies Inc	UnitedHealth Group Inc	Fidelity National Info Svcs
Baker Hughes, a GE Co	CNO Financial Group Inc	Sprint Corp	Boston Properties Inc	Blackrock Inc

Table 3
Validating our measures of corporate cultural values

This table validates our measures of corporate cultural values based on the QA section of calls. In Panel A, LnPatent, R&D spending, and innovation strength are used to validate the cultural value of innovation. In Panel B, restatement and backdating are used to validate the cultural value of integrity. In Panel C, product quality, product safety, and top brand are used to validate the cultural value of quality. In Panel D, diversity is used to validate the cultural value of respect. In Panel E, employee involvement and best employer are used to validate the cultural value of teamwork. Ordinary least squares (OLS) regressions are used when the dependent variables are LnPatent, and R&D spending, and probit regressions are used for all other validating variables. Definitions of the variables are provided in Appendix B. Heteroskedasticity-consistent standard errors (in parentheses) are clustered at the firm level. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

Panel A: Validating the cultural value of innovation

Variable	LnPatent (1)	LnPatent (2)	LnPatent (3)	R&D spending (4)	R&D spending (5)	R&D spending (6)	Innovation strength (7)	Innovation strength (8)	Innovation strength (9)
Innovation	0.245***	0.249***	0.148***	0.015***	0.019***	0.012***	0.381***	0.376***	0.261***
	(12.14)	(12.32)	(7.85)	(8.09)	(14.21)	(7.57)	(4.93)	(4.98)	(2.71)
Size	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
ROA	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Ind FE/Yr FE	No	No	Yes	No	No	Yes	No	No	Yes
Intercept	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	25,667	25,667	25,667	57,125	57,125	57,125	12,495	12,495	12,495
R^2 /Pseudo R^2	0.035	0.041	0.238	0.006	0.456	0.553	0.033	0.049	0.143

Panel B: Validating the cultural value of integrity

Variable	Restatement (1)	Restatement (2)	Restatement (3)	Backdating (4)	Backdating (5)	Backdating (6)
Integrity	-0.193***	-0.181***	-0.131*	-0.291***	-0.229**	-0.438***
	(-2.91)	(-2.70)	(-1.82)	(-2.82)	(-2.21)	(-3.78)
Size	No	Yes	Yes	No	Yes	Yes
ROA	No	Yes	Yes	No	Yes	Yes
Ind FE/Yr FE	No	No	Yes	No	No	Yes
Intercept	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	51,461	51,461	51,461	18,122	18,122	18,122
Pseudo R^2	0.001	0.001	0.028	0.002	0.006	0.062

Panel C: Validating the cultural value of quality

Variable	Product quality (1)	Product quality (2)	Product quality (3)	Product safety (4)	Product safety (5)	Product safety (6)	Top brand (7)	Top brand (8)	Top brand (9)
Quality	-0.047	-0.000	-0.019	0.229***	0.253***	0.207***	0.229***	0.415***	0.275***
	(-1.24)	(-0.01)	(-0.30)	(4.60)	(4.65)	(2.65)	(5.85)	(7.10)	(3.74)

Size	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
ROA	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Ind FE/Yr FE	No	No	Yes	No	No	Yes	No	No	Yes
Intercept	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	19,596	19,596	19,596	22,705	22,705	22,705	43,019	43,019	43,019
Pseudo R^2	0.000	0.083	0.279	0.008	0.123	0.272	0.013	0.441	0.488

Panel D: Validating the cultural value of respect

Variable	Diversity (1)	Diversity (2)	Diversity (3)
Respect	-0.020	0.266***	0.203***
	(-0.24)	(3.78)	(2.83)
Size	No	Yes	Yes
ROA	No	Yes	Yes
Ind FE/Yr FE	No	No	Yes
Intercept	Yes	Yes	Yes
Obs.	20,720	20,720	20,720
Pseudo R^2	0.000	0.161	0.331

Panel E: Validating the cultural value of teamwork

Variable	Employee involvement (1)	Employee involvement (2)	Employee involvement (3)	Best employer (4)	Best employer (5)	Best employer (6)
Teamwork	0.671***	1.076***	0.845***	0.128	0.712***	0.481**
	(5.56)	(8.20)	(5.74)	(0.85)	(4.03)	(2.31)
Size	No	Yes	Yes	No	Yes	Yes
ROA	No	Yes	Yes	No	Yes	Yes
Ind FE/Yr FE	No	No	Yes	No	No	Yes
Intercept	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	18,431	18,431	18,431	55,428	55,428	55,428
Pseudo R^2	0.009	0.056	0.137	0.001	0.149	0.250

Table 4
Summary statistics of alternative measures of corporate cultural values

This table presents an overview of alternative measures of corporate cultural values. The suffix _full refers to corporate culture measures based on the entire call (i.e., including both management presentation and QA sections). The suffix _seed refers to corporate culture measures based on a simple count of the seed words (including the value word) in the QA section of calls. The suffix _10k refers to corporate culture measures based on applying the word embedding model to the MD&A section of 10-Ks. Panel A presents the summary statistics. Panel B presents the correlations between our main measure and alternative measures of the cultural value of innovation. Panel C presents the correlations between our main measure and alternative measures of the cultural value of integrity. Panel D presents the correlations between our main measure and alternative measures of the cultural value of quality. Panel E presents the correlations between our main measure and alternative measures of the cultural value of respect. Panel F presents the correlations between our main measure and alternative measures of the cultural value of respect.

Panel A: Summary statistics for alternative measures of corporate cultural values

Variable	Obs.	Mean	10th Percentile	Median	90th Percentile	SD
Innovation_full	78,214	1.995	0.999	1.900	3.129	0.825
Integrity_full	78,214	0.320	0.132	0.275	0.563	0.192
Quality_full	78,214	1.589	0.822	1.430	2.614	0.717
Respect_full	78,214	0.575	0.291	0.520	0.927	0.275
Teamwork_full	78,214	0.301	0.078	0.231	0.620	0.238
Innovation_seed	77,541	0.568	0.208	0.520	0.994	0.315
Integrity_seed	77,541	0.006	0.000	0.000	0.019	0.012
Quality_seed	77,541	0.397	0.088	0.312	0.829	0.318
Respect_seed	77,541	0.063	0.000	0.043	0.147	0.068
Teamwork_seed	77,541	0.001	0.000	0.000	0.003	0.004
Innovation_10k	174,506	0.372	0.046	0.323	0.745	0.282
Integrity_10k	174,506	0.144	0.000	0.108	0.325	0.143
Quality_10k	174,506	0.915	0.292	0.826	1.636	0.564
Respect_10k	174,506	0.234	0.025	0.196	0.472	0.185
Teamwork_10k	174,506	0.047	0.000	0.019	0.126	0.079

Panel B: Correlations between our main and alternative measures of *innovation*

	Innovation	Innovation_full	Innovation_seed	Innovation_10k
Innovation	1.000			
Innovation_full	0.844***	1.000		
Innovation_seed	0.631***	0.555***	1.000	
Innovation_10k	0.214***	0.242***	0.278***	1.000

Panel C: Correlations between our main and alternative measures of *integrity*

	Integrity	Integrity_full	Integrity_seed	Integrity_10k
Integrity	1.000			
Integrity_full	0.830***	1.000		
Integrity_seed	0.265***	0.213***	1.000	
Integrity_10k	0.161***	0.179***	0.037***	1.000

Panel D: Correlations between our main and alternative measures of *quality*

	Quality	Quality_full	Quality_seed	Quality_10k
Quality	1.000			
Quality_full	0.867***	1.000		
Quality_seed	0.732***	0.658***	1.000	
Quality_10k	0.194***	0.220***	0.272***	1.000

Panel E: Correlations between our main and alternative measures of *respect*

	Respect	Respect_full	Respect_seed	Respect_10k
Respect	1.000			
Respect_full	0.813***	1.000		
Respect_seed	0.443***	0.303***	1.000	
Respect_10k	0.166***	0.190***	0.092***	1.000

Panel F: Correlations between our main and alternative measures of *teamwork*

	Teamwork	Teamwork_full	Teamwork_seed	Teamwork_10k
Teamwork	1.000			
Teamwork_full	0.871***	1.000		
Teamwork_seed	0.213***	0.195***	1.000	
Teamwork_10k	0.158***	0.192***	0.086***	1.000

Table 5
Validating alternative measures of corporate cultural values: Using the entire call

This table validates corporate culture measures based on the entire call (i.e., including both management presentation and QA sections). In Panel A, LnPatent, R&D spending, and innovation strength are used to validate the cultural value of innovation. In Panel B, restatement and backdating are used to validate the cultural value of integrity. In Panel C, product quality, product safety, and top brand are used to validate the cultural value of quality. In Panel D, diversity is used to validate the cultural value of respect. In Panel E, employee involvement and best employer are used to validate the cultural value of teamwork. OLS regressions are used when the dependent variables are LnPatent, and R&D spending, and probit regressions are used for all other validating variables. Definitions of the variables are provided in Appendix B. Heteroskedasticity-consistent standard errors (in parentheses) are clustered at the firm level. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

Panel A: Horse race between innovation and innovation_full

Variable	LnPatent (1)	LnPatent (2)	LnPatent (3)	R&D spending (4)	R&D spending (5)	R&D spending (6)	Innovation strength (7)	Innovation strength (8)	Innovation strength (9)
Innovation	0.176***	0.171***	0.075***	0.006*	0.023***	0.016***	0.230	0.274*	0.223
	(5.97)	(5.83)	(2.99)	(1.67)	(9.45)	(7.01)	(1.63)	(1.89)	(1.36)
Innovation_full	0.066***	0.075***	0.074***	0.009***	-0.004**	-0.005***	0.144	0.096	0.038
	(2.87)	(3.23)	(3.45)	(3.59)	(-2.30)	(-2.65)	(1.04)	(0.73)	(0.25)
Size	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
ROA	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Ind FE/Yr FE	No	No	Yes	No	No	Yes	No	No	Yes
Intercept	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	25,667	25,667	25,667	57,125	57,125	57,125	12,495	12,495	12,495
R^2 /Pseudo R^2	0.036	0.043	0.240	0.006	0.456	0.553	0.035	0.050	0.143

Panel B: Horse race between integrity and integrity_full

Variable	Restatement (1)	Restatement (2)	Restatement (3)	Backdating (4)	Backdating (5)	Backdating (6)
Integrity	-0.452***	-0.457***	-0.344***	-0.433**	-0.387**	-0.392**
	(-4.17)	(-4.21)	(-3.09)	(-2.25)	(-2.02)	(-2.02)
Integrity_full	0.284***	0.305***	0.253**	0.158	0.176	-0.055
	(2.83)	(3.01)	(2.39)	(0.88)	(0.98)	(-0.29)
Size	No	Yes	Yes	No	Yes	Yes
ROA	No	Yes	Yes	No	Yes	Yes
Ind FE/Yr FE	No	No	Yes	No	No	Yes
Intercept	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	51,461	51,461	51,461	18,122	18,122	18,122
Pseudo R^2	0.001	0.001	0.028	0.003	0.007	0.062

Panel C: Horse race between quality and quality_full

Variable	Product quality (1)	Product quality (2)	Product quality (3)	Product safety (4)	Product safety (5)	Product safety (6)	Top brand (7)	Top brand (8)	Top brand (9)
Quality	0.198**	0.274***	0.118	0.070	0.038	0.165	0.188**	0.211*	0.041
	(2.31)	(3.01)	(1.09)	(0.66)	(0.34)	(1.26)	(2.41)	(1.78)	(0.32)
Quality_full	-0.237***	-0.266***	-0.141	0.151	0.205**	0.043	0.033	0.203**	0.246**
	(-3.28)	(-3.55)	(-1.55)	(1.60)	(2.11)	(0.40)	(0.46)	(2.00)	(2.24)
Size	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
ROA	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Ind FE/Yr FE	No	No	Yes	No	No	Yes	No	No	Yes
Intercept	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	19,596	19,596	19,596	22,705	22,705	22,705	43,019	43,019	43,019
Pseudo R^2	0.003	0.085	0.280	0.009	0.124	0.272	0.012	0.443	0.493

Panel D: Horse race between respect and respect_full

Variable	Diversity (1)	Diversity (2)	Diversity (3)
Respect	0.431***	-0.150	-0.131
	(3.36)	(-1.26)	(-1.22)
Respect_full	-0.459***	0.432***	0.370***
	(-3.80)	(3.80)	(3.37)
Size	No	Yes	Yes
ROA	No	Yes	Yes
Ind FE/Yr FE	No	No	Yes
Intercept	Yes	Yes	Yes
Obs.	20,720	20,720	20,720
Pseudo R^2	0.002	0.163	0.333

Panel E: Horse race between teamwork and teamwork_full

Variable	Employee involvement (1)	Employee involvement (2)	Employee involvement (3)	Best employer (4)	Best employer (5)	Best employer (6)
Teamwork	0.752***	0.763***	0.604**	0.406**	0.335	-0.009
	(2.69)	(2.73)	(2.02)	(2.01)	(1.26)	(-0.03)
Teamwork_full	-0.078	0.303	0.241	-0.286	0.389	0.518
	(-0.30)	(1.20)	(0.89)	(-1.42)	(1.60)	(1.56)
Size	No	Yes	Yes	No	Yes	Yes
ROA	No	Yes	Yes	No	Yes	Yes
Ind FE/Yr FE	No	No	Yes	No	No	Yes
Intercept	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	18,431	18,431	18,431	55,428	55,428	55,428
Pseudo R^2	0.009	0.057	0.138	0.001	0.149	0.251

Table 6
Validating alternative measures of corporate cultural values: Using a simple count

This table validates corporate culture measures based on a simple count of the seed words (including the value word) in the QA section of calls. In Panel A, LnPatent, R&D spending, and innovation strength are used to validate the cultural value of innovation. In Panel B, restatement and backdating are used to validate the cultural value of integrity. In Panel C, product quality, product safety, and top brand are used to validate the cultural value of quality. In Panel D, diversity is used to validate the cultural value of respect. In Panel E, employee involvement and best employer are used to validate the cultural value of teamwork. OLS regressions are used when the dependent variables are LnPatent, and R&D spending, and probit regressions are used for all other validating variables. Definitions of the variables are provided in Appendix B. Heteroskedasticity-consistent standard errors (in parentheses) are clustered at the firm level. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

Panel A: Horse race between innovation and innovation_seed

Variable	LnPatent (1)	LnPatent (2)	LnPatent (3)	R&D spending (4)	R&D spending (5)	R&D spending (6)	Innovation strength (7)	Innovation strength (8)	Innovation strength (9)
Innovation	0.243***	0.279***	0.169***	0.060***	0.029***	0.022***	0.258***	0.313***	0.190*
	(10.18)	(11.01)	(7.57)	(20.56)	(14.67)	(10.99)	(3.10)	(3.82)	(1.93)
Innovation_seed	0.005	-0.101**	-0.076*	-0.144***	-0.033***	-0.036***	0.410**	0.210	0.243
	(0.12)	(-2.35)	(-1.93)	(-22.59)	(-8.96)	(-10.15)	(2.12)	(1.27)	(1.10)
Size	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
ROA	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Ind FE/Yr FE	No	No	Yes	No	No	Yes	No	No	Yes
Intercept	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	25,667	25,667	25,667	57,125	57,125	57,125	12,495	12,495	12,495
R^2 /Pseudo R^2	0.035	0.042	0.239	0.077	0.459	0.557	0.039	0.050	0.144

Panel B: Horse race between integrity and integrity_seed

Variable	Restatement (1)	Restatement (2)	Restatement (3)	Backdating (4)	Backdating (5)	Backdating (6)
Integrity	-0.216***	-0.203***	-0.149**	-0.280***	-0.222**	-0.438***
	(-3.15)	(-2.94)	(-2.01)	(-2.64)	(-2.07)	(-3.69)
Integrity_seed	1.186	1.157	0.917	-0.738	-0.486	0.053
	(1.33)	(1.30)	(1.00)	(-0.46)	(-0.30)	(0.03)
Size	No	Yes	Yes	No	Yes	Yes
ROA	No	Yes	Yes	No	Yes	Yes
Ind FE/Yr FE	No	No	Yes	No	No	Yes
Intercept	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	51,461	51,461	51,461	18,122	18,122	18,122
Pseudo R^2	0.001	0.001	0.028	0.001	0.006	0.062

Panel C: Horse race between quality and quality_seed

Variable	Product quality (1)	Product quality (2)	Product quality (3)	Product safety (4)	Product safety (5)	Product safety (6)	Top brand (7)	Top brand (8)	Top brand (9)
Quality	0.024	0.003	-0.075	0.072	0.155**	0.140	0.372***	0.419***	0.357***
	(0.47)	(0.06)	(-0.88)	(1.12)	(2.21)	(1.33)	(7.99)	(5.63)	(3.95)
Quality_seed	-0.186*	-0.009	0.142	0.433***	0.285*	0.177	-0.428***	-0.012	-0.253
	(-1.68)	(-0.07)	(0.86)	(3.32)	(1.88)	(0.94)	(-4.54)	(-0.07)	(-1.41)
Size	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
ROA	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Ind FE/Yr FE	No	No	Yes	No	No	Yes	No	No	Yes
Intercept	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	19,596	19,596	19,596	22,705	22,705	22,705	43,019	43,019	43,019
Pseudo R^2	0.001	0.083	0.279	0.011	0.124	0.272	0.018	0.441	0.489

Panel D: Horse race between respect and respect_seed

Variable	Diversity (1)	Diversity (2)	Diversity (3)
Respect	-0.178**	0.240***	0.240***
	(-1.98)	(3.07)	(3.03)
Respect_seed	1.344***	0.215	-0.277
	(4.23)	(0.80)	(-1.13)
Size	No	Yes	Yes
ROA	No	Yes	Yes
Ind FE/Yr FE	No	No	Yes
Intercept	Yes	Yes	Yes
Obs.	20,720	20,720	20,720
Pseudo R^2	0.003	0.161	0.332

Panel E: Horse race between teamwork and teamwork_seed

Variable	Employee involvement (1)	Employee involvement (2)	Employee involvement (3)	Best employer (4)	Best employer (5)	Best employer (6)
Teamwork	0.645***	1.033***	0.778***	0.105	0.679***	0.422**
	(5.21)	(7.73)	(5.14)	(0.70)	(3.81)	(1.97)
Teamwork_seed	5.024	9.635	12.516*	5.168	7.704	10.021
	(0.86)	(1.57)	(1.95)	(0.95)	(1.14)	(1.37)
Size	No	Yes	Yes	No	Yes	Yes
ROA	No	Yes	Yes	No	Yes	Yes
Ind FE/Yr FE	No	No	Yes	No	No	Yes
Intercept	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	18,431	18,431	18,431	55,428	55,428	55,428
/Pseudo R^2	0.009	0.057	0.138	0.001	0.148	0.250

Table 7
Validating alternative measures of corporate cultural values: Using 10-Ks

This table validates corporate culture measures based on applying the word embedding model to the MD&A section of 10-Ks. In Panel A, LnPatent, R&D spending, and innovation strength are used to validate the cultural value of innovation. In Panel B, restatement and backdating are used to validate the cultural value of integrity. In Panel C, product quality, product safety, and top brand are used to validate the cultural value of quality. In Panel D, diversity is used to validate the cultural value of respect. In Panel E, employee involvement and best employer are used to validate the cultural value of teamwork. OLS regressions are used when the dependent variables are LnPatent, and R&D spending, and probit regressions are used for all other validating variables. Definitions of the variables are provided in Appendix B. Heteroskedasticity-consistent standard errors (in parentheses) are clustered at the firm level. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

Panel A: Horse race between innovation and innovation_10k

Variable	LnPatent (1)	LnPatent (2)	LnPatent (3)	R&D spending (4)	R&D spending (5)	R&D spending (6)	Innovation strength (7)	Innovation strength (8)	Innovation strength (9)
Innovation	0.309***	0.323***	0.194***	0.024***	0.022***	0.013***	0.339***	0.343***	0.228**
	(12.71)	(13.12)	(8.76)	(10.92)	(13.58)	(7.38)	(4.26)	(4.30)	(2.15)
Innovation_10k	-0.025	-0.065	-0.085**	-0.072***	-0.023***	-0.016***	0.291	0.205	0.158
	(-0.51)	(-1.35)	(-2.01)	(-16.57)	(-8.29)	(-6.03)	(1.35)	(0.94)	(0.64)
Size	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
ROA	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Ind FE/Yr FE	No	No	Yes	No	No	Yes	No	No	Yes
Intercept	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	19,836	19,836	19,836	43,817	43,817	43,817	11,037	11,037	11,037
R^2 /Pseudo R^2	0.048	0.060	0.287	0.029	0.468	0.563	0.033	0.050	0.149

Panel B: Horse race between integrity and integrity_10k

Variable	Restatement (1)	Restatement (2)	Restatement (3)	Backdating (4)	Backdating (5)	Backdating (6)
Integrity	-0.167**	-0.151**	-0.061	-0.033	0.076	-0.095
	(-2.25)	(-2.01)	(-0.75)	(-0.25)	(0.57)	(-0.63)
Integrity_10k	-0.047	-0.073	-0.031	-0.215	-0.206	-0.338*
	(-0.44)	(-0.70)	(-0.28)	(-1.20)	(-1.12)	(-1.80)
Size	No	Yes	Yes	No	Yes	Yes
ROA	No	Yes	Yes	No	Yes	Yes
Ind FE/Yr FE	No	No	Yes	No	No	Yes
Intercept	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	39,191	39,191	39,191	13,895	13,895	13,895
Pseudo R^2	0.000	0.001	0.030	0.000	0.013	0.084

Panel C: Horse race between quality and quality_10k

Variable	Product quality (1)	Product quality (2)	Product quality (3)	Product safety (4)	Product safety (5)	Product safety (6)	Top brand (7)	Top brand (8)	Top brand (9)
Quality	-0.058	0.006	-0.012	0.268***	0.276***	0.295***	0.274***	0.462***	0.363***
	(-1.44)	(0.12)	(-0.18)	(5.01)	(4.67)	(3.37)	(5.69)	(6.87)	(4.04)
Quality_10k	-0.016	-0.046	-0.010	0.051	0.086	0.078	-0.007	-0.028	-0.028
	(-0.44)	(-1.27)	(-0.21)	(0.99)	(1.57)	(1.43)	(-0.12)	(-0.37)	(-0.39)
Size	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
ROA	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Ind FE/Yr FE	No	No	Yes	No	No	Yes	No	No	Yes
Intercept	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	17,556	17,556	17,556	20,264	20,264	20,264	31,555	31,555	31,555
Pseudo R^2	0.001	0.082	0.284	0.011	0.122	0.270	0.019	0.407	0.487

Panel D: Horse race between respect and respect_10k

Variable	Diversity (1)	Diversity (2)	Diversity (3)
Respect	-0.035	0.195***	0.156**
	(-0.45)	(2.91)	(2.26)
Respect_10k	0.038	0.311***	0.095
	(0.33)	(3.11)	(1.05)
Size	No	Yes	Yes
ROA	No	Yes	Yes
Ind FE/Yr FE	No	No	Yes
Intercept	Yes	Yes	Yes
Obs.	18,376	18,376	18,376
Pseudo R^2	0.000	0.140	0.320

Panel E: Horse race between teamwork and teamwork_10k

Variable	Employee involvement (1)	Employee involvement (2)	Employee involvement (3)	Best employer (4)	Best employer (5)	Best employer (6)
Teamwork	0.661***	1.026***	0.782***	0.116	0.713***	0.418*
	(5.05)	(7.34)	(5.01)	(0.71)	(3.73)	(1.78)
Teamwork_10k	-0.185	0.131	0.324	-0.447	-0.273	0.384
	(-0.49)	(0.33)	(0.73)	(-0.61)	(-0.33)	(0.44)
Size	No	Yes	Yes	No	Yes	Yes
ROA	No	Yes	Yes	No	Yes	Yes
Ind FE/Yr FE	No	No	Yes	No	No	Yes
Intercept	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	16,355	16,355	16,355	42,352	42,352	42,352
Pseudo R^2	0.009	0.055	0.140	0.001	0.155	0.275

Table 8
Deal incidence and merger pairing

This table examines the relation between corporate cultural values and acquisitiveness, and between cultural fit and merger pairing. The acquirer sample consists of 7,875 completed deals between 2003 and 2017 from the Thomson One Banker SDC database. The acquirer-target sample consists of 734 completed deals where both the acquirer and its target firms are public with available control firms. Panel A examines the relation between a firm's cultural values and its probability of being an acquirer. The dependent variable is equal to one for the acquirer, and zero for other firm-years in the full Compustat sample or matched acquirers that form the control group. Panel B examines the relation between cultural fit measures and acquirer-target firm pairing. The dependent variable is equal to one for the acquirer-target firm pair, and zero for the control firm pairs. The coefficients are estimated from linear probability models (LPM) and conditional logit models (Clogit). Definitions of the variables are provided in Appendix B. All specifications based on the matched samples include deal fixed effects. Robust standard errors are reported in the parentheses. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

Panel A: Deal incidence

	Full sample	Industry and size-matched		Industry, size, and B/M-matched	
	LPM	LPM	Clogit	LPM	Clogit
Variable	(1)	(2)	(3)	(4)	(5)
Innovation	0.022***	0.028***	0.168***	0.031***	0.189***
	(0.006)	(0.005)	(0.028)	(0.005)	(0.029)
Integrity	-0.046***	-0.059***	-0.390***	-0.050***	-0.329***
	(0.017)	(0.018)	(0.100)	(0.018)	(0.101)
Quality	-0.029***	-0.041***	-0.265***	-0.045***	-0.292***
	(0.006)	(0.005)	(0.030)	(0.005)	(0.031)
Respect	-0.022*	-0.036***	-0.208***	-0.032***	-0.180***
	(0.013)	(0.012)	(0.067)	(0.012)	(0.068)
Teamwork	-0.018	0.019	0.036	0.019	0.062
	(0.015)	(0.016)	(0.088)	(0.016)	(0.090)
Firm size	0.006***	0.370***	2.217***	0.351***	2.122***
	(0.002)	(0.008)	(0.076)	(0.007)	(0.069)
Leverage	-0.092***	-0.149***	-0.935***	-0.125***	-0.771***
	(0.012)	(0.013)	(0.084)	(0.014)	(0.089)
ROA	0.091***	0.036	0.256**	0.017	0.139
	(0.014)	(0.022)	(0.120)	(0.025)	(0.130)
Sales growth	0.001**	0.004***	0.463***	0.004***	0.480***
	(0.000)	(0.001)	(0.033)	(0.001)	(0.034)
Past return	0.033***	0.048***	0.243***	0.049***	0.248***
	(0.004)	(0.005)	(0.027)	(0.006)	(0.028)
Top 5 institutions	0.137***	0.230***	1.508***	0.242***	1.604***
	(0.019)	(0.019)	(0.113)	(0.019)	(0.117)
Ind FE/Yr FE	Yes	No	No	No	No
Deal FE	No	Yes	Yes	Yes	Yes
Intercept	Yes	Yes		Yes	
Obs.	49,681	40,387	40,382	39,249	39,244
R^2/Pseudo R^2	0.0729	0.0651	0.0705	0.0699	0.0772

Panel B: Merger pairing

Variable	Industry and size-matched (1)	(2)	Industry, size, and B/M-matched (3)	(4)
Cultural similarity	4.812***		4.604***	
	(1.083)		(1.110)	
Cultural distance		-0.668***		-0.683***
		(0.102)		(0.104)
Acquirer Characteristics				
Firm size	-0.395	-0.309	-0.600**	-0.561*
	(0.272)	(0.273)	(0.302)	(0.301)
Leverage	0.286**	0.286**	0.249**	0.257**
	(0.122)	(0.120)	(0.117)	(0.114)
ROA	-0.187	-0.225	0.075	-0.004
	(0.275)	(0.276)	(0.304)	(0.306)
Sales growth	-0.094	-0.100	-0.119	-0.125
	(0.093)	(0.094)	(0.096)	(0.098)
Past return	1.974***	1.972***	1.989***	1.986***
	(0.341)	(0.343)	(0.365)	(0.366)
Top 5 institutions	2.159***	2.152***	1.847***	1.872***
	(0.256)	(0.258)	(0.239)	(0.243)
Target Characteristics				
Firm size	2.778***	2.824***	2.615***	2.681***
	(0.191)	(0.194)	(0.191)	(0.191)
Leverage	-1.125***	-1.155***	-0.955***	-0.996***
	(0.279)	(0.282)	(0.310)	(0.311)
ROA	-0.107	-0.115	-0.009	-0.039
	(0.457)	(0.462)	(0.547)	(0.552)
Sales growth	0.454***	0.438***	0.426***	0.410***
	(0.120)	(0.122)	(0.140)	(0.142)
Past return	0.164	0.162	0.092	0.098
	(0.111)	(0.114)	(0.122)	(0.125)
Top 5 institutions	1.157***	1.107***	1.230***	1.163***
	(0.361)	(0.352)	(0.395)	(0.383)
Deal Characteristics				
Same state	0.885***	0.881***	0.890***	0.891***
	(0.132)	(0.133)	(0.141)	(0.141)
HP similarity	27.041***	26.887***	27.599***	27.434***
	(1.781)	(1.783)	(1.921)	(1.925)
Deal FE	Yes	Yes	Yes	Yes
Obs.	7,049	7,049	6,799	6,799
R^2/Pseudo R^2	0.289	0.293	0.295	0.300

Table 9
Deal outcomes

This table examines the relation between acquirer-target culture fit and deal outcomes. The sample consists of 810 completed deals between 2003 and 2017 from the Thomson One Banker SDC database. We further require that acquirers do not engage in any other significant deals in the year after deal completion when the dependent variables are one-year buy-and-hold abnormal returns (BHAR1), the change in ROA from deal announcement to one year after deal completion (ΔROA1), integration, and retention after deal completion. Panel A reports the regression results where the dependent variable is combined CAR(-1,1). Panel B reports the regression results where the dependent variable is BHAR1. Panel C reports the regression results where the dependent variable is ΔROA1. Panel D reports the regression results where the dependent variable is integration. Panel E reports the regression results where the dependent variable is retention. Acquirer and target two-digit SIC industry and year fixed effects are included. Definitions of the variables are provided in Appendix B. Heteroskedasticity-consistent standard errors (in parentheses) are clustered at the acquirer level. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

Panel A: Culture fit and combined CAR(-1, 1)

	Combined CAR(-1, 1)	Combined CAR(-1, 1)	Combined CAR(-1, 1)	Combined CAR(-1, 1)	Combined CAR(-1, 1)
	(1)	(2)	(3)	(4)	(5)
Cultural similarity	4.130				
	(4.898)				
Cultural distance		-1.254**			
		(0.508)			
Ext-Int conflict			-0.732		
			(0.607)		
Ext-Int conflict2				-0.819*	
				(0.497)	
People focus					1.493
					(1.012)
Acquirer characteristics					
Firm size	-1.235***	-1.209***	-1.195***	-1.178***	-1.216***
	(0.240)	(0.238)	(0.237)	(0.238)	(0.238)
Leverage	7.072***	6.743***	7.020***	6.929***	7.045***
	(2.303)	(2.266)	(2.303)	(2.305)	(2.300)
ROA	0.619	0.648	0.574	0.572	0.613
	(4.834)	(4.773)	(4.784)	(4.783)	(4.774)
Sales growth	-0.900	-0.852	-0.941	-0.936	-1.086
	(0.866)	(0.852)	(0.857)	(0.852)	(0.862)
Past return	0.894	0.903	0.927	0.954	0.941
	(0.905)	(0.903)	(0.901)	(0.898)	(0.895)
Top5 institutions	-2.781	-2.690	-2.525	-2.451	-2.719
	(2.477)	(2.487)	(2.489)	(2.498)	(2.467)
Target characteristics					
Firm size	0.660***	0.642**	0.665***	0.667***	0.694***
	(0.255)	(0.255)	(0.254)	(0.254)	(0.255)
Leverage	1.793	1.735	1.725	1.649	1.632
	(1.792)	(1.783)	(1.795)	(1.794)	(1.802)
ROA	4.354***	4.258***	4.291***	4.216***	4.384***
	(1.606)	(1.608)	(1.607)	(1.581)	(1.598)

Sales growth	0.269	0.350	0.306	0.322	0.284
	(0.766)	(0.765)	(0.767)	(0.762)	(0.764)
Past return	-1.099**	-1.110**	-1.097**	-1.091**	-1.069**
	(0.541)	(0.541)	(0.542)	(0.543)	(0.542)
Top5 institutions	-3.162	-3.103	-3.064	-3.139	-3.089
	(2.107)	(2.081)	(2.112)	(2.109)	(2.123)
Deal characteristics					
All cash	2.676***	2.736***	2.617***	2.625***	2.585***
	(0.693)	(0.681)	(0.692)	(0.693)	(0.691)
All stock	-0.978	-0.941	-0.984	-0.980	-1.016
	(0.791)	(0.786)	(0.788)	(0.788)	(0.787)
Diversifying	0.146	0.058	0.117	0.118	0.157
	(0.569)	(0.573)	(0.575)	(0.568)	(0.575)
Tender offer	-0.635	-0.721	-0.647	-0.627	-0.639
	(0.626)	(0.629)	(0.627)	(0.626)	(0.619)
Relative size	1.031***	1.057***	1.080***	1.075***	1.031***
	(0.330)	(0.325)	(0.329)	(0.327)	(0.329)
Same state	-0.214	-0.248	-0.175	-0.190	-0.211
	(0.548)	(0.548)	(0.545)	(0.545)	(0.549)
HP similarity	1.684	1.393	1.872	1.981	1.427
	(3.024)	(2.992)	(2.987)	(3.011)	(3.029)
Ind FE/Yr FE	Yes	Yes	Yes	Yes	Yes
Intercept	Yes	Yes	Yes	Yes	Yes
Obs.	810	810	810	810	810
R^2	0.341	0.346	0.342	0.343	0.343

Panel B: Culture fit and BHAR1

	BHAR1 (1)	BHAR1 (2)	BHAR1 (3)	BHAR1 (4)	BHAR1 (5)
Cultural similarity	0.488				
	-0.326				
Cultural distance		0.043			
		(0.045)			
Ext-Int conflict			-0.013		
			(0.059)		
Ext-Int conflict2				-0.035	
				(0.045)	
People focus					-0.088
					(0.079)
Acquirer/target/deal controls	Yes	Yes	Yes	Yes	Yes
Ind FE/Yr FE	Yes	Yes	Yes	Yes	Yes
Intercept	Yes	Yes	Yes	Yes	Yes
Obs.	636	636	636	636	636
R^2	0.274	0.272	0.271	0.272	0.273

Panel C: Culture fit and ΔROA1

	ΔROA1 (1)	ΔROA1 (2)	ΔROA1 (3)	ΔROA1 (4)	ΔROA1 (5)
Cultural similarity	-3.046				
	(9.218)				
Cultural distance		0.477			
		(0.978)			
Ext-Int conflict			0.765		
			(1.118)		
Ext-Int conflict2				-0.200	
				(0.983)	
People focus					4.672***
					(1.389)
Acquirer/target/deal controls	Yes	Yes	Yes	Yes	Yes
Ind FE/Yr FE	Yes	Yes	Yes	Yes	Yes
Intercept	Yes	Yes	Yes	Yes	Yes
Obs.	566	566	566	566	566
R^2	0.463	0.464	0.464	0.463	0.472

Panel D: Culture fit and post-merger integration problems

	Integration (1)	Integration (2)	Integration (3)	Integration (4)	Integration (5)
Cultural similarity	-0.326				
	(1.324)				
Cultural distance		0.073			
		(0.159)			
Ext-Int conflict			0.538***		
			(0.207)		
Ext-Int conflict2				0.441***	
				(0.164)	
People focus					-0.718***
					(0.263)
Acquirer/target/deal controls	Yes	Yes	Yes	Yes	Yes
Ind FE/Yr FE	Yes	Yes	Yes	Yes	Yes
Intercept	Yes	Yes	Yes	Yes	Yes
Obs.	579	579	579	579	579
Pseudo R^2	0.259	0.259	0.269	0.269	0.268

Panel E: Culture fit and post-merger retention problems

	Retention (1)	Retention (2)	Retention (3)	Retention (4)	Retention (5)
Cultural similarity	0.695				
	(1.239)				
Cultural distance		-0.196			
		(0.149)			
Ext-Int conflict			0.083		
			(0.162)		
Ext-Int conflict2				0.058	
				(0.140)	
People focus					-0.268
					(0.262)
Acquirer/target/deal controls	Yes	Yes	Yes	Yes	Yes
Ind FE/Yr FE	Yes	Yes	Yes	Yes	Yes
Intercept	Yes	Yes	Yes	Yes	Yes
Obs.	609	609	609	609	609
Pseudo R^2	0.144	0.146	0.144	0.144	0.145

Table 10
Acculturation

This table examines acculturation after deal completion. The sample consists of 605 (415) deals one year (three years) after deal completion from the Thomson One Banker SDC database. We further require that acquirers do not engage in any other significant deals in the year (three years) after deal completion. OLS regression results are reported. Definitions of the variables are provided in Appendix B. Heteroskedasticity-consistent standard errors (in parentheses) are clustered at the acquirer level. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

	$Innovation_{t+1}$	$Innovation_{t+3}$	$Integrity_{t+1}$	$Integrity_{t+3}$	$Quality_{t+1}$	$Quality_{t+3}$	$Respect_{t+1}$	$Respect_{t+3}$	$Teamwork_{t+1}$	$Teamwork_{t+3}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Acquirer innovation	0.737***	0.693***	0.018*	0.021	0.023	0.049	0.016	0.031	0.021**	0.018
	(0.036)	(0.056)	(0.011)	(0.014)	(0.031)	(0.037)	(0.016)	(0.020)	(0.010)	(0.014)
Target innovation	0.041	0.033	-0.003	-0.005	0.024	0.017	-0.013	-0.006	-0.005	0.011
	(0.028)	(0.050)	(0.008)	(0.014)	(0.027)	(0.041)	(0.011)	(0.018)	(0.009)	(0.014)
Acquirer integrity	-0.131	-0.187	0.516***	0.397***	-0.067	-0.101	-0.043	-0.077	0.027	-0.058
	(0.141)	(0.223)	(0.053)	(0.078)	(0.118)	(0.166)	(0.055)	(0.082)	(0.043)	(0.064)
Target integrity	0.022	-0.021	0.088***	0.129***	0.029	0.029	0.039	0.128**	-0.016	0.014
	(0.117)	(0.171)	(0.030)	(0.047)	(0.096)	(0.126)	(0.046)	(0.062)	(0.029)	(0.047)
Acquirer quality	-0.005	-0.017	0.000	-0.009	0.632***	0.547***	0.000	0.008	0.017	0.037**
	(0.042)	(0.061)	(0.011)	(0.019)	(0.049)	(0.062)	(0.019)	(0.021)	(0.013)	(0.018)
Target quality	0.014	0.054	-0.012	-0.012	0.053*	0.065*	0.005	-0.011	-0.013	-0.020*
	(0.031)	(0.042)	(0.008)	(0.012)	(0.029)	(0.035)	(0.012)	(0.015)	(0.008)	(0.012)
Acquirer respect	-0.092	0.094	0.020	0.087**	0.136	0.235**	0.677***	0.602***	-0.010	0.005
	(0.103)	(0.157)	(0.030)	(0.043)	(0.087)	(0.115)	(0.041)	(0.057)	(0.027)	(0.041)
Target respect	-0.023	-0.069	0.017	0.010	-0.143**	-0.177*	0.071**	0.083*	0.012	-0.004
	(0.069)	(0.114)	(0.021)	(0.033)	(0.059)	(0.095)	(0.030)	(0.045)	(0.021)	(0.031)
Acquirer teamwork	0.267**	0.242	0.050	0.026	0.170	0.057	0.061	0.016	0.668***	0.610***
	(0.127)	(0.185)	(0.035)	(0.053)	(0.115)	(0.142)	(0.056)	(0.064)	(0.041)	(0.063)
Target teamwork	0.080	0.193	0.011	0.036	0.149*	0.274**	-0.029	-0.015	0.082***	0.129***
	(0.082)	(0.117)	(0.022)	(0.036)	(0.078)	(0.117)	(0.035)	(0.046)	(0.027)	(0.038)
Acquirer/target/deal controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ind FE/Yr FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Intercept	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	605	415	605	415	605	415	605	415	605	415
R^2	0.829	0.804	0.654	0.626	0.828	0.807	0.669	0.660	0.768	0.756

Internet Appendix for "Measuring Corporate Culture Using Machine Learning"

Table IA1. Overview of our sample

The Thomson Reuters' SE database covers the period from January 1, 2001 to May 25, 2018, and Factiva covers the period from January 1, 2002 to December 31, 2017. Panel A provides the distribution of types of calls in the SE database. Panel B provides an example of backfilling by Thomson Reuters when an acquisition takes place, the target firm name and ticker symbol are backfilled by its acquirer's. Panel C provides the distribution of earnings calls ("Earning Conference Call/Presentation") over time from SE and Factiva. Panel D compares firm-years from the SE database and those from Factiva.

Panel A: Types of calls

Call type	eventTypeId	# calls	Percentage
Earning Conference Call/Presentation	1	270,879	69.26%
Conference Presentation	7	73,158	18.71%
Corporate Conference Call/Presentation	5	15,531	3.97%
Sales Conference Call/Presentation	33	7,546	1.93%
Type Missing	25	6,743	1.72%
M&A Conference Call/Presentation	34	6,150	1.57%
Analyst Meeting	31	5,451	1.39%
Shareholder Meeting	11	3,070	0.78%
Guidance Conference Call/Presentation	30	1,435	0.37%
Other Corporate Conference Event	8	898	0.23%
Other		230	0.06%
Total		391,091	100%

Panel B: A backfilling example

Transcript ID	Event title	Backfilled company name (Provided by SE)	Extracted company name from the body of a call
703530.F	Q4 2002 Duratek, Inc. Earnings Conference Call	EnergySolutions Inc	Duratek, Inc.
703538.F	Q3 2003 Duratek, Inc. Earnings Conference Call	EnergySolutions Inc	Duratek, Inc.
986680.F	Q4 2004 Duratek, Inc. Earnings Conference Call	EnergySolutions Inc	Duratek, Inc.
1054564.F	Q1 2005 Duratek, Inc. Earnings Conference Call	EnergySolutions Inc	Duratek, Inc.

Panel C: Temporal distribution of earnings calls

Year	# Calls from SE (for training embedding model)	# Firms from SE	# Calls from Factiva	# Firms from Factiva	# Total firms (after padding and three-year rolling window)
2001	417	112	194	188	300
2002	5,913	1,984	1,155	568	2,577
2003	9,035	2,510	1,546	660	3,455
2004	10,214	2,678	2,943	810	3,977

2005	11,072	2,933	2,511	848	4,366
2006	11,606	3,083	2,005	822	4,702
2007	12,221	3,311	1,503	577	5,049
2008	12,737	3,359	1,681	519	5,050
2009	12,524	3,229	1,579	465	4,809
2010	12,606	3,271	1,971	788	4,879
2011	12,651	3,365	2,903	924	5,056
2012	12,115	3,266	4,296	1,021	5,188
2013	11,279	3,101	2,877	859	5,131
2014	12,139	3,393	1,431	684	5,142
2015	12,288	3,415	637	248	4,941
2016	11,716	3,230	425	172	4,719
2017	12,729	3,355	359	178	4,271
2018	4,057	2,943	52	45	3,929
Total	187,319	52,538	30,068	10,376	77,541

Panel D: Comparing firm characteristics between SE and Factiva firms

	Mean	SD	Median	Mean	SD	Median
	SE			Factiva		
Total assets	10,918.247	78,383.043	1,083.146	33,685.318***	174,287.968	1,332.433***
Leverage	0.230	0.233	0.167	0.274***	0.258	0.210***
ROA	-0.017	0.345	0.029	-0.043***	0.378	0.018***
Sales growth	0.385	20.440	0.071	0.244	3.593	0.069*
Past return	0.160	0.893	0.084	0.124***	0.701	0.058***
Top 5 institutions	0.280	0.127	0.278	0.231***	0.186	0.234***

Table IA2. Cultural values, seed words, and dictionary

Cultural values				
Innovation	Integrity	Quality	Respect	Teamwork
Guiso, Sapienza, and Zingales (2015) seed words				
creativity, excellence, improvement, passion, pride, leadership, growth, performance, efficiency, results	ethics, accountability, honesty, fairness, responsibility, transparency	customer, commitment, dedication, value, expectations	diversity, inclusion, development, talent, employees, dignity, empowerment	collaboration, cooperation
The dictionary				
execution, innovation, excellence, efficiency, productivity, creativity, fanatical, focus, effectiveness, differentiation, agility, relentlessly, technological_advancement, passion, speed_agility, powerfully, responsiveness, relentless, innovate, competitiveness, virtuous_cycle, improvement, tenacity, adaptability, relentless_pursuit, competence, entrepreneurial_culture, scalability, obsession, uncompromising, blocking_tackling, unrelenting, growth, blocking_tackle, versatility, nimbleness, technological_advance, secret_sauce, raise_bar, imperative, dependability, technologically, vitality, momentum, progress, efficient, resonance, distinctive,	accountability, transparency, ethic, oversight, transparency_accountability, integrity, responsibility_accountability, governance, rigor, utmost, responsibility, continuity, zero_tolerance, seriousness, credibility, sense_urgency, consistency, alignment, moral, humility, assure, autonomy, accountable, thoroughness, hold_accountable, ethical, cohesiveness, urgency, constituent, ethically, ensure, fairness, empathetic, advice, counterproductive, candor, organization, honesty, sympathetic, regulatory_oversight, proper, onus, decision_making, fiduciary_responsibility, interdisciplinary, advocacy, socially_responsible, transparent, paramount_importance,	quality, value, commitment, customer, reputation, loyal, choice, satisfaction, client, committed, commitment, tireless, loyalty, relationship, perseverance, dedication, longevity, stickiness, proposition, willingness, solve_problem, reliable, capable, affordability, dedication_commitment, hardwork, promise, appreciative, uniquely, convince, engagement, necessity, friendliness, ultimately, safe_reliable, demanding, solution, speed, insure, supplier, continually, content, creditworthy, testimonial, credibly, peace_mind, conviction, taste, bargain, hassle_free, hardworking, affordably, vendor, selection, discriminating, flexibility,	talent, skill, employee, talented, talented_workforce, empowerment, highly_skilled, highly_motivate, talented_dedicated, employer_choice, empowered, talent_pool, competency, culturally, incredibly_talented, alumnus, attract_talent[29], expertise, skillset, student_faculty, compassion, skilled_workforce, attract_retain, experienced_talented, technologist, gifted, cadre, talented_experienced, humanity, faculty, generosity, workplace, equality, boot_ground, personify, philanthropy, seasoned_professional, coaching, novice, diversity,	collaborative, cooperative, collaborate, cooperation, teamwork, collegial, working, collaboratively, coordinate, coordination, cooperatively, collaborating, coordinated_effort, collaboration, teaming, unite, collective, engage, involvement, partner, comarketing, joint, cordial, coordinated, symbiotic, jointly, consultative, engaged, professionalism, supportive, involved, multi_disciplinary, confrontational, shared_vision, shoulder_shoulder, mutually, tripartite, rapport, multifaceted, stakeholder, interaction, intimately, intimately_involve, amicable, unyielding,

[29] Words that are frequently used together are treated as a single phrase in our analysis (e.g., attract and retain, depth and breadth). We refer readers to Mikolov et al. (2013) for the method of learning phrases from data.

obsessive, steadfast_commitment, relevance, virtuous_circle, fundamental, transformation, entrepreneurship, intellect, bedrock, differentiated, core_competency, storytelling, stride, strategy, executional, craftsmanship, mantra, foundational, critically_important, distinctiveness, differentiating, authenticity, improve, demonstrating, inspiring, simplicity, brilliantly, tenacious, vision, prowess, driving, differentiator, sophistication, uniqueness, ingenuity, learning, thoughtfulness, efficiently_effectively, innovativeness, superlative, unsurpassed, specialization, innovating, devotion, positioning, relevancy, foundation, scaling, underpinning, innovative, vibrancy, unequaled, obsess, greatness, inspiration, initiative, superb, repeatable, superior, holistic_approach, powerful, demonstrate, skill_set, vigor, key_differentiator, locally_relevant, enabler, passionately, vigilance, mission, efficiently, innate, breakthrough, effectively_efficiently, brand_awareness, mastery, advancement, scaleability,	elect_official, privacy, intricacy, properly, compromise, scrutiny, policymaker, rigger, legitimacy, consciousness, rigor_discipline, insistence, social_responsibility, rest_assure, administratively, culpability, intimately_familiar, ministerial, adherence, bureaucracy, fiduciary_duty, dishonest, timeliness, partisan, completeness, identity, watchdog, safety_soundness, credential, deference, acknowledgement, compliance, ambiguity, individuality, criticism, lawmaker, governmental_authority, professionally, controllership, documentation, impartial, sincerity, clear_concise, hierarchy, legislator, thoughtful, abide, citizenship, practice, cohesive, reputational_risk, supervision, informed, ethics, independence, authoritative, prescriptive, accreditor, responsible, prudence, humbly, testimony, reassurance, reputational, predictability, sacred, prejudice, fraud_abuse, legitimately, recommendation, proactive, codify, dissent, outsider, whistleblower, necessary, philosophy,	covet, bankable, dedicated, uptime, specification, criticality, economically, repeatability, customization, referenceability, stop_shop, insist, commit, determined, honor, bandwidth, maniacally_focused, workmanship, durability, conductivity, commendation, hungry, affordable, referenceable, gratitude, afford, engender, bona_fide, choose, density, accessibility, safely, zero_defect, score, dedicated_talented, eager, equipped, marketer, financially_viable, remain_steadfast, maintainability, desirability, deepen_relationship, economical, testament, scale, proficiency, serve, quantity, advertiser, sticky, heartfelt, sincerely, protect, offer, means, smile, owner, precious, customize, motivator, flexible, accolade, compatibility, satisfaction_score, resistance, immersive_experience, dissatisfaction, curate, save_money, lifetime, image, propensity, tradition, compliment, convenient, return, priority, referencable, spec, style, easily_accessible, seeker,	recruit, highly_respected, faculty_staff, highly_qualified, genuinely, educated, workforce, experienced, recruiting, topnotch, motivating, dignity, employer, seniority, cultural_fit, consultant, educator, skilled, attract, salespeople, worker, compassionate, deserving, caliber, employee_morale, demonstrated, job_seeker, socially, inculcate, training, academically, institutionalized, businessman, retrain, closeness, ingrained, fortitude, recruiter, deserve, ambassador, accomplished, student_teacher, teaching, attentive, engrain, counseling, unmatched, recruiting_training, capability, practitioner, stature, institutionalize, intern, teach, freelance, individually_collectively, stewardship, geographically_disperse, salesperson, dexterity, teacher, deep_domain, nationality, instructor, armed_force, recruiting_retention, tenured, diverse, utmost_confidence, attitude, cross_pollinate, bilingual, ingrain, honor_privilege, incentiviz,	harmonious, outreach, ngos, educate, consultive, cultivate, orchestrate, pitching, organize, cohesively, organized, applaud, multidisciplinary, multi_faceted, informal, fruitful, formal_informal, spirited, highly_engaged, collaborator, embrace, unison, consultative_selling, interact, aligned, longstand, culture, unified, constituency, coalition, longstanding, comarket, hand_glove, resourcefulness, consortium, gatekeeper, teammate, spearhead, heavily_involved, partnering, mentoring, handshake, nurture, deeply_embed, partnership, negotiator, world_renowned, team, trusted, corporative, renowned, academic_institution, deeply_involved, longstanding_relationship, consortia, assist, multi_facet, highly_regarded, knowledgeable, camaraderie, socialize, respectful, participatory, tireless_effort, jointly_develop, council, dialog, communicative,

enduring, consistently, passionate, demonstrable, innovate_elevate, enhance, smarter, absolutely_convinced, pioneering, big_believer, lesson_learn, critical_mass, push_boundary, unsurpass, core, resourceful, scalable, excitement, virtuous, enthusiasm, superbly, enable, aptitude, evolution, ergonomic, exceptional, unparalleled, elegance, imaginative, transform, freshness, sensibility, execute, proven, unwavering, word_mouth, core_competence, acuman, reproducible, permeate, uncompromised, perfection, entrepreneurial, develop, breadth, engaging, technological_breakthrough, sharpen, tremendously, brilliance, imagination, exquisite, push_envelope, advantage, personalization, amaze, motivational, globalization, containment, smoothness, reinvent, excel, optimize, fundamentally, create, irresistible, empower, streamlined, energize, highly_differentiate, hunger, simplification, newness, ethos, emotional_connection, relentlessly_pursue, traction, provenance, miniaturization, curation, reinforcement, hone, unbeatable, enhancement, awareness, esthetic, energetic,	fiduciary, forcefully, inappropriately, documented, informed_decision, drafting, intellectually, referee, supervise, trustworthy, harm, discipline, regulator, accountant, accuracy, external_auditor, disciplinary, guide_principle, comprehensively, burdensome, auditing, appropriately, procedurally, supervisory, humanitarian, practicality, principle, activist, checklist, comply_regulation, nonsense, transparently, credible, candid, unambiguous, abundantly_clear, suggestion, careful_thoughtful, unintended_consequence, critical, attempt, testify, urgently, organizational_structure, bureaucratic, ensure_continuity, persuasive, thoroughly, accreditation, traceability, unnecessary, reporting, public_affair, lobbyist, centrally, misconduct, overrule, overcome_obstacle, seriously, disagreement, determination, morally, improper, complacency, critique, constitutionally, anonymous, administrator, external_consultant, delicate, assurance, misuse, suitability, arrogance, auditor, politician,	designer, convenience, evidence, elegant, purity, gouge, exacting, digitally, universally, safely_reliably, delivering, reliably, indispensible, personalize, parental, user, imagery, clientele, measurability, satisfy, consumer, retention, deliverability, uniformity, network, notch, standardization, waver, latency, efficacy, timely_manner, proactively, rating, supporter, viewer, flawless, legendary, creator, provider, sufficiently, audience, serviceability, greet, patience, validate, receptivity, creditworthiness, authenticate, aspire, interoperable, database, demand, tailor, secure, shipper, betterment, intent, buyer, prospects, feature, assortment, feature_functionality, goal, delivery, centricity, endeavor, requirement, arsenal, affinity, proprietary, durable, merchant, engineer, uninterrupted, deliverable, profile, installer, medal, objective, lean, rigorous, eye_ball, winning, qualified, premier, availability, control, vender, reception, premiere, meticulous, innerworking, showcase,	emphasize_importance, spouse, hourly_employee, decentralization, multilingual, diversification, wholeheartedly, generalist, technician, counselor, resident, learner, intellectual, dedicate, curricula, individualized, discriminate, complimentary, spotlight, overarch, commonality, incentivize, internship, luminary, indoctrinate, possess, inexperienced, unbiased, deepen, graduate, differentiate, prospective_student, track_record, attuned, message_resonate, spiritual, youthful, seasoned, man_woman, reassure, salesman, trainer, rehire, character, hire, unionized, acclimate, retraining, appealing, condolence, curricular, employer_sponsor, religion, heightened_awareness, geographic_diversity, praise, attracted, complementary, logical_extension, childcare, veteran, egg_basket, recruitment, wellbeing, transferable, concierge, geographic_footprint, broaden, degrees, supervisor, amenity, researcher, merit,	contentious, intertwine, integrative, nonprofit, intimacy, lobbying, constructively, holistic, cosponsor, sponsor, competent, respected, investigator, coordinating, reputable, dialogue, teams, enlist, cultural, earnestly, implementer, grassroots, mutually_acceptable, negotiating, endorse, consensual, decentralized, instill, communicating, interdependent, academic_collaborator, unwavering_commitment, consultancy, businesspeople, integrator, acknowledgment, commend, consultation, cultivation, keen, influencer, warmly, independently, decentralize, responsive, facet, professionalize, unifying, invigorate, university_college, broaden_deepen, influential, worker_council, cooperate, codevelopment, friendship, multifacet, reenergized, referral_source, constructive, concert, interoperability, unwaver, acquaint, consort, faithful, multimodal, symbiotic_relationship, alliance, enlightening,

glamorous, crack_code, exemplary, creative, envy, nurturing, functional, proof_pudding, resonate, science, zeal, believer, outperformance, journey, cleanliness, hallmark, proof, unrivaled, genuine, moore_law, environmentally, inspirational, superiority, complexity, headway, thirst, remain_committed, storyline, nimble, tirelessly, architecturally, extendibility, mass_customization, recipe, courage, technology, incredible, enormously, underscore_importance, memorable, incredibly, technological, incrementality, sacrifice, key_enabler, ambition, cutting_edge, reengineer, creatively, globalize, skillful, leading_edge, artificial_intelligence, achievement, visibly, infrastructure, emotional, tangibly, heart_soul, selectivity, heroic, platform, foundational_element, breeding, sustainably, optimized, originality, sensory, thinker, design, halo_effect, usability, effortless, merchandising, invasion, modularity, playbook, aesthetic, rewarding, highly_scalable, dominance, environmentally_friendly,	objectively, delegate, committees, overreach, profession, discomfort, causation, lawyer, thorough, constitutional, forthright, shortcut, organizational, bipartisan, edict, profess, organizationally, unethical, inadvertent, reliability, distraction, subcommittee, appropriateness, act_responsibly, veracity, gracious, disservice, rightfully, bylaw, impatience, espouse, unanimous, regulatory_affair, citizen, unequivocal, purview, frustrate, frown, prejudge, distract, actionable, meticulously, abundance_caution, paperwork, wrongdoing, crisper, boards, requisite, superintendent, doctrine, hassle, complacent, harassment, misrepresent, commissioners, accuracy_completeness, advising, statistician, preach, uphold, conscientious, inconsistency, timely_fashion, validity, supermajority, functioning, safeguard, inappropriate, govern_body, practical, flaw, adviser, questionnaire, pleading, persuade, preparedness, inadvertently, preauthorization, legality, precaution, delicate_balance,	operationally, quotient, skills, interface, elite, providing, lifeline, operational, checkpoint, inception, irrevocable, reengagement, joining, underscores, measurable, groundbreak, deluxe, directed, preexisting, nist, kaizen, vanguard, vignette, relationships, consumers, valence, expectations	fellowship, assimilate, resiliency, beloved, clienteling, geographical_diversity, members, student, conflict, distinguish, retiree, multinational, breadth_diversity, knack, resilience, diversify, endure, coalesce, scheduler, confer, atmosphere, freedom, geographically_diverse, dispatcher, acknowledge, socioeconomic, impart, matters, desirable, coaches, geographic_diversification, perceive, residency, parochial, nonunion, curriculum, interconnectivity, overemphasize, attractiveness, solidify, wide_variety, classroom, geographic_dispersion, complementarity, societal, doctorate, highly_complementary, relevant, professor, multitude, equitably, footprint, structure, segmentation, tutoring, hiring, pursuit, multiplicity, established, vested, applicability, literacy, cares, accordant, grants, varied, choices, acceptability, subspecialty, virtue, diploma, scholarship, strength, prospective, situational, robustness,	lifelong, enablement, alongside, unification, gratified, conversational, nonexclusive, receptive, harmonize, firsthand, codevelop, concerted, interagency, shared, merging, partnerships, tribal, solicit, multiproduct, harmonized, integral, meritocracy, siloed, university, procurement, conceptualization, federation, seamlessly, centerpiece, privileged, informally, longtime, intertwined, socialization, memorialize, scholarly, conceive, referral, dovetail_nicely, dispute_resolution, backoffice, reengage, teamsters, culmination, ministry, keenly_interested, grassroot, intimate, teammates, humane, seamlessly_integrate, seamless_integration, scientific_advice, multilevel, educational_institution, immerse, diplomatic, civic, tightly_integrate, reviewer, orchestrated, expansive, sponsorship, credentialing, social, systematic, multiplatform, quest, multisite, collegiate, multistate, blessed,

unstoppable, invention, relentlessly_focused, exemplify, approachable, concerted_effort, manufacturability, modernize, knowhow, encouragement, proficient, analytical_tool, scaleable, optimization, decisiveness, synergy, productively, revitalize, authentic, simplify, road_map, mover_advantage, analytic, admiration, expansion, powerhouse, unleash, streamline, heighten_awareness, newfound, interactivity, harness, comprehensiveness, openness, unquestionable, excitement_enthusiasm, paradigm_shift, attentiveness, incredibly_powerful, accomplishment, transcend, motivate, styling, excellency, preeminence, personality, attraction, extensibility, innovator, widespread_adoption, roadmap, immersive, reenergiz, creativeness, rediscover, thrive, precision, theme, inspire, ambience, multichannel, localization, nourish, revolutionize, reinvigoration, sustained, digitization, embody, broaden_appeal, reimagine, cornerstone, authorship, functionality, motivated, revitalization, metric, charismatic, agile, unrivalled,	counsels, humanly_possible, adhere, comparative_effectiveness, oath, eliminate_unnecessary, committee, formalization, machination, heed, enforce, rigorously, systematic_approach, confidentially, environmentalist, vetting, complaint, workflow, discretion, caretaker, substantiation, unequivocally, judiciary, strict, singularly_focused, violate, corrupt, legally, underwriter, record_keeping, mediocrity, rules, litigator, prerequisite, bogus, esteemed, indictment, conscience, commonsense, disinterested, honorable, prerogative, commissioner, tenet, probation, examiner, unanimously, govern, sovereignty, secrecy, solidarity, organizational_effectiveness, policy, democratic, deliberation, grievance, enforcement, affirmatively, eligibility, baseless, confidentiality, criminal, hierarchical, delegation, sustainability, remuneration, corruption, representation, fraud_prevention, concurrence, plea, comply, endorsement, affirmative, travesty, strict_adherence, finality, vehemently, affair,		impressions, receptiveness, affirmation, workable, regardless, externship, indigenous, coursework, disband, schools, involve, courseware, varsity, endanger, mindful, geographically_diversify, reshape, differentiatable, ideological, presbyterian, homogeneity, preserve, respective, closely_aligned, override, doctoral, celebrated, broadness, beholden, establishment, leveragability, geoscience, overarching, reassignment, relation, undergraduate, preexist, discrimination, establishing, reside, correctional, balanced, instructional, validation, broadening, baccalaureate, sanctity, geographic, handyman, sharing, redefinition, disperse, demonstration, membership, governmental, realignment, adaptation, inheritance, inherent, implementation, maturation, reunion, encompas, essentials, bachelors, undergrad, surveyor, developmental, representative, overly_dependent, friends, development, underpin, respect, exploitation, enchanted, commercialization, players,	honored, nondisclosure, postmarketing, kudos, regionalization, consult, coworker, conceptual, harmoniously, copromotion, multiregional, revitalized, enterprisewide, scholar, expanded, licensing, multipronged, morale, participative, synergistic, selecting, laudable, resourcing, congruent, alliances, orchestration, teamster, verticalization, alumni, synergize, academia, consulting, cohesion, distributional, outsourcing, multicountry, resonant, modularization, offshoring, interactive, synergetic, fledgling, transnational, externalization, seamlessness, interwoven, attune, troika

automation, reinvention, refocus, uplifting, symbolize, decisive, motivation, tailoring, ingenious, enhanced, creation, refreshing, forefront, responsibly, analytical, thrilling, visionary, ecosystem, solutioning, constantly_innovate, integration, dynamism, evangelize, rapid_prototyping, ruthless, inspired, driven, fluidity, bolder, spirit, pride, premiumization, adoption, lifestyle, enlightened, proliferation, reinvigorate, industrialization, prominence, informatic, energized, champion, incredibly_valuable, megatrend, defensibility, creditability, stamina, personalized, unparallel, revolutionary, visualization, diversified, multi_dimensional, appearance, poise, immeasurable, wellness, seamless, disruptive, platforming, rejuvenate, mobility, convergence, experimentation, eagerness, livelihood, invent, rejuvenation, redesign, visual, digitalization, streamlining, complement, ideation, ignite, crave, enhancing, brand_ambassador, robotic_automation, reenergize, inventive, reimagin, transformational, preeminent, savvy, readiness, enrich, specialized, novel, novelty,	standards, dissemination, separateness, distrust, equitable, contention, adjudication, bylaws, soundness, guideline, dissident, negligent, centralize, strictest, opposition, revoke, purported, handbook, violation, promulgate, preventative, reproducibility, prudency, crime, compliancy, directive, correctness, overt, decisionmaking, activism, guardrail, confidential, preside, preventive, signatory, vulnerability, alleged, courtroom, policyholder, inspector, guidelines, standardize, councils, ego, revocation, sexual, parliamentary, administration, disciplined, stringent, assured, bulletproof, custodian, unqualified, occupational, investigations, non_negotiable, formalize, sincerest, procedures, verifies, reappoint, environmentally_responsible, innocent, symmetry, standard, underwriting, hotline, retain, vow, practices, audit, roundtable, formalized, mystery_shop, overwrite, organizing, solidity, recordkeeping, laud, egovernment, coding, interrogation, law_enforcement,		sponsored, commercialize, ambassadors, dependent, exclusion, inclusion	

scorecard, social_medium, emblematic, exciting, dashboard, evolutionary, enthusiastically, evolved, patented, connectivity, showplace, invigorated, acclaim, supercharge, undisputed, regeneration, maximization, augmented_reality, foresight, discerning, bankability, evolving, streamlines, forge, scientist, suite, paramount, artistic, acumen, transformative, zest, reenforce, swift, pioneer, wholistic, pioneers, reborn, groundbreaking, specialness, restrengthen, transforming, capstone, solutions, initiatives, concepts	commander, protection, discredit, police, quorum, plaintiff, motto, counterparty, taskforce, rights, sanitation, sympathy, salute, permanence, entrust, accountants, indemnify, reviewing, reassume, restitution, formative, incentivization, operability, oversight_committee, indemnity, reputed, malice, renumeration, affairs, allegiance, incentivise, fidelity, duly, consistence, directorship, articles, cardinal, unstated, inspectors			

Table IA3
Validating our measures of corporate cultural values: Including all five values

This table validates our measures of corporate cultural values. We extend Table 3 by including all five culture values in each regression. In Panel A, LnPatent, R&D spending, and innovation strength are used to validate the cultural value of innovation. In Panel B, restatement and backdating are used to validate the cultural value of integrity. In Panel C, product quality, product safety, and top brand are used to validate the cultural value of quality. In Panel D, diversity is used to validate the cultural value of respect. In Panel E, employee involvement and best employer are used to validate the cultural value of teamwork. OLS regressions are used when the dependent variables are LnPatent, and R&D spending, and probit regressions are used for all other validating variables. Definitions of the variables are provided in Appendix B. Heteroskedasticity-consistent standard errors (in parentheses) are clustered at the firm level. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

Panel A: Validating the cultural value of innovation

Variable	LnPatent (1)	LnPatent (2)	LnPatent (3)	R&D spending (4)	R&D spending (5)	R&D spending (6)	Innovation strength (7)	Innovation strength (8)	Innovation strength (9)
Innovation	0.278***	0.269***	0.156***	-0.005**	0.012***	0.004**	0.366***	0.322***	0.158
	(12.54)	(12.51)	(7.76)	(-2.49)	(7.70)	(2.41)	(3.90)	(3.62)	(1.32)
Integrity	-0.440***	-0.467***	-0.256***	0.018**	0.003	-0.003	-0.283	-0.430	0.024
	(-10.61)	(-10.81)	(-6.20)	(2.28)	(0.54)	(-0.47)	(-0.85)	(-1.20)	(0.05)
Quality	0.019	0.032*	0.027	-0.000	-0.002	0.005***	0.112	0.157	0.283**
	(1.08)	(1.83)	(1.49)	(-0.15)	(-1.25)	(3.00)	(1.08)	(1.45)	(2.14)
Respect	-0.233***	-0.224***	-0.149***	0.020***	0.009**	0.014***	-0.076	-0.007	0.213
	(-7.63)	(-7.41)	(-4.80)	(3.39)	(2.16)	(3.44)	(-0.43)	(-0.04)	(0.82)
Teamwork	-0.082*	0.001	0.083*	0.209***	0.080***	0.064***	-0.330	-0.089	-0.138
	(-1.94)	(0.02)	(1.87)	(18.89)	(11.90)	(10.96)	(-0.99)	(-0.26)	(-0.35)
Size	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
ROA	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Ind FE/Yr FE	No	No	Yes	No	No	Yes	No	No	Yes
Intercept	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	25,667	25,667	25,667	57,125	57,125	57,125	12,495	12,495	12,495
R^2 /Pseudo R^2	0.055	0.060	0.243	0.124	0.470	0.563	0.039	0.055	0.152

Panel B: Validating the cultural value of integrity

Variable	Restatement (1)	Restatement (2)	Restatement (3)	Backdating (4)	Backdating (5)	Backdating (6)
Innovation	-0.043**	-0.045**	-0.055**	0.080**	0.084**	-0.008
	(-2.08)	(-2.15)	(-2.38)	(2.09)	(2.19)	(-0.19)
Integrity	-0.120*	-0.115*	-0.078	-0.268**	-0.192*	-0.389***
	(-1.76)	(-1.69)	(-1.06)	(-2.48)	(-1.77)	(-3.28)
Quality	0.062***	0.062***	0.015	-0.085**	-0.105***	-0.098**
	(3.07)	(3.04)	(0.65)	(-2.37)	(-2.91)	(-2.41)
Respect	-0.146***	-0.145***	-0.157***	-0.038	-0.053	-0.010
	(-2.76)	(-2.75)	(-3.01)	(-0.47)	(-0.66)	(-0.11)

Teamwork	-0.083	-0.067	0.024	-0.052	-0.081	-0.034
	(-1.28)	(-0.97)	(0.35)	(-0.50)	(-0.75)	(-0.29)
Size	No	Yes	Yes	No	Yes	Yes
ROA	No	Yes	Yes	No	Yes	Yes
Ind FE/Yr FE	No	No	Yes	No	No	Yes
Intercept	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	51,461	51,461	51,461	18,122	18,122	18,122
Pseudo R^2	0.002	0.002	0.029	0.003	0.008	0.064

Panel C: Validating the cultural value of quality

Variable	Product quality (1)	Product quality (2)	Product quality (3)	Product safety (4)	Product safety (5)	Product safety (6)	Top brand (7)	Top brand (8)	Top brand (9)
Innovation	0.410***	0.337***	0.074	-0.263***	-0.133***	0.034	0.388***	0.180***	0.086
	(9.82)	(7.91)	(1.27)	(-5.18)	(-2.58)	(0.52)	(9.81)	(3.01)	(1.19)
Integrity	-0.013	-0.321*	0.185	-0.304**	0.145	-0.576**	0.776***	-0.684***	-0.238
	(-0.09)	(-1.81)	(0.81)	(-2.15)	(0.86)	(-2.37)	(5.99)	(-2.90)	(-0.92)
Quality	-0.208***	-0.179***	-0.049	0.320***	0.331***	0.211**	0.114**	0.315***	0.191**
	(-4.26)	(-3.27)	(-0.69)	(5.19)	(4.87)	(2.51)	(2.43)	(4.56)	(2.49)
Respect	-0.173*	-0.076	0.108	0.137	-0.021	-0.033	-0.228*	-0.048	0.136
	(-1.66)	(-0.69)	(0.77)	(1.00)	(-0.14)	(-0.17)	(-1.94)	(-0.27)	(0.69)
Teamwork	-0.427***	0.164	-0.269	0.655***	-0.034	0.187	-0.781***	0.392	0.656**
	(-2.73)	(0.96)	(-0.98)	(3.65)	(-0.16)	(0.81)	(-4.80)	(1.56)	(2.47)
Size	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
ROA	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Ind FE/Yr FE	No	No	Yes	No	No	Yes	No	No	Yes
Intercept	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	19,596	19,596	19,596	22,705	22,705	22,705	43,019	43,019	43,019
Pseudo R^2	0.026	0.100	0.280	0.021	0.126	0.274	0.053	0.450	0.492

Panel D: Validating the cultural value of respect

Variable	Diversity (1)	Diversity (2)	Diversity (3)
Innovation	0.408***	0.297***	0.207***
	(8.31)	(7.67)	(5.52)
Integrity	0.709***	0.260***	0.162
	(6.05)	(2.68)	(1.64)
Quality	-0.114***	-0.022	0.055
	(-2.97)	(-0.66)	(1.62)
Respect	-0.218**	-0.035	-0.032
	(-2.42)	(-0.44)	(-0.43)
Teamwork	-0.445***	0.323***	0.317***
	(-4.38)	(3.46)	(3.55)
Size	No	Yes	Yes

ROA	No	Yes	Yes
Ind FE/Yr FE	No	No	Yes
Intercept	Yes	Yes	Yes
Obs.	20,720	20,720	20,720
Pseudo R^2	0.031	0.181	0.342

Panel E: Validating the cultural value of teamwork

Variable	Employee involvement (1)	Employee involvement (2)	Employee involvement (3)	Best employer (4)	Best employer (5)	Best employer (6)
Innovation	0.239***	0.153***	-0.055	0.250***	0.084	0.018
	(5.55)	(3.55)	(-1.08)	(4.83)	(1.34)	(0.22)
Integrity	-0.240	-0.561***	-0.272	-0.296	-0.816***	-0.756***
	(-1.56)	(-3.30)	(-1.50)	(-1.30)	(-2.87)	(-2.79)
Quality	0.178***	0.237***	0.238***	0.038	0.109	0.150
	(4.17)	(5.23)	(4.46)	(0.66)	(1.61)	(1.63)
Respect	-0.248**	-0.096	-0.037	-0.017	0.163	0.230
	(-2.34)	(-0.86)	(-0.30)	(-0.11)	(0.87)	(1.08)
Teamwork	0.442***	0.833***	0.794***	-0.091	0.568***	0.428**
	(3.41)	(5.80)	(5.23)	(-0.59)	(3.44)	(2.03)
Size	No	Yes	Yes	No	Yes	Yes
ROA	No	Yes	Yes	No	Yes	Yes
Ind FE/Yr FE	No	No	Yes	No	No	Yes
Intercept	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	18,431	18,431	18,050	55,428	55,428	46,276
Pseudo R^2	0.034	0.079	0.143	0.021	0.160	0.256

Table IA4
Horse race between our main and alternative corporate culture measures

This table compares our main culture measures with alternative culture measures based on: i) the entire call (_full); 2) a simple count of the seed words (including the value word) in the QA section of calls (_seed); and iii) applying the word embedding model to the MD&A section of 10-Ks (_10k). In Panel A, LnPatent, R&D spending, and innovation strength are used to validate the cultural value of innovation. In Panel B, restatement and backdating are used to validate the cultural value of integrity. In Panel C, product quality, product safety, and top brand are used to validate the cultural value of quality. In Panel D, diversity is used to validate the cultural value of respect. In Panel E, employee involvement and best employer are used to validate the cultural value of teamwork. OLS regressions are used when the dependent variables are LnPatent, and R&D spending, and probit regressions are used for all other validating variables. Definitions of the variables are provided in Appendix B. Heteroskedasticity-consistent standard errors (in parentheses) are clustered at the firm level. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

Panel A: Horse race between innovation, innovation_full, innovation_seed and innovation_10k

Variable	LnPatent (1)	LnPatent (2)	LnPatent (3)	R&D spending (4)	R&D spending (5)	R&D spending (6)	Innovation strength (7)	Innovation strength (8)	Innovation strength (9)
Innovation	0.236***	0.266***	0.140***	0.046***	0.034***	0.028***	0.122	0.201	0.124
	(5.75)	(6.47)	(4.11)	(10.00)	(10.29)	(8.75)	(0.71)	(1.22)	(0.69)
Innovation_full	0.063**	0.083***	0.074***	0.019***	-0.003	-0.004*	0.096	0.071	0.020
	(2.23)	(2.94)	(2.94)	(6.44)	(-1.20)	(-1.95)	(0.67)	(0.51)	(0.13)
Innovation_seed	0.028	-0.107**	-0.072	-0.149***	-0.035***	-0.037***	0.404**	0.230	0.300
	(0.54)	(-2.03)	(-1.54)	(-19.73)	(-7.50)	(-8.64)	(1.98)	(1.29)	(1.24)
Innovation_10k	-0.036	-0.064	-0.087**	-0.050***	-0.019***	-0.011***	0.216	0.172	0.137
	(-0.73)	(-1.30)	(-2.01)	(-13.12)	(-6.60)	(-4.32)	(0.99)	(0.78)	(0.56)
Size	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
ROA	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Ind FE/Yr FE	No	No	Yes	No	No	Yes	No	No	Yes
Intercept	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	19,836	19,836	19,836	43,817	43,817	43,817	11,037	11,037	11,037
R^2/Pseudo R^2	0.049	0.062	0.288	0.095	0.471	0.567	0.040	0.052	0.151

Panel B: Horse race between integrity, integrity_full, integrity_seed and integrity_10k

Variable	Restatement (1)	Restatement (2)	Restatement (3)	Backdating (4)	Backdating (5)	Backdating (6)
Integrity	-0.374***	-0.390***	-0.263**	-0.329	-0.362	-0.191
	(-2.97)	(-3.09)	(-2.06)	(-0.87)	(-0.93)	(-0.44)
Integrity_full	0.221*	0.259**	0.246**	0.573*	0.614*	0.366
	(1.94)	(2.26)	(2.09)	(1.87)	(1.95)	(0.98)
Integrity_seed	0.023	-0.112	-0.422	-4.419	-4.698	-5.012
	(0.02)	(-0.11)	(-0.41)	(-1.12)	(-1.14)	(-1.18)
Integrity_10k	-0.062	-0.092	-0.042	0.014	-0.037	0.012
	(-0.58)	(-0.87)	(-0.38)	(0.05)	(-0.14)	(0.04)
Size	No	Yes	Yes	No	Yes	Yes

ROA	No	Yes	Yes	No	Yes	Yes
Ind FE/Yr FE	No	No	Yes	No	No	Yes
Intercept	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	39,191	39,191	39,191	13,895	13,895	13,895
Pseudo R^2	0.001	0.001	0.031	0.004	0.006	0.129

Panel C: Horse race between quality, quality_full, quality_seed and quality_10k

Variable	Product quality (1)	Product quality (2)	Product quality (3)	Product safety (4)	Product safety (5)	Product safety (6)	Top brand (7)	Top brand (8)	Top brand (9)
Quality	0.226**	0.248**	0.184	0.073	0.095	0.270*	0.342***	0.343**	0.403**
	(2.37)	(2.43)	(1.46)	(0.67)	(0.79)	(1.73)	(3.35)	(2.12)	(2.32)
Quality_full	-0.216***	-0.242***	-0.217**	0.053	0.104	-0.016	0.063	0.097	0.065
	(-2.71)	(-2.94)	(-2.17)	(0.52)	(1.00)	(-0.14)	(0.70)	(0.72)	(0.45)
Quality_seed	-0.170	0.017	0.038	0.397***	0.213	0.109	-0.410***	0.065	-0.345
	(-1.41)	(0.13)	(0.22)	(2.62)	(1.26)	(0.52)	(-3.17)	(0.34)	(-1.55)
Quality_10k	0.001	-0.037	-0.005	0.033	0.075	0.075	0.024	-0.037	-0.009
	(0.02)	(-1.00)	(-0.10)	(0.63)	(1.36)	(1.36)	(0.39)	(-0.48)	(-0.12)
Size	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
ROA	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Ind FE/Yr FE	No	No	Yes	No	No	Yes	No	No	Yes
Intercept	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	17,556	17,556	17,556	20,264	20,264	20,264	31,555	31,555	31,555
Pseudo R^2	0.003	0.084	0.285	0.014	0.123	0.270	0.024	0.407	0.488

Panel D: Horse race between respect, respect_full, respect_seed and respect_10k

Variable	Diversity (1)	Diversity (2)	Diversity (3)
Respect	0.175	-0.234*	-0.122
	(1.26)	(-1.79)	(-1.05)
Respect_full	-0.354***	0.404***	0.346***
	(-2.91)	(3.46)	(3.17)
Respect_seed	1.135***	0.330	-0.252
	(3.52)	(1.21)	(-1.00)
Respect_10k	0.047	0.292***	0.086
	(0.41)	(2.91)	(0.95)
Size	No	Yes	Yes
ROA	No	Yes	Yes
Ind FE/Yr FE	No	No	Yes
Intercept	Yes	Yes	Yes
Obs.	18,376	18,376	18,376
Pseudo R^2	0.004	0.141	0.322

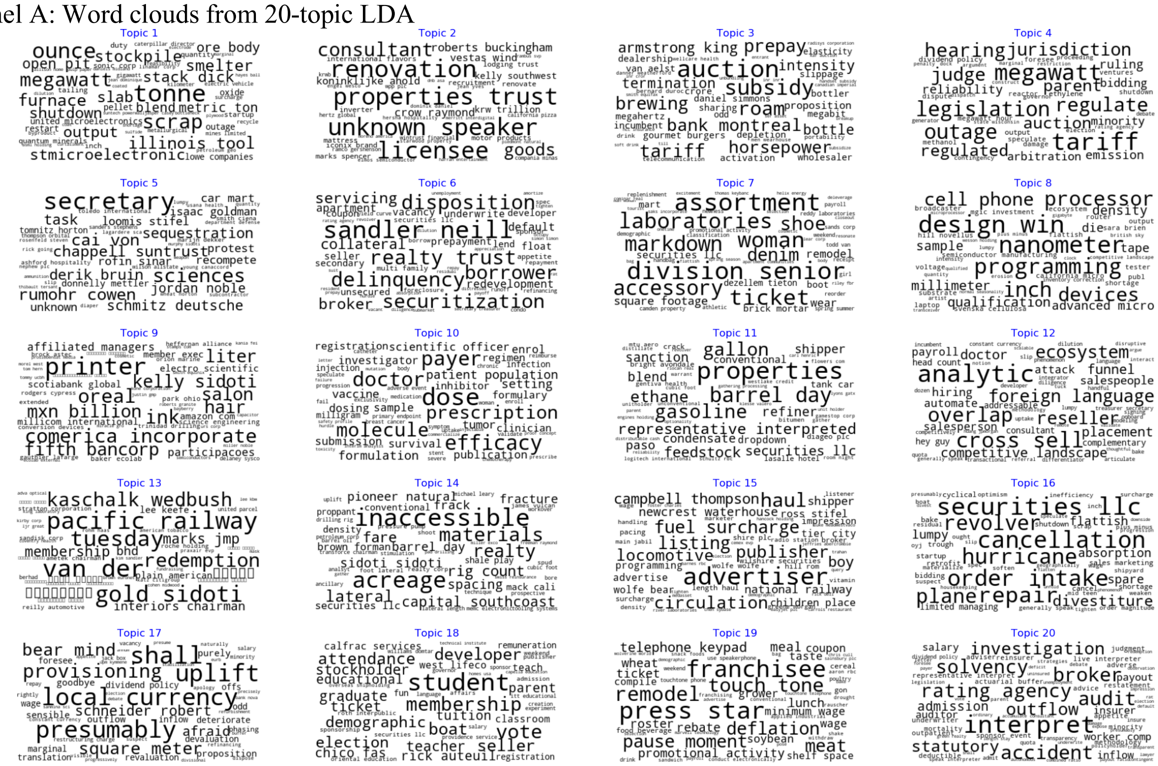
Panel E: Horse race between teamwork, teamwork_full, teamwork_seed and teamwork_10k

Variable	Employee involvement (1)	Employee involvement (2)	Employee involvement (3)	Best employer (4)	Best employer (5)	Best employer (6)
Teamwork	0.808***	0.817***	0.614*	0.424*	0.444	0.038
	(2.85)	(2.83)	(1.93)	(1.69)	(1.30)	(0.09)
Teamwork_full	-0.168	0.163	0.100	-0.319	0.265	0.353
	(-0.64)	(0.62)	(0.35)	(-1.32)	(0.87)	(0.87)
Teamwork_seed	5.816	9.899	13.195*	0.951	2.002	7.306
	(0.92)	(1.49)	(1.88)	(0.16)	(0.26)	(0.88)
Teamwork_10k	-0.188	0.080	0.270	-0.400	-0.296	0.355
	(-0.49)	(0.20)	(0.60)	(-0.54)	(-0.35)	(0.40)
Size	No	Yes	Yes	No	Yes	Yes
ROA	No	Yes	Yes	No	Yes	Yes
Ind FE/Yr FE	No	No	Yes	No	No	Yes
Intercept	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	16,355	16,355	16,355	42,352	42,352	42,352
Pseudo R^2	0.009	0.055	0.142	0.001	0.155	0.276

Table IA5
Results from topic modeling

This table presents the results from LDA, a topic modeling method based on Bayesian statistics, on the QA section of calls. LDA is an alternative machine learning method for extracting features from textual data. Before fitting LDA models, we pre-process the data by removing numerical digits, less frequent words (n < 5) and top 2,000 common words. We fit three different LDA models, with the number of topics being 20, 100, and 200. For each topic, we generate word clouds that shows the top 100 words with the highest probability. Panel A presents the word clouds from a 20-topic LDA model. Panel B presents a randomly chosen 20 word clouds from a 100-topic LDA model. Panel C presents a randomly chosen 20 word clouds from a 200-topic LDA model.

Panel A: Word clouds from 20-topic LDA



Panel B: Randomly selected 20 word clouds from 100-topic LDA

Panel C: Randomly selected 20 word clouds from 200-topic LDA

Table IA6
M&A sample overview

The acquirer sample consists of 7,875 completed deals over the period 2003 to 2017 from the Thomson One Banker SDC database. The sample is formed as the intersection of the Compustat database, Thomson One Banker SDC database, and the earnings call datasets. The pair sample consists of 810 completed deals where both the acquirer and its target firm are public. The sample selection criteria are as follows: 1) the deal is classified as "Acquisition of Assets (AA)", "Acquisition of Majority Interest (AM)," or "Merger (M)" by the data provider; 2) the acquirer is a U.S. public firm listed on the AMEX, NYSE, or NASDAQ; 3) the acquirer holds less than 50% of the shares of the target firm before deal announcement and ends up owning 100% of the shares of the target firm through the deal; 4) the deal value is at least $1 million (in 1995 dollar value); 5) the relative size of the deal (i.e., the ratio of transaction value over book value of acquirer total assets) is at least 1%; 6) the target firm is domiciled in the U.S.; 7) the target firm is a public firm, a private firm, or a subsidiary; 8) multiple deals announced by the same acquirer on the same day are excluded; and 9) basic financial and stock return information is available for the acquirer. 10) culture variables are available for the acquirer (as well as for the target for the pair sample)

Year	Acquirer sample	Pair sample
2003	435	35
2004	569	54
2005	649	70
2006	701	68
2007	659	71
2008	460	48
2009	322	50
2010	497	61
2011	499	32
2012	565	51
2013	542	54
2014	650	62
2015	555	72
2016	453	68
2017	319	14
Total	7,875	810

Table IA7
Summary statistics of the acquirer and pair samples for deal incidence and merger pairing analysis

The acquirer sample consists of 7,875 completed deals over the period 2003 to 2017. The pair sample consists of 734 completed deals where both the acquirer and its target firm are public with available control firms. Panel A presents the summary statistics of acquirers. Panel B presents the summary statistics of the pair sample. Panel C presents correlations between corporate culture variables and acquirer characteristics. Panel D presents correlations between cultural similarity measures and other similarity measures. Definitions of the variables are provided in Appendix B. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

Panel A: Summary statistics of the acquirer sample

Variable	Obs.	Mean	10th Percentile	Median	90th Percentile	SD
Innovation	7,875	1.410	0.686	1.307	2.276	0.628
Integrity	7,875	0.240	0.096	0.204	0.433	0.152
Quality	7,875	1.244	0.614	1.090	2.107	0.607
Respect	7,875	0.445	0.210	0.398	0.728	0.230
Teamwork	7,875	0.217	0.056	0.167	0.444	0.174
Combined CAR(-1, 1)	1,068	2.642	-3.937	1.435	10.92	6.785
BHAR1	7,596	-0.017	-0.598	-0.011	0.556	0.486
ΔROA1	6,752	-2.290	-10.32	-0.523	4.741	10.60
Integration	7,875	0.714	0	1	1	0.452
Retention	7,875	0.338	0	0	1	0.473
Divestiture	7,875	0.026	0	0	0	0.160
Total assets	7,875	4,920	152.2	1,166	10,207	12,377
Leverage	7,875	0.212	0	0.168	0.499	0.196
ROA	7,875	0.036	-0.031	0.040	0.115	0.080
Sales growth	7,875	0.212	-0.069	0.123	0.542	0.402
Past return	7,875	0.229	-0.263	0.160	0.749	0.487
Top 5 institutions	7,875	0.269	0.121	0.273	0.404	0.117
All cash	7,875	0.368	0	0	1	0.482
All stock	7,875	0.050	0	0	0	0.217
Tender offer	7,875	0.024	0	0	0	0.153
Same industry	7,875	0.556	0	1	1	0.497
Same state	7,875	0.221	0	0	1	0.415
Relative size	7,875	0.205	0.017	0.078	0.522	0.351
Private target	7,875	0.514	0	1	1	0.500
Subsidiary target	7,875	0.330	0	0	1	0.470
HP similarity	7,875	0.009	0	0	0	0.041

Panel B: Summary statistics of the pair sample for merger pairing analysis

	Obs.	Mean	10th Percentile	Median	90th Percentile	SD	Mean	10th Percentile	Median	90th Percentile	SD
				Acquirers					Target Firms		
Innovation	734	1.564	0.804	1.499	2.401	0.616	1.487	0.774	1.411	2.282	0.615
Integrity	734	0.254	0.109	0.225	0.432	0.141	0.260	0.098	0.227	0.462	0.160
Quality	734	1.314	0.658	1.180	2.162	0.591	1.367	0.671	1.232	2.289	0.646
Respect	734	0.444	0.230	0.414	0.693	0.192	0.478	0.231	0.447	0.746	0.232
Teamwork	734	0.220	0.066	0.178	0.416	0.168	0.262	0.070	0.197	0.558	0.210
Firm size	734	8.424	6.082	8.394	10.796	1.845	6.573	4.342	6.484	9.017	1.812
Leverage	734	0.209	0.001	0.153	0.507	0.193	0.206	0.000	0.133	0.550	0.225
ROA	734	0.038	-0.020	0.046	0.121	0.110	-0.035	-0.226	0.022	0.106	0.270
Sales growth	734	0.181	-0.078	0.094	0.494	0.460	0.158	-0.149	0.074	0.441	0.696
Past return	734	0.190	-0.297	0.146	0.623	0.530	0.138	-0.470	0.048	0.747	0.707
Top 5 institutions	734	0.270	0.155	0.266	0.381	0.102	0.300	0.167	0.294	0.435	0.114
			Acquirer-Target Firm Pairs								
Cultural similarity	734	0.957	0.903	0.974	0.995	0.053					
Cultural distance	734	0.837	0.316	0.768	1.454	0.463					

Panel C: Correlation between cultural values and firm characteristics of the acquirer sample

	Innovation	Integrity	Quality	Respect	Teamwork	Firm Size	Leverage	ROA	Sales Growth	Past Return	Top 5 Institutions
Innovation	1.000										
Integrity	0.169***	1.000									
Quality	0.486***	0.170***	1.000								
Respect	0.249***	0.244***	0.218***	1.000							
Teamwork	0.340***	0.304***	0.367***	0.345***	1.000						
Firm Size	0.049***	-0.040***	-0.117***	-0.064***	-0.160***	1.000					
Leverage	-0.303***	-0.118***	-0.250***	-0.137***	-0.227***	0.323***	1.000				
ROA	0.006	-0.042***	-0.069***	-0.044***	-0.087***	0.157***	-0.215***	1.000			
Sales Growth	-0.072***	-0.018	-0.049***	-0.004	0.048***	-0.105***	0.016	0.022*	1.000		
Past Return	0.006	-0.027**	0.002	-0.012	0.028**	-0.071***	-0.077***	0.062***	0.079***	1.000	
Top 5 Institutions	0.022*	-0.000	0.026**	0.048***	0.006	-0.075***	-0.023**	-0.010	-0.078***	-0.071***	1.000

Panel D Correlations between culture similarity and other similarity measures of the pair sample

	Cultural similarity	Cultural distance	Same state	HP similarity	Same industry
Cultural similarity	1.000				
Cultural distance	-0.473***	1.000			
Same state	0.010	-0.032	1.000		
HP similarity	0.029	-0.136***	0.186***	1.000	
Same industry	0.063*	-0.125***	0.004	0.182***	1.000

Table IA8
Dominant culture and merger pairing

This table examines the relation between dominant culture fit and merger pairing. The acquirer-target sample consists of 734 completed deals where both the acquirer and its target firms are public with available control firms. The dependent variable is equal to one for the acquirer-target firm pair, and zero for the control firm pairs. The coefficients are estimated from conditional logit models. Definitions of the variables are provided in Appendix B. All specifications based on the matched samples include deal fixed effects. Robust standard errors are reported in the parentheses. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

Variable	Industry, Size (1)	Industry, Size, B/M (2)
Acquirer innovation dominant * Target innovation dominant	0.022	0.056
	(0.210)	(0.208)
Acquirer integrity dominant * Target integrity dominant	0.469**	0.449*
	(0.237)	(0.243)
Acquirer quality dominant * Target quality dominant	0.463**	0.475**
	(0.221)	(0.230)
Acquirer respect dominant * Target respect dominant	-0.062	-0.094
	(0.240)	(0.236)
Acquirer teamwork dominant * Target teamwork dominant	0.433*	0.416*
	(0.231)	(0.228)
Acquirer innovation dominant	0.164	0.097
	(0.108)	(0.107)
Acquirer integrity dominant	-0.258**	-0.176
	(0.119)	(0.126)
Acquirer quality dominant	-0.169	-0.239*
	(0.129)	(0.135)
Acquirer respect dominant	-0.088	-0.069
	(0.127)	(0.131)
Acquirer teamwork dominant	-0.015	-0.031
	(0.137)	(0.139)
Target innovation dominant	-0.114	-0.036
	(0.126)	(0.127)
Target integrity dominant	-0.173	-0.226*
	(0.126)	(0.128)
Target quality dominant	-0.078	-0.149
	(0.117)	(0.122)
Target respect dominant	0.167	0.182
	(0.112)	(0.114)
Target teamwork dominant	0.102	0.103
	(0.117)	(0.119)
Acquirer Characteristics		
Firm size	2.805***	2.653***
	(0.188)	(0.190)
Leverage	-1.136***	-1.013***
	(0.280)	(0.310)

ROA	-0.187	-0.001
	(0.461)	(0.544)
Sales growth	0.433***	0.392***
	(0.122)	(0.141)
Past return	0.186	0.102
	(0.113)	(0.124)
Top 5 institutions	1.056***	1.105***
	(0.340)	(0.373)
Target Characteristics		
Firm size	2.187***	1.894***
	(0.255)	(0.237)
Leverage	-0.145	0.100
	(0.270)	(0.299)
ROA	-0.259	-0.514*
	(0.285)	(0.310)
Sales growth	0.274**	0.238**
	(0.116)	(0.116)
Past return	-0.099	-0.122
	(0.097)	(0.101)
Top 5 institutions	2.088***	2.100***
	(0.334)	(0.352)
Same state	0.898***	0.905***
	(0.133)	(0.141)
HP similarity	27.600***	28.086***
	(1.806)	(1.947)
Deal FE	Yes	Yes
Obs.	7,049	6,799
Pseudo R^2	0.288	0.294

Table IA9
Summary statistics of the pair sample for ex post deal outcome analysis

The pair sample for ex post deal outcome analysis consists of 810 completed deals over the period 2003 to 2017. Panel A presents the summary statistics. Panel B presents correlations between deal outcome variables and cultural fit variables. Definitions of the variables are provided in Appendix B. Heteroskedasticity-consistent standard errors (in parentheses) are clustered at the acquirer level. ***, **, * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

Panel A: Summary statistics of the pair sample for ex post deal outcome analysis

Variable	Obs.	Mean	10th Percentile	Median	90th Percentile	SD
Cultural similarity	810	0.958	0.903	0.974	0.995	0.049
Cultural distance	810	0.842	0.326	0.772	1.487	0.454
Ext-Int conflict	810	0.151	0	0	1	0.358
Ext-Int conflict2	810	0.246	0	0	1	0.431
People focus	810	0.061	0	0	0	0.239
Combined CAR(-1, 1)	810	3.012	-3.793	1.695	11.63	6.596
BHAR1	787	-0.030	-0.572	-0.029	0.455	0.420
ΔROA1	711	-2.715	-10.82	-0.791	4.931	10.14
Integration	810	0.780	0	1	1	0.414
Retention	810	0.484	0	0	1	0.500
Acquirer size	810	8.421	6.141	8.376	10.79	1.789
Acquirer leverage	810	0.207	0.0004	0.149	0.506	0.194
Acquirer ROA	810	0.045	-0.020	0.047	0.118	0.069
Acquirer sales growth	810	0.162	-0.073	0.092	0.471	0.321
Acquirer past return	810	0.179	-0.292	0.150	0.632	0.405
Acquirer top5 institutions	810	0.261	0.144	0.266	0.376	0.110
Target size	810	6.520	4.330	6.434	8.906	1.738
Target leverage	810	0.202	0	0.127	0.548	0.224
Target ROA	810	-0.022	-0.226	0.023	0.107	0.176
Target sales growth	810	0.132	-0.140	0.075	0.437	0.315
Target past return	810	0.122	-0.462	0.042	0.760	0.546
Target top5 institutions	810	0.297	0.155	0.292	0.438	0.122
All cash	810	0.451	0	0	1	0.498
All stock	810	0.181	0	0	1	0.386
Tender offer	810	0.190	0	0	1	0.393

Same industry	810	0.686	0	1	1	0.464
Relative size	810	0.519	0.029	0.227	1.290	0.942
Same state	810	0.254	0	0	1	0.436
HP similarity	810	0.060	0	0.037	0.148	0.083

Panel B: Correlation between deal outcome variables and cultural fit variables

	Combined CAR(-1,1)	BHAR1	ΔROA1	Integration	Retention	Culture similarity	Culture distance	Ext-Int conflict	Ext-Int conflict2	People focus
Combined CAR(-1,1)	1.000									
BHAR1	-0.022	1.000								
ΔROA1	-0.052	0.241***	1.000							
Integration	0.066*	-0.022	-0.026	1.000						
Retention	0.056	-0.020	-0.014	0.265***	1.000					
Culture similarity	0.066*	0.052	-0.043	-0.005	0.051	1.000				
Culture distance	-0.105***	0.064*	0.007	0.001	-0.045	-0.471***	1.000			
Ext-Int conflict	-0.064*	0.002	0.006	0.068*	0.044	-0.083**	0.226***	1.000		
Ext-Int conflict2	-0.108***	-0.008	-0.007	0.075**	0.012	-0.106***	0.227***	0.730***	1.000	
People focus	0.081**	-0.059	0.036	-0.041	-0.001	-0.077**	-0.103***	-0.096**	-0.132***	1.000