Duality in Diversity: Cultural Heterogeneity, Language, and Firm Performance *

Matthew Corritore¹, Amir Goldberg¹, and Sameer B. Srivastava²

¹Stanford Graduate School of Business
²Haas School of Business, University of California, Berkeley

WORKING PAPER
July, 2017

Abstract
This article deepens our understanding of how the culture of an organization can reflect its underlying capacity for execution and creative exploration and thereby foreshadow how it will perform in the future. Existing literature often understands cultural diversity as presenting a trade-off between task coordination and creative problem-solving. In contrast, we conceptually unpack cultural heterogeneity into two distinct forms: compositional and content-based. We propose that the former undermines coordination and therefore portends worsening firm profitability, while the latter facilitates creativity and therefore predicts higher market expectations of future growth. To evaluate these propositions, we use unsupervised learning to identify cultural content in employee reviews of nearly 500 publicly traded firms on the Glassdoor website and then develop novel, time-varying measures of cultural heterogeneity. Using coarsened exact matching to reduce imbalance between firms exhibiting higher and lower levels of compositional and content-based heterogeneity, we find support for our two core propositions.

*We thank Glassdoor for generously providing us with the data for this study. We also thank seminar participants at University of Illinois School of Labor and Employment Relations, University of Chicago, Booth School of Business, IESE Business School, the MORS seminar at Kellogg School of Management, Northwestern University, the marketing seminar at the Arison School of Business at IDC Herzliya, Israel, and participants at the Lugano Organizations Conference, the Stanford/Berkeley Organizational Behavior Student Conference, the Consortium on Competitiveness and Cooperation Conference for Doctoral Student Research at the Wharton School, University of Pennsylvania, the Carnegie School of Organizational Learning Conference, and the International Conference on Computational Social Science.
Whether deliberatively cultivated or naturally arising, every organization develops a culture—a system of meanings and norms shared by its members. The culture of a firm can have important consequences for the success of its members and the organization as a whole through its effects on individual motivation and commitment, interpersonal coordination, and group creativity and innovation (Chatman and O’Reilly, 2016). Although organizational scholars often ask how the content of organizational culture relates to performance—for example, how different beliefs and norms promote or inhibit various outcomes—a growing literature has been focused on the effects of cultural heterogeneity on organizations’ productivity and vitality. Research in this vein asks: When is a diversity of ideas and beliefs conducive to organizational success and when is it instead detrimental?

Different literatures have provided varied and inconsistent answers to this question. For example, research on the strength of organizational cultures emphasizes the importance of cultural agreement between organizational members. According to this line of work, incompatibilities in employees’ beliefs and normative expectations can impede their ability to coordinate tasks (Denison and Mishra, 1995; Weber and Camerer, 2003; Kotter and Heskett, 1992), thereby producing a negative relationship between cultural heterogeneity and firm performance. In contrast, the literature on diversity in work groups and organizations acknowledges the negative consequences of disagreement but also identifies conditions under which it can serve as a source of advantage, for example in promoting collective learning about new or changed marketplace conditions (Fiol, 1994; Van Knippenberg and Schippers, 2007; Page, 2007). Separately, work in cultural sociology also conceptualizes heterogeneity as a reflection of the variety of cultural resources available to organizational members, arguing that such breadth is beneficial for addressing change, uncertainty, and divergent environmental demands (Stark, 2011; Hallett and Ventresca, 2006; Swidler, 1986).

We reconcile these divergent perspectives by drawing on the core insight that cultural heterogeneity is not a unidimensional construct. Rather, we argue that it can arise in two conceptually distinct ways: from disagreement among members’ cultural perceptions or from a broad cultural
repertoire shared across organizational members. As illustration, imagine a stylized world in which there exist only two possible cultural beliefs, $A$ and $B$. Imagine further two hypothetical organizations. In the first organization half of the employees espouse belief $A$ and the other half espouse belief $B$. In the second organization, all employees espouse both beliefs. The two organizations appear similarly culturally heterogeneous—in both the two beliefs are equally prevalent. But the nature of this diversity is different. In the first organization, heterogeneity stems from divergent cultural beliefs. We refer to this as \textit{compositional heterogeneity}. The second organization is characterized by \textit{content heterogeneity}: the culture consists of a diverse range of beliefs. Although compositional and content heterogeneity are interrelated, they are analytically distinct. An organization with high content heterogeneity need not be compositionally heterogeneous.

Integrating insights from organizational theory and cultural sociology, we argue that compositional and content heterogeneity are related to different organizational outcomes. Consistent with a broad body of work on the organizational consequences of cultural strength, we theorize that compositional heterogeneity will be more directly linked to a firm’s coordination and execution capabilities and thus to indicators of performance such as profitability. By contrast, drawing on group learning research and the toolkit theory of culture, we posit that content heterogeneity will be more closely tied to a firm’s capacity for creative exploration and therefore to expectations of the firm’s market value.

Empirically, the methods most commonly used to study organizational culture—chiefly self-reports (O’Reilly et al., 1991) and participant-observation (Kunda, 2009; Turco, 2016)—are impractical for generating fine-grained, dynamic measures of compositional and content-based heterogeneity. Traditional culture surveys are often unwieldy, yielding low response rates, and can only be feasibly collected on an intermittent basis. At best, they yield static snapshots of changing organizational culture. In part for this reason, prior work examining the link between culture and firm performance has tended to rely on cross-sectional designs and simply side-stepped questions about the causal relationship between the two.
To overcome these limitations, we apply the tools of computational linguistics to derive novel, time-varying measures of compositional and content heterogeneity for a sample of nearly 500 publicly traded companies on Glassdoor (www.glassdoor.com)—a career intelligence website that allows employees to evaluate and comment on their firms. Given that cultural content can appear in a wide variety of reviews, we use unsupervised learning to identify distinct cultural topics in the nearly one million sentences that contain the word “culture” and its synonyms. We then train a topic model, which we fit to all employee reviews in our sample, and derive our cultural measures based on these identified topics. To move closer to causal estimates of the relationship between these two forms of cultural heterogeneity and firm performance, we apply coarsened exact matching (Iacus et al., 2012) to identify pairs of firms that vary on the cultural dimensions of interest but are otherwise observationally equivalent.

Cultural Heterogeneity and Organizational Performance

Culture is often understood by organizational scholars as a “system of publicly and collectively accepted meanings” (Pettigrew, 1979, p. 574) that a group—including a formal organization—develops in response to challenges of external adaptation and internal integration (Schein, 2010). These meanings manifest both in the form of deeply rooted assumptions and beliefs about the world, as well as in the normative and behavioral expectations that they prescribe (Schein, 2010; Mobasseri et al., forthcoming). Cultural heterogeneity, concomitantly, implies a variety of beliefs and behavioral norms in an organization.

Organizational theorists see cultural heterogeneity as both a blessing and as a curse. Although work on organizational culture is vast and fragmented (Chatman and O’Reilly, 2016), two conflicting themes on the relationship between cultural heterogeneity and performance pervade these literatures. The first, most strongly associated with research on cultural strength, sees cultural heterogeneity as an impediment to organizational performance. This line of work tends to con-
ceptualize culture as a solution to a complex coordination problem. Heterogeneous cultures, it is argued, are detrimental to organizational performance because they undermine interpersonal integration and erode internal cohesion.

Arguments that link cultural strength to organizational performance fall under two broad categories. First, cultural homogeneity is assumed to promote interpersonal coordination by facilitating goal alignment and behavioral consistency (Kreps, 1996; Gordon and DiTomaso, 1992; Weber and Camerer, 2003). Consequently, firms with a strong culture can exercise social control over group members more cost effectively than can firms with a weaker culture (O’Reilly and Chatman, 1996). Second, the absence of a unified and shared culture can generate fragmentation and a sense of personal estrangement. This can lead both to intraorganizational conflict, as well as to a decline in morale and a dampening of individual commitment and motivation (Martin, 1992; Jehn et al., 1999).

Heterogeneous cultures are particularly detrimental to organizational performance when they generate tensions between members, fragmenting them into conflicting identity groups and subcultures. Ideological disagreement between musicians and administrators in the Atlanta Symphony Orchestra, for example, led the organization into acute paralysis (Glynn, 2000). Whereas the musicians valued artistic excellence, administrators saw their primary objective as promoting economic efficiency. Research on the costs of cultural diversity in organizations echoes findings by economists and political scientists on the negative implications of cultural fractionalization at the nation-state level. This work finds, for example, that ethnic and cultural diversity foments civil conflict (Esteban et al., 2012), reduces generalized trust (Glaeser et al., 2000; Putnam, 2007) and undermines economic growth (Alesina et al., 2003).

In contrast, a second line of research—which hails predominantly from the organizational learning perspective—sees cultural heterogeneity as an advantage even while acknowledging its potential downsides. These benefits are more readily discerned when culture is conceptualized as a set of cognitive resources that organizational members deploy in the course of adapting to external changes
and competitive pressures. Two core assumptions underlie this perspective. First, organizations exist to solve complex and multifaceted problems, which are too complex for any individual to tackle singlehandedly. Second, creativity—the application of novel and useful solutions to problems (Amabile, 1996)—stems from the ability to recombine existing ideas in unconventional ways (Fleming, 2001; Uzzi et al., 2013; de Vaan et al., 2015).

Drawing on these insights, researchers who emphasize the benefits of diversity for performance argue that it does so by promoting a capacity for creative problem solving. This capacity derives not only from the fact that culturally diverse teams draw on a breadth of ideas and interpretative lenses, but also from the superadditive effects of this breadth: the novelty that emerges when ideas intersect and recombine (Page, 2007). The combination, for example, of a profit-oriented banking culture and a development-oriented social mission enabled the banks in Battilana and Dorado’s (2010) study to pioneer commercial microfinance in Bolivia in the early 1990s.

Culturally heterogeneous organizations learn more effectively by virtue of their cultural breadth. Such learning results in creativity and innovation when different cultural resources can be fused together to generate novel solutions. Cultural homogeneity, in contrast, is assumed to be detrimental to a firm’s capacity for creativity for three core reasons. First, employees in culturally uniform organizations are believed to be slower to recognize the need for change than their counterparts in organizations with weak cultures (Lant and Mezias, 1992). Second, whereas cultural strength can foster first-order learning—for example, determining how to more efficiently execute tasks that are known to be important for marketplace success, it can inhibit second-order learning—identifying which new tasks to take on in response to a new or changing competitive landscape (Denison, 1984). Third, the more consensual the culture of an organization, the less likely it is to nurture the development of counter-cultures that can be sources of new ideas and growth opportunities (Martin and Siehl, 1983).
Compositional and Content Cultural Heterogeneity

Taken together, these literatures suggest that cultural heterogeneity presents a fundamental trade-off: Culturally diverse firms are better at creative problem-solving, but this capability comes at the cost of decreased coordination and efficiency. In other words, cultural heterogeneity will be harmful for exploitation—harnessing existing and well-understood opportunities—and beneficial for exploration—the pursuit of new and unknown opportunities (March, 1991; Sørensen, 2002).

However, empirical support for such a trade-off is inconclusive. Sørensen (2002), for example, theorizes that the negative effects of cultural heterogeneity on firm performance should attenuate in volatile contexts, which ought to favor firms with a greater capacity for adaptation and exploration. The analyses he reports do not, however, yield consistent support for this contention. Similarly, Kotrba et al. (2012) report that the relationship between cultural heterogeneity and firm performance is contingent on various other cultural attributes and also varies by performance indicator (e.g., market-to-book ratio versus return on assets).

We argue that these inconclusive empirical findings reflect a theoretical shortcoming—the assumption that cultural heterogeneity is unidimensional. Previous conceptions of cultural heterogeneity have emphasized diversity in terms of demographic composition, highlighting the benefits and costs of disagreement among organizational members for different performance outcomes. The logic—predominantly in the literature on cultural strength—is that diversity manifests as disagreement in beliefs across individuals, leading to fragmentation and friction. Yet implicit in the arguments that see the creative potential of cultural diversity is an alternative conception of heterogeneity—one that emphasizes the breadth of cultural resources available to organizational members. The logic behind this conception is that diversity helps unify beliefs by providing members a common set of ideas with which to create new superadditive knowledge (Tadmor et al., 2012b). We contend that prevailing conceptions of cultural heterogeneity fail to differentiate between these two logics, and in turn have conflated two dimensions of cultural heterogeneity.
Building on insights from cultural sociology, we propose that heterogeneity can arise in two conceptually distinct ways. The first is the familiar demographic route: it can surface from cultural disagreement among organizational members. Yet a second way that heterogeneity can emerge is when people subscribe to multiple and potentially incompatible beliefs and values. As illustration, consider that firms commonly list a variety of values on their formal mission statements. For example, Netflix, the online DVD rental and streaming service, has an influential culture statement, publicly available online, that is 126 pages long and includes value statements ranging from freedom and autonomy to curiosity and responsibility.¹ Some of these values appear to be at odds with one another. For example, Netflix emphasizes selflessness and teamwork but also a relentless pursuit of individual performance. “We’re a team, not a family...” the culture slide deck states, “...adequate performance gets a generous severance package.”

Formal mission statements are often aspirational and do not necessarily represent the lived experiences of organizational members. For example, to our knowledge, no one in Enron officially endorsed malfeasance. There are nevertheless good reasons to expect that, like formal mission statements, enacted cultures can also exhibit a diversity of ideas, whether or not they are congruent with the officially espoused culture. Support for this view comes from research in institutional theory, which has demonstrated that people, especially in complex societies, are chronically exposed to multiple and incongruent institutional orders. The normative assumptions governing relationships in the family, for example, are very different from those governing market transactions (Friedland and Alford, 1991). People habitually draw symbolic boundaries between familial and economic relationships to resolve this incongruence (Zelizer, 2007).

Extending these arguments across levels of analysis, cultural sociologists have similarly proposed that organizations, like individuals, often operate in multi-institutional environments (Boltanski and Thévenot, 2006). For example, many companies cultivate a family-like ethos but draw on a

¹Downloaded more than 10 million times, the slide deck was hailed by Facebook’s Chief Operating Officer, Sheryl Sandberg, as the “most important document ever to come out of the valley.”
market logic to manage labor relations. The individuals comprising these organizations invariably intersect the different cultural orders upon which these organizations are founded. Medical professionals, for example, are required to navigate the tensions between competing cultural logics that understand medicine either through the lens of science or care-giving. The former emphasizes scientific authority and diagnostic success, whereas the latter conceptualizes quality health care as compassionate and preventive. Each prescribes different criteria for evaluating the legitimacy and desirability of behaviors and outcomes (Dunn and Jones, 2010). Institutional plurality begets friction and fragmentation in organizations when different individuals subscribe to distinct institutions, adopt different identities, and see the organizational mission through internally consistent but incongruent lenses. Organizations can overcome this tension when members find ways to fuse these different cultural components (Battilana and Dorado, 2010; Zilber, 2002; Besharov, 2014).

A second, related insight from cultural sociology is that humans are indeed cognitively equipped to internalize and selectively deploy multiple, coinciding cultural frames. This approach conceives of culture as a loosely-held repertoire or cultural “toolkit” (Swidler, 1986). Research by cognitive and cultural psychologists generally supports this conceptualization. It finds that individuals hold multiple and potentially inconsistent cognitive schemas (DiMaggio, 1997) and are capable of identifying with multiple cultural identities (Morris et al., 2015). Different situations invoke the deployment of different cultural lenses. Participants in Swidler’s (2001) study of romantic relationships, for example, at times described their bonds through a prism of love and selfless commitment and at others emphasized their relationships’ rational and instrumental foundations. Individuals, in other words, do not necessarily subscribe to a single and internally coherent cultural order. Rather, they embody multiple cultural models by virtue of exposure to different institutional settings (Hallett and Ventresca, 2006; Harding, 2007).

Bringing together the two broad perspectives—one that emphasizes heterogeneity’s roots in group demography and the other that focuses on its origins in competing institutional orders and the breadth of cultural repertoires available to individual actors—we argue that cultural het-
heterogeneity comprises two analytically distinct dimensions: compositional and content-based. By *compositional heterogeneity* we refer to cultural disagreement among the individuals who make up the organization. By content heterogeneity, we mean the breadth of cultural beliefs to which those individuals subscribe. In other words, compositional heterogeneity is the result of variation between individuals, whereas content heterogeneity relates to variation within individuals.

To illustrate this distinction, we return again to our stylized example in which there exist only two possible cultural beliefs, $A$ and $B$. Figure 1 illustrates the demographic makeup of two hypothetical organizations, each represented by a circle. The organization on the left, comprising individuals who either adopt belief $A$ or belief $B$, exhibits high compositional heterogeneity but low content heterogeneity. Its culture is neither strong nor broad. The organization on the right, in contrast, is characterized by high content heterogeneity and low compositional heterogeneity. It has both a strong and broad culture. An analogy to national cultures might be useful for further illustrating these two forms of cultural heterogeneity. Compositionally heterogenous cultures are akin to multi-ethnic societies wherein different groups of people subscribe to different cultural understandings and identities, while compositionally homogeneous cultures correspond to mono-ethnic societies. Content heterogeneity, on the other hand, relates to the breath of cultural resources available to members of the culture. Pre-modern societies are typically culturally narrow by this construction: they tend to be ordered along a single dominant and internally coherent cultural system (usually anchored in a supernatural religious cosmology (Durkheim, 1915)). Modern complex societies are broader, exhibiting a plurality of non-overlapping institutional orders.

Distinguishing between these two components of heterogeneity helps to uncover an important insight: broad organizational cultures need not be characterized by high levels of interpersonal disagreement. Members of an organization can share the same diverse cultural toolkit. This diversity can breed behavioral consistency if a situation invokes the same cultural frame for different individuals (Fiol, 1994). Netflix, for example, has developed a variety of complementary human resource practices. The company places a strong emphasis on hiring and dismissal on the basis of cultural
fit and invests in instituting formal procedures and behavioral norms that are consistent with its espoused mission and values (McCord, 2014). Netflix, in other words, invests in generating low compositional heterogeneity and cultivating high content heterogeneity. Insofar as these practices are effective, they should produce a culture that is both consensual and broad.

Seen in this light, the trade-off between organizational coordination and problem-solving capacity no longer seems inescapable. If cultural heterogeneity comprises two dimensions—compositional and content-based—then culture can facilitate coordination without necessarily undermining creative problem-solving and innovation. Consistent with this logic, we argue that each dimension of heterogeneity should promote different types of organization outcomes. Drawing on the literature on cultural strength we propose that compositional heterogeneity will weaken an organization’s coordination and cohesion and will therefore undermine its capacity for effective execution. We therefore hypothesize that:

**HYPOTHESIS 1 (H1): All else equal, compositional heterogeneity will be negatively related to a firm’s capacity for efficient execution.**

Content heterogeneity, in contrast provides organizational members with a broad set of cultural resources. The important distinction here is that this diversity occurs within—by virtue of the breadth of the cultural toolkit—rather than among individuals. Because creativity and innovation stem from the recombination of hitherto unrelated ideas, a wide cultural repertoire should be conducive to individual creativity. Research on multi-cultural individuals finds that, by virtue of chronic exposure to different national cultures, they present a capacity for higher integrative complexity and creative output (Tadmor et al., 2012a). This ability to combine multiple perspectives was presumably similarly conducive to creativity for the machine operators in Stark’s (2011) ethnography of a late 1980s Hungarian factory. Operating in an organizational cultural climate that sees skill as the ultimate principle of value but that also promotes an anti-bureaucratic ethos that appreciates networking and relationship-building, these individuals were successful in forg-
ing innovative partnerships and pursuing semi-private enterprise under Hungary’s late-communist “second-economy” legislation. Drawing again on Page’s (2007) metaphor of scientific problem-solving, creative entrepreneurship emerges in organizations—such as the Hungarian factory, new-media startup, or financial trading room that are the focus of Stark’s ethnographic studies—not because chemists outperform economists but because a fusion of economic and chemist rationales lead to the discovery and seizure of new opportunities. We therefore hypothesize that:

**HYPOTHESIS 2 (H2):** All else equal, content heterogeneity will be positively related to a firm’s capacity for creativity and innovation.

**Language as a Window into Cultural Heterogeneity**

In defining both constructs—compositional and content heterogeneity—we begin with the premise that organizational culture can be detected in the language used by members (Pinker, 2007; Crémér et al., 2007). The relationship between language and culture is a complex one. A useful way of conceptualizing this relationship draws on a distinction between the behavioral and cognitive dimensions of culture (Mobasseri et al., forthcoming). Work focused on the behavioral dimension tends to think of language as set of norms that facilitate interpersonal coordination and that people who seek to fit in to an organization typically aim to follow. Weber and Camerer (2003), for example, experimentally demonstrate that linguistic conventions formed by a group of individuals solving a coordination task increases group efficiency but also serves as an impediment once groups with different conventions are fused. In more recent work, Srivastava et al. (2017) and Goldberg et al. (2016) develop a language-based measure of cultural fit and, using an email corpus and personnel records from a mid-sized firm, demonstrate that compliance with linguistic norms is positively related to individual attainment.

Language can also serve as a reflection of speakers’ underlying beliefs and assumptions. The language used in public discourse by activists and civil society organizations, for example, reflects
deep-current cultural shifts in Americans’ perceptions about nuclear energy in the nineteen eighties (Gamson and Modigliani, 1989) or about Muslims post the September 11th Attacks (Bail, 2012). Building on these insights, we propose that organizational culture can not only be detected by observing the degree of linguistic compliance that members exhibit when communicating with each other—for example, in emails or text messages—but also in the language they use to describe the organization as a whole. In particular, we focus on the topics that members use when describing their culture to each other and to outsiders. When explicitly talking about culture, organizational members consciously articulate the assumptions and beliefs they believe are prevalent in their organization. Unlike previous conceptions of organizational culture or climate that focus on categories such as innovation or transparency that are predefined by researchers or informants (e.g., senior leaders in the firm) (Ehrhart and Naumann, 2004; O’Reilly et al., 1991), ours neither privileges one set of cultural topics over others nor assumes that researchers and informants understand the culture better than the typical organizational member does. Instead, our approach assumes that all topics used in discourse about the organization’s culture are potentially informative.

Following Weber (2005), we distinguish between cultural toolkits at the actor-level and the broader cultural register at the firm- or field-level. Given a set of topics that organizational members use to describe culture in a given period, we define compositional heterogeneity as the dissimilarity of topics that group members mention in their characterizations. In other words, organizations exhibit greater compositional heterogeneity when their members tend to disagree with each other when describing the culture. We define content heterogeneity as the dissimilarity in or breadth of those topics. Organizations exhibit greater content heterogeneity when their members have access to and draw from a more diverse cultural toolkit.

Although we have theorized about the independent effects of compositional and content heterogeneity on firm performance, we acknowledge that the two constructs are interrelated. For example, in the extreme case of an organization in which every member describes the culture using a single cultural topic, there will necessarily be no compositional heterogeneity across members. However,
we do not observe such extreme cases in the data. In fact, our empirical measures of compositional and content heterogeneity are negatively correlated, possibly because individual reviewers on average discussing more topics creates less space for disagreement across reviewers. We leave it to future research to disentangle the extent to which observed correlations between these constructs follow from theory versus are produced by measurement error.

METHOD

Data Sources and Sample

The data include all employer reviews written by employees in the United States from January 2008 to July 2015 on the website Glassdoor. Glassdoor is a career intelligence website that attracts a diverse audience primarily as a job search platform. It has an estimated 17 million unique users per month. While their identities as employees are authenticated by Glassdoor, reviewers are anonymous, thus making the reviews less susceptible to bias stemming from fear of employer retribution. Reviews are either unsolicited or contributed by users searching for jobs in exchange for unlimited site access (see Appendix B for details). Popadak (2013) used similar employee review text to construct longitudinal culture measures.

We restricted the firm sample to: 1) publicly-traded companies for which we have access to performance data from Compustat, and 2) firms with at least 50 employee reviews in one or more quarters to ensure that there were a sufficient number of reviews to calculate our culture measures. A small number of reviews were later dropped from this sample because they did not contain at least five words given weight by the LDA culture model – only firm/quarters with at least 25 reviews were used in estimated models. The resulting sample contains 512,246 reviews across 492 organizations. We lagged all predictors by one quarter to assuage concerns of reverse causality and standardized the culture measures.

\[^2\text{Accessible at www.glassdoor.com.}\]
Measures

Dependent Variables

Our hypotheses focus on the link between cultural heterogeneity and firms’ capacity for efficient execution (H1) and creativity and innovation (H2). Because these capabilities are difficult to observe, our dependent variables instead focus on firm performance outcomes that are proximally related to these underlying capabilities. In particular, we link the capacity for efficient execution to firm profitability and the capacity for creativity and innovation to market expectations of a firm’s growth potential. As a robustness check, we also report results that link content heterogeneity to an outcome that is even more proximally related to the capacity for creativity and innovation: the volume of patents produced by a firm.

Firm profitability is measured using Return on Assets (ROA), defined as income before extraordinary items over total assets. Market expectation of a firm’s growth potential is measured using Tobin’s Q (TQ), defined as the market value of a firm’s assets relative to their book value. A firm’s capacity for effective exploration and innovation is reflected in the Tobin’s Q measure (Kogan et al., 2012). R&D intensity, patent counts, and patent impact as measured by citation counts each have a statistically and economically significant effect on Tobin’s Q (Hall et al., 2005). Formally:

\[
TQ = \frac{\text{market value of assets}}{0.9 \times \text{book value of assets} + 0.1 \times \text{market value of assets}}
\]

Where market value of assets is defined as:

\[
MV = \text{book assets} + (\text{market value of common equity} - \text{common quity} - \text{deferred taxes})
\]
Language-Based Measures of Cultural Heterogeneity

We develop language-based measures of cultural heterogeneity to capture variation in compositional and content heterogeneity. We measure compositional heterogeneity and content heterogeneity using free response text written by employees reviewing the firm. Following prior text analysis work, we treat each review as a “bag of words,” which assumes that we can identify topical content even after discarding word order. We then represent each review as a vector of unigram counts, which identifies how many times the review includes individual words. Together, these individual words comprise a set of the most popular words that appear across the entire text corpus.

Our empirical strategy consists of two primary steps: 1) training a linguistic topic model to identify distinct dimensions of organizational culture, and 2) fitting that model to our analytic sample to identify the cultural dimensions mentioned in each employee review. We use a Latent Dirichlet Allocation (LDA) topic model (see Appendix A for technical details). LDA inputs a document-term matrix, for which the rows are reviews and the columns are unigram counts, and identifies distinct topics across the corpus by observing words that tend to co-occur frequently within each review. LDA then outputs a document-topic matrix, for which each review is assigned to a probabilistic mixture of topics, or a probability distribution giving the percentages across all topics $c \in C$ that the model estimates comprising the review. The model predicts that two reviews with similar probability distributions contain similar content.

Identifying a Set of Cultural Dimensions

Training the LDA model allows us to learn what topics employees across many organizations collectively consider germane to organizational culture. Our model training approach requires a key assumption: when employees write about firm culture, they sometimes explicitly use the word
“culture” or a synonym, and sometimes do not. Regardless, we can use the presence of a culture synonym as a label that indicates a given sentence contains content relevant to culture. Training the LDA model on text with these explicit references allows the model to identify a set of cultural topics. The model is then fit to reviews in our analytic sample to identify the cultural topics in text containing either explicit or implicit culture references (see Appendix A for details).

The topics identified by the LDA model have face validity as cultural dimensions that capture beliefs, norms, and artifacts. One way to validate LDA topics is to examine the words that are most highly weighted within each topic. Table 1 shows the highest-weighted words for four hand-picked and four randomly selected LDA topics. The first set were hand-picked based on their highly distinctive culture content and refer to emphasis on quality versus quantity in production, the entrepreneurial environment, travel and multiculturalism, and the nature of social interactions, respectively. The randomly selected set is representative of average LDA topics. These topics refer to how employee performance is recognized, the cultural dynamics of mergers, fun and laid-back coworkers, and challenging work, respectively. The LDA topics are clearly germane to organizational culture, which supports our approach of training the linguistic model on sentences that contain a culture synonym.

In simply inspecting the 500 culture topics, it is sometimes difficult to identify exactly what distinguishes one topic from others. Our goal, however, is not to select LDA model parameters (e.g., the number of topics to output) that maximize the coherence or distinctiveness of the topics. We are not interested in the cultural content per se but rather the breadth of the content discussed and the extent to which reviewers agree or disagree about the content. As such, we output a large number of topics to ensure we tease apart conceptually meaningful distinctions between cultural topics. Our cultural heterogeneity measures are highly-correlated and our results consistent using different numbers of topics (i.e. 25, 50, 100, and 250), which indicates that measuring compositional and content heterogeneity with regards to the same linguistic baseline is what is required to make comparisons between organizations.
Measuring Compositional Heterogeneity

We measure compositional heterogeneity by assessing the degree to which a firm’s employees in a given quarter characterize the firm using dissimilar cultural topics (Appendix A provides a series of measurement validation checks). After fitting the LDA model to the reviews in our analytic sample, each review $i$ is represented as a probability distribution $p$ indicating the relative proportion of each cultural topic $c$ estimated as present in the review text.

We define compositional heterogeneity for a given firm/quarter as the mean Jensen-Shannon (JS) divergence between the LDA probability distributions for all unordered pairs of reviews $i, j$ for that firm/quarter, formally:

$$A = \frac{\sum_{i,j} JS(p_i, p_j)}{\sum_{i,j}}, \text{ for all } \{i,j \mid i < j\} \tag{3}$$

where the JS-divergence between the two probability distributions is defined as:

$$JS(p_i, p_j) = \frac{1}{2} KL(p_i, M) + \frac{1}{2} KL(p_j, M) \tag{4}$$

and where $M = \frac{1}{2}(p_i + p_j)$ and $KL(p_i, M)$ is the Kullback-Leibler divergence of $M$ from $p_i$: 

18
JS-divergence is a symmetric measure of the dissimilarity of two probability distributions. It is well-suited for comparing sparse, power-law distributions of words observed in natural language and has been used previously to measure the similarity of organizational members’ language use (Goldberg et al., 2016; Srivastava et al., 2017).

**Measuring Content Heterogeneity**

We measure content heterogeneity by assessing the degree to which a firm’s employees discuss a broad versus narrow set of cultural topics (see Appendix A). Each review $i$ is represented as a probability distribution $p$ indicating the relative proportion of each cultural topic $c$ estimated as present in the review text. We apply the Herfindahl index, a popular measure of concentration, to these probability distributions, and calculate the mean Herfindahl score across all reviews for a given firm/quarter. Formally:

$$
\bar{H} = \frac{\sum_i \sum_{c \in C} (p_i^c)^2}{\sum_i} \tag{6}
$$

After taking the inverse, high values indicate that employees discuss a broader range of cultural topics, while low values indicate a narrower, concentrated set of topics. We take the natural log of the mean Herfindahl Index because the measure has a highly right-skewed distribution. Content heterogeneity for a firm/quarter is formally defined as::
Our language-based model of cultural heterogeneity has two advantages over survey-based culture measures. First, it allows us to measure dimensions of organizational culture longitudinally for a large, diverse set of organizations, which would be extremely difficult using more expensive and logistically-demanding survey methods. Second, the model inductively identifies topics that employees consider germane to organizational culture – we do not require the researcher to make a priori assumptions about the cultural topics that broadly characterize organizations.

Since the employees who write Glassdoor reviews were not selected through random sampling from the population of firm employees, Appendix B includes robustness checks to address the impact of any non-random selection of employees into writing Glassdoor reviews that could bias our findings. We find no evidence that either the number or composition of reviewers systematically changes with firm performance, and no evidence that the cultural heterogeneity measures themselves vary with the number of or composition of reviewers.

**Analytical Strategy and Estimation**

In addition to lagging our independent variables, we use Coarsened Exact Matching (CEM) to (partially) address concerns about endogeneity (Iacus et al., 2012). CEM identifies firm observations that vary on the culture variable of interest but have the same or very similar values for each control variable. Conditional on identifying all variables that affect the relationship between cultural heterogeneity and performance, CEM helps to correct for selection bias. We acknowledge, however, that the relationship between culture and firm performance is likely to be complicated
and bi-directional. CEM moves us closer to causal estimates; however, in the absence of exogenous variation in cultural heterogeneity or a compelling instrumental variable, we stop short of making a strong causal claim.

We implemented CEM as follows. For each culture variable (compositional and content heterogeneity), we first divided observations into high and low categories on the culture measure, defined by a value above or below the industry median.\(^3\) We then matched these firms identified as high or low on the culture variable with others that were the same or very similar on observed characteristics to achieve covariate balance. CEM allows us to match exactly on some covariates and coarsely on other covariates when it is infeasible to produce exact matches. We matched exactly on industry, year, and quarter. We matched coarsely on firm assets as a firm size control and on number of reviews to account for the level of coverage on the Glassdoor website. Coarse matches were identified using the binning algorithm default for the \texttt{cem} command in Stata.\(^4\) An attractive feature of CEM is it can produce matched strata with an unequal number of high and low culture observations so as to maximize the total number of matched observations and thus increase estimation efficiency in the subsequent analysis. The \texttt{cem} command produces simple weights to adjust for these differences during estimation, which we apply in all models. We also include strata fixed effects in all models, meaning we model variation in the performance outcomes between high and low culture observations within each stratum of matched observations.

\section*{RESULTS}

Table 2 reports univariate statistics and bivariate correlations for the final analytical sample. In line with expectations, compositional heterogeneity and content heterogeneity have a moderately high

---

\(^3\)We identify high and low culture groups using industry medians because the distributions of compositional and content heterogeneity vary substantially across some industries, where industry is defined by two-digit SIC code.

\(^4\)Sturge’s rule is the default algorithm, which is commonly used to determine the bin width when representing a probability distribution as a histogram.
negative correlation. Consistent with Hypothesis 1, compositional heterogeneity has a significant negative association with ROA. Consistent with Hypothesis 2, content heterogeneity has a significant positive correlation with Tobin’s Q. Note, however, both measures of cultural heterogeneity have moderately high correlations with firm size and the number of GlassDoor reviews.

We report CEM model results as our main findings. CEM creates 265 matched strata when matching on compositional heterogeneity, and 255 matched strata when matching on content heterogeneity. The matching strategy is successful in that it eliminates statistically significant differences in the observed covariates. Table 3 shows t-tests on covariate means before versus after matching for both the compositional heterogeneity and content heterogeneity matching. Any large t-statistics for firm assets and number of Glassdoor reviews before matching are sharply reduced after matching, such that we cannot reject the null hypothesis that there are no differences in means between the high and low culture groups.

Table 4 shows the coarsened exact matching results. Model 1 shows that compositional heterogeneity has a significantly negative association with ROA, providing support for Hypothesis 1. This model predicts a one standard deviation increase in compositional heterogeneity reduces ROA by approximately 0.14 standard deviations.

Given their moderately high correlation, we tested whether compositional heterogeneity exerts an effect on ROA independent of content heterogeneity. When matching high and low compositional heterogeneity observations, we mandated that matches have the same high/low content heterogeneity value. While sharply reducing the number of matches, Model 2 shows that the compositional heterogeneity association still holds, although the coefficient is only marginally significant.

Model 3 shows that content heterogeneity has a significantly positive association with Tobin’s Q, supporting Hypothesis 2. This model predicts a one standard deviation increase in content heterogeneity increases Tobin’s Q by approximately 0.18 standard deviations. We tested whether content heterogeneity exerts an effect on Tobin’s Q independent of compositional heterogeneity. When matching high and low content heterogeneity observations, we mandated that matches have
the same high/low compositional heterogeneity value. Although this strategy reduces the number of matches, Model 4 shows that the content heterogeneity association is still positive and significant.

**Robustness Check**

Given that we draw inferences about firms’ underlying capabilities based on performance outcomes but do not directly observe those capabilities, we ran a robustness check related to content heterogeneity based on patenting behavior. While a firm’s capacity for creativity and innovation is reflected in the Tobin’s Q measure (Kogan et al., 2012), we can measure this capacity more directly by modeling patenting outcomes. One simple measure of innovation capacity is firm patent counts, which has a statistically and economically significant effect on Tobin’s Q (Hall et al., 2005). All else equal, we expect firms with high content heterogeneity to patent more prolifically.

Table 5 shows a negative binomial model of annual patent counts for firms with at least 40 patents in a given year, collected from the U.S. Patent and Trademark Office website. R&D expenditure is defined as R&D spending relative to sales. Unsurprisingly, larger firms that invest more in R&D produce more patents. Consistent with our arguments, firms higher in content heterogeneity tend to patent more, though the coefficient is only marginally significant.

**DISCUSSION AND CONCLUSION**

The tension between integration and diversity is a recurring theme in organizational research. From a perspective that understands organizations as solutions to complex coordination problems, cultural heterogeneity is seen mostly as a source of dissonance and friction; through the lens of organizational learning and cultural sociology, however, it is seen as a necessary condition for creativity and innovation. The goal of this article has been to clarify and deepen our understanding of how these two facets of cultural heterogeneity relate to an organization’s underlying capacity for execution and creative exploration. We did so by shifting the analytical frame away from a
unidimensional understanding of cultural diversity based on group demography.

In contrast, we theorized that cultural heterogeneity has two distinct but related manifestations: compositional heterogeneity, or the extent to which organizational members diverge in how they understand the culture; and content heterogeneity, or the breadth of components comprising the culture. We argued that the former is negatively associated with effective coordination and execution and thus portends worsening firm profitability, while the latter is positively associated with new idea generation and predicts higher market expectations of future growth. To overcome the limitations of traditional approaches to measuring organizational culture—for example, indirect self-reports such as the Organizational Culture Profile (O’Reilly et al., 1991)—we used unsupervised learning to discern cultural content in employee reviews of nearly 500 publicly traded firms on the Glassdoor website. We then developed novel, time-varying measures of the two theorized forms of cultural heterogeneity based on these identified cultural topics. Using coarsened exact matching to help alleviate concerns about endogeneity that have plagued efforts to link culture to firm performance, we found support for our two core propositions.

Organizational Culture and Firm Performance

Findings from this investigation make important contributions to the wide-ranging literature—spanning organizational sociology (Sørensen, 2002) and economics (Weber and Camerer, 2003)—that examines the relationship between organizational culture and firm performance. In particular, we highlight the theoretical ambiguity in previous conceptions of cultural heterogeneity, which have tended to emphasize diversity as reflected in group demographics and suggested that heterogeneity will be harmful for exploitation yet beneficial for exploration. We trace the incongruities in these conceptualizations and the mixed empirical support for the hypothesized link to exploration-exploitation to the fact that prior work has conflated two distinct dimensions of cultural heterogeneity. Bringing together insights from the literatures on organizational cultural strength (Kotter and Heskett, 1992; Sørensen, 2002), group diversity and learning (Page, 2007; Jehn et al., 1999),
and cultural sociology (Boltanski and Thévenot, 2006; Swidler, 2001), we propose that heterogeneity can arise through two distinct routes: one based on disagreement among group members (compositional) and one based on the breadth of cultural resources available to group members (content-based). We conceptually link compositional heterogeneity to execution and task coordination and content heterogeneity to creativity and innovation. Our empirical results lend support for this more nuanced understanding of how cultural heterogeneity relates to firm capabilities and thus to different facets of firm performance: profitability and market expectations of future growth. In addition, whereas prior research attempting to study how culture impacts performance have mostly relied on cross-sectional designs, our approach based on coarsened exact matching (Iacus et al., 2012) moves closer to providing causal estimates of the link.

**Language as a Window into Culture**

This work has important implications for research that applies computational methods to the study of culture (Askin and Mauskapf, forthcoming), and, more specifically, for the burgeoning literature that uses language as a window into culture (Goldberg et al., 2016; Kramsch, 1998; Pinker, 2007). Whereas prior work has drawn inferences about how individuals fit into the social groups and organizations to which they belong based on analyses of interactional language use with other group members (e.g., Doyle et al. 2017, Srivastava et al. 2017), the present study derives measures of organizational culture based on the language that employees use in describing culture to each other and to the outside world.

In addition, given that culture is a multifaceted construct that people invoke in a variety of different ways, our analytical approach, which relies on unsupervised learning, provides a means to inferring cultural content from variegated textual descriptions of organizations. Finally, our approach to measuring culture does not rely on cultural categories defined by informants or researchers (Denison, 1984; O’Reilly et al., 1991). Instead, our approach allows cultural categories to emerge from the full set of naturally occurring topics that employees use in describing organiza-
tional culture. To put it differently, our method provides a means to identifying cultural content even when people have varying conceptions and definitions of culture as a construct.

Cultural Toolkit Theory

The approach developed in this paper also offers a way to inject greater theoretical precision to cultural toolkit theory (Swidler, 1986, 2013)—one of the most influential perspectives in contemporary cultural sociology but also one that has been criticized as vague and susceptible to slippage in terminology and use of underlying constructs (Lamont, 1992; Small et al., 2010).

Our approach provides a means to build upon Weber’s (2005) attempt to sharpen and apply toolkit theory to the realm of organizations and the fields in which they are embedded. Weber develops an analytic approach, which combines qualitative semiotic analysis with quantitative techniques for dimensionality reduction, to distinguish between the cultural register of an organizational field and the cultural toolkits of specific organizations. This approach allows him to characterize the cultural similarity between organizations. Our approach provides a complementary, and arguably more objective and efficient, means to achieving these same ends. It can be readily extended to identify the range of cultural topics in use within an organizational field or industry and to discern when organizational actors begin to draw upon specific cultural resources in response to a changing institutional landscape. Moreover, it can be used to characterize the changing distance between firms in the space of cultural topics.

The Role of Culture in Mergers, Acquisitions, and Alliances

That our approach can be used to derive measures of the cultural similarity or divergence between pairs of firms points to its potential utility in examining the role of culture in the success or failure of mergers and acquisitions, (Bauer and Matzler, 2014; Stahl and Voigt, 2008; Van den Steen, 2010; Weber and Camerer, 2003), as well as joint ventures and alliances (Park and Ungson, 1997; Pothukuchi et al., 2002). Whereas prior work has relied on static, survey-based measures of culture
(Stahl and Voigt, 2008), formal models (Van den Steen, 2010), or laboratory experiments that may lack external validity (Weber and Camerer, 2003), our approach allows for the construction of time-varying measures of cultural alignment between firms based on employee reviews.

More importantly, our theoretical model points to several new avenues for assessing the effects of cultural similarity between firms on post-merger or post-alliance success: (1) the degree of overlap in cultural topics used to describe the firms; (2) the extent to which the firms exhibit comparable levels of compositional heterogeneity; (3) the extent to which the firms have similar levels of content heterogeneity; and (4) the joint effect of similarity or difference in the two dimensions of culture between the firms. Together, these more fine-grained measures have the potential to provide a more nuanced understanding of interorganizational cultural compatibility and its consequences for integration and coordination across firm boundaries.

Limitations and Directions for Future Research

Given the nature of the data we analyze, this study has at least three key limitations, which also point to avenues for future research. First, as noted above, we draw inferences about firms’ underlying capabilities based on their performance outcomes; however, we do not directly observe or measure these latent capacities (except to a limited extent in the robustness check that links content heterogeneity to patenting outcomes). A natural extension of our patenting analysis would be to build more directly on Balsmeier et al. (2016), who conceptually distinguish between and develop measures of firm patents that are more explorative versus exploitative. Although we expect compositional heterogeneity to affect the capacity for exploitation in myriad ways that might not be detected through patenting behavior, we anticipate—on balance—that compositional heterogeneity will be negatively associated with the exploitative patents a firm succeeds in filing. By contrast, we expect that—all else equal—content heterogeneity will be positively associated with a firm’s success in filing explorative patents.

Second, the coarsened exact matching approach we use (Iacus et al., 2012) achieves balance
between our treatment and control groups by matching on observed firm attributes. It does not, however, address potential threats to causal identification stemming from unobserved heterogeneity. Future research in this vein—especially studies that draw on data sets spanning longer time horizons and thus affording a window into changing firm cultures—could account for time-invariant unobserved heterogeneity by estimating within-firm models. Over the time horizon of our data set, employee descriptions of firm culture in Glassdoor reviews simply do not exhibit sufficient temporal variance to support the use of within-firm estimates.

Finally, although we report robustness checks that help to dispel concerns that our findings can be accounted for by compositional shifts in the kinds of employees who choose to comment about firm culture prior to changes in firm performance, we cannot fully rule out the potentially confounding role of selection effects. We leave to future research the task of more thoroughly accounting for selection dynamics in employee reviews. For example, researchers could draw on national survey panels to identify a representative set of employees at firms included in the Glassdoor data and ask them to rate their firm using the same pro and con questions used by Glassdoor.

**Conclusion**

This study paves the way for novel investigations of the role of culture in organizational performance. It highlights that cultural heterogeneity can be a double-edged sword, with its compositional form foreshadowing a decline in profitability and its content-based form heralding heightened market expectations of future firm growth. Moreover, it highlights the value of language as a window into changing organizational culture.
References

Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg

Amabile, T. M.

Askin, N. and M. Mauskapf

Bail, C. A.

Balsmeier, B., L. Fleming, and G. Manso

Battilana, J. and S. Dorado

Bauer, F. and K. Matzler

Besharov, M. L.

Blei, D. M., A. Y. Ng, and M. I. Jordan

Boltanski, L. and L. Thévenot

Chatman, J. A. and C. A. O’Reilly

Crémer, J., L. Garicano, and A. Prat
de Vaan, M., B. Vedres, and D. Stark

Denison, D. R.
1984. Bringing corporate culture to the bottom line. *Organizational dynamics*, 13(2):5–22

Denison, D. R. and A. K. Mishra

DiMaggio, P.

Doyle, G., A. Goldberg, S. B. Srivastava, and M. C. Frank

Dunn, M. B. and C. Jones

Durkheim, E.

Ehrhart, M. G. and S. E. Naumann

Esteban, J., L. Mayoral, and D. Ray

Fiol, C. M.

Fleming, L.

Friedland, R. and R. R. Alford

Gamson, W. A. and A. Modigliani

Glynn, M. A.

Goldberg, A., S. B. Srivastava, V. G. Manian, and C. Potts

Gordon, G. G. and N. DiTomaso

Hall, B. H., A. Jaffe, and M. Trajtenberg

Hallett, T. and M. J. Ventresca

Harding, D. J.

Iacus, S. M., G. King, G. Porro, and J. N. Katz

Jehn, K. A., G. B. Northcraft, and M. A. Neale

Kogan, L., D. Papanikolaou, A. Seru, and N. Stoffman

Kotrba, L. M., M. A. Gillespie, A. M. Schmidt, R. E. Smerek, S. A. Ritchie, and D. R. Denison

Kotter, J. P. and J. L. Heskett
Kramsch, C. . .

Kreps, D. M.

Kunda, G.

Lamont, M. . .

Lant, T. K. and S. J. Mezias

March, J. G.

Martin, J.

Martin, J. and C. Siehl

McCord, P.

Mobasseri, S., A. Goldberg, and S. B. Srivastava

Morris, M. W., C.-y. Chiu, and Z. Liu

O’Reilly, C. A., J. Chatman, and D. F. Caldwell
O’Reilly, C. A. and J. A. Chatman

Page, S. E.

Park, S. H. and G. R. Ungson

Pettigrew, A. M.

Pinker, S.

Popadak, J. A.
2013. A corporate culture channel: How increased shareholder governance reduces firm value.


Putnam, R. D.

Schein, E. H.

Small, M. L., D. J. Harding, and M. . -. Lamont
2010. Reconsidering culture and poverty.

Sørensen, J. B.

Srivastava, S. B., A. Goldberg, V. G. Manian, and C. Potts
2017. Enculturation trajectories: Language, cultural adaptation, and individual outcomes in organizations. Management Science, Published online in Articles in Advance 02 Mar 2017(0):null.
Stahl, G. K. and A. Voigt

Stark, D.

Swidler, A.

Swidler, A.

Swidler, A. X.

Tadmor, C. T., A. D. Galinsky, and W. W. Maddux

Tadmor, C. T., P. Satterstrom, S. Jang, and J. T. Polzer

Turco, C.

Uzzi, B., S. Mukherjee, M. Stringer, and B. Jones

Van den Steen, E.

Van Knippenberg, D. and M. C. Schippers

Weber, K.

Weber, R. A. and C. F. Camerer
Zelizer, V. A.

Zilber, T. B.
Figure 1: Stylized example of compositional and content heterogeneity. Two hypothetical organizations are represented by a circle each. Individuals making up the organization are represented by letters, corresponding to the cultural beliefs they espouse. Organization A exhibits low compositional and content heterogeneity, whereas organization B exhibits high compositional and content heterogeneity.
### Table 1: Highest-Weighted Words for Select LDA Culture Topics

<table>
<thead>
<tr>
<th>Selected Topic #</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>qualiti product high produc deliv commit counter deliveri emphasi quantiti speed result output compromis highest assur expect deliver content outcom sacrif emphas bean cost craft creat sloppi</td>
</tr>
<tr>
<td>16</td>
<td>entrepreneuri thrive initi spirit dynam motiv collabor starter creativ individu passion empow suit unstructur structur ideal autonomi succeed entrepreneur mindset autonom ambigu flourish</td>
</tr>
<tr>
<td>33</td>
<td>travel opportun world countri meet experi chanc interact abroad global multicultur visit oversea airlin relocreloc deploy flight foreign adventur airport</td>
</tr>
<tr>
<td>35</td>
<td>social fun event activ interact lot frequent aspect regular includ bond scene gather anti respons network regularli outing media committe host justic conduct conscion mingl introvert societi awkward outgo</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Topic #</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>473</td>
<td>perform reward recognit recogn incent individu consist mediocr contribut resul</td>
</tr>
<tr>
<td>415</td>
<td>compani exist virtual anymor bought parent basic built longer sold</td>
</tr>
<tr>
<td>399</td>
<td>fun cowork great cool amaz outgo hang toy lightheart easygo brilliant train hip</td>
</tr>
<tr>
<td>481</td>
<td>challeng work interest reward present demand inher tackl workplac</td>
</tr>
</tbody>
</table>
Table 2: Univariate Statistics and Bivariate Correlations

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>S.D.</th>
<th>Number of Observations</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ROA</td>
<td>1.52</td>
<td>2.52</td>
<td>2595</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>TQ</td>
<td>1.77</td>
<td>0.76</td>
<td>2595</td>
<td>0.46**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Lag Compositional Heterogeneity</td>
<td>0.04</td>
<td>0.93</td>
<td>2595</td>
<td>-0.22*** -0.31***</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Lag Content Heterogeneity</td>
<td>-0.05</td>
<td>0.90</td>
<td>2595</td>
<td>0.058** 0.061**</td>
<td>-0.54***</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Lag Log of Assets</td>
<td>10.38</td>
<td>1.91</td>
<td>2595</td>
<td>-0.15*** -0.45***</td>
<td>0.42***</td>
<td>-0.0024</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Lag Log Number of Reviews</td>
<td>4.76</td>
<td>0.67</td>
<td>2595</td>
<td>0.070*** 0.070***</td>
<td>0.062**</td>
<td>-0.17***</td>
<td>0.20***</td>
<td>1</td>
</tr>
</tbody>
</table>

* p < 0.05, ** p < 0.01, *** p < 0.001
Table 3: t-Tests on Covariate Means Before and After Matching

<table>
<thead>
<tr>
<th></th>
<th>Before matching</th>
<th>Matched sample</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t-test on means</td>
<td></td>
<td>(low-high compositional heterogeneity)</td>
<td>(low-high compositional heterogeneity)</td>
</tr>
<tr>
<td>Lag Log of Assets</td>
<td>-4.92</td>
<td>-0.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lag Log Number of Reviews</td>
<td>-6.49</td>
<td>-0.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strata</td>
<td></td>
<td>2515</td>
<td>265</td>
<td></td>
</tr>
<tr>
<td>Matched strata</td>
<td></td>
<td>265</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multivariate L1 distance</td>
<td>0.90</td>
<td>0.66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low compositional heterogeneity obs.</td>
<td>1639</td>
<td>360</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High compositional heterogeneity obs.</td>
<td>1644</td>
<td>375</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>t-test on means</td>
<td></td>
<td>(low-high content heterogeneity)</td>
<td>(low-high content heterogeneity)</td>
</tr>
<tr>
<td>Lag Log of Assets</td>
<td>0.55</td>
<td>1.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lag Log Number of Reviews</td>
<td>10.40</td>
<td>0.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strata</td>
<td></td>
<td>2496</td>
<td>255</td>
<td></td>
</tr>
<tr>
<td>Matched strata</td>
<td></td>
<td>255</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multivariate L1 distance</td>
<td>0.91</td>
<td>0.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low content heterogeneity obs.</td>
<td>1637</td>
<td>337</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High content heterogeneity obs.</td>
<td>1616</td>
<td>362</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4: Coarsened Exact Matching – High vs. Low Cultural Heterogeneity

<table>
<thead>
<tr>
<th></th>
<th>Column (1)</th>
<th>Column (2)</th>
<th>Column (3)</th>
<th>Column (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lag Compositional Heterogeneity</td>
<td>-0.33*</td>
<td>-0.40+</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.47)</td>
<td>(1.79)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lag Content Heterogeneity</td>
<td></td>
<td></td>
<td>0.11*</td>
<td>0.17*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(2.18)</td>
<td>(2.08)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.39</td>
<td>0.59**</td>
<td>1.01***</td>
<td>1.04***</td>
</tr>
<tr>
<td></td>
<td>(1.24)</td>
<td>(3.03)</td>
<td>(10.20)</td>
<td>(15.82)</td>
</tr>
</tbody>
</table>

Matching Weights
- yes
Stratum FEs
- yes
Other Cultural Dimension Matched
- no, yes, no, yes
Firm/Quarters
- 735, 422, 699, 398

Absolute $t$ statistics in parentheses
Standard errors clustered by firm
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patent Count</td>
<td></td>
</tr>
<tr>
<td>Lag Content Heterogeneity</td>
<td>0.12$^+$</td>
</tr>
<tr>
<td></td>
<td>(1.70)</td>
</tr>
<tr>
<td>Lag Log Number of Reviews</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>(0.66)</td>
</tr>
<tr>
<td>Lag .R&amp;D Expenditures</td>
<td>0.066$^*$</td>
</tr>
<tr>
<td></td>
<td>(2.11)</td>
</tr>
<tr>
<td>Lag Log of Assets</td>
<td>0.58$^{***}$</td>
</tr>
<tr>
<td></td>
<td>(4.57)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.82</td>
</tr>
<tr>
<td></td>
<td>(1.13)</td>
</tr>
<tr>
<td>Year FEs</td>
<td>yes</td>
</tr>
<tr>
<td>Industry FEs</td>
<td>yes</td>
</tr>
<tr>
<td>Firm/Years</td>
<td>334</td>
</tr>
</tbody>
</table>

Absolute $t$ statistics in parentheses
Standard errors clustered by firm
$^+$ $p < 0.10$, $^*$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$
Appendix A : Measuring Cultural Heterogeneity Using Latent Dirichlet Allocation

All analyzed text was first preprocessed according to standard text analysis conventions. We removed common stop words and punctuation, discarded word order, and stemmed the words using the Porter stemming algorithm.

To train the Latent Dirichlet Allocation (LDA) model, we constructed a document-term matrix for which the rows represent distinct sentences observed across all available reviews for all organizations that contain the word “culture” or a close synonym (environment, atmosphere, attitude, climate, value, philosophy, belief). This results in 904,613 sentences. We identify the 4,000 most popular unigrams in these sentences. Less popular words outside of this set were increasingly proper noun references, badly misspelled, or nonsense words. After the researchers manually removed proper nouns, the document-term matrix tracked the frequency of 3,870 words.

This set of training sentences was analyzed using LDA. LDA is a model of the probabilistic generation of a text corpus. Documents are represented as random mixtures of topics, and each topic is characterized as a probability distribution over words (Blei et al., 2003). We parameterized LDA to identify 500 topics present in these culture sentences. Each topic is characterized by a weighted set of words that tend to co-occur within documents.

After identifying cultural topics using this training set of sentences with explicit cultural references, we fit the LDA model to the reviews in our analytic sample. In contrast to clustering methods, LDA is a mixed membership approach, which assigns each document to a probability distribution over multiple topics. Figure A1 illustrates LDA’s assignment of each review in the analytic sample to a mixture of multiple culture topics.
Our measures of cultural heterogeneity are constructed using these topic probability distributions over each review. Figure A2 illustrates that firm/quarters with low compositional heterogeneity feature reviews with more similar topic probability distributions. Conversely, high compositional heterogeneity firm/quarters have reviews with more dissimilar topic probability distributions. Figure A3 illustrates that low content heterogeneity firm/quarters have reviews with more concentrated topic distributions on average, while high content heterogeneity firm/quarters have reviews with on average more uniformly distributed topic distributions.
Figure A2: Stylized Example of Compositional Heterogeneity

<table>
<thead>
<tr>
<th>Review</th>
<th>Low Compositional Heterogeneity</th>
<th>High Compositional Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="chart1" alt="Chart" /></td>
<td><img src="chart2" alt="Chart" /></td>
</tr>
<tr>
<td>2</td>
<td><img src="chart3" alt="Chart" /></td>
<td><img src="chart4" alt="Chart" /></td>
</tr>
<tr>
<td>3</td>
<td><img src="chart5" alt="Chart" /></td>
<td><img src="chart6" alt="Chart" /></td>
</tr>
</tbody>
</table>

Figure A3: Stylized Example of Content Heterogeneity

<table>
<thead>
<tr>
<th>Review</th>
<th>Low Content Heterogeneity</th>
<th>High Content Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="chart7" alt="Chart" /></td>
<td><img src="chart8" alt="Chart" /></td>
</tr>
<tr>
<td>2</td>
<td><img src="chart9" alt="Chart" /></td>
<td><img src="chart10" alt="Chart" /></td>
</tr>
<tr>
<td>3</td>
<td><img src="chart11" alt="Chart" /></td>
<td><img src="chart12" alt="Chart" /></td>
</tr>
</tbody>
</table>
Measure Variation

Organizational culture is stable but not invariant over time (Kotter and Heskett, 1992). As such, we examine the sources of variation in our measures of cultural heterogeneity. Table A1 decomposes the variance in the measures into three components: within a firm over time, across firms within an industry, and across industries. Compositional heterogeneity exhibits more within-firm variation than content heterogeneity. This accords with the notion that changes in leaders’ emphasis on nurturing or maintaining beliefs about firm culture can drive variation in the consistency with which employees describe the culture. In contrast, the breadth of topics that constitute a firm’s culture is more stable over time. Figures A4 and A5 plot the within firm variation in compositional and content heterogeneity, respectively, moving from time $t - 1$ to $t$. This visual evidence shows that cultural heterogeneity is relatively stable but not invariant over time.

<table>
<thead>
<tr>
<th>Table A1: Variance Decomposition of Cultural Heterogeneity Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>By Total Variation</strong></td>
</tr>
<tr>
<td>------------------------</td>
</tr>
<tr>
<td>Within Firms</td>
</tr>
<tr>
<td>Within Industries</td>
</tr>
<tr>
<td>Between Industries</td>
</tr>
</tbody>
</table>
Figure A4: Within Firm Variation in Cultural Agreement

Figure A5: Within Firm Variation in Cultural Breadth
Additionally, we examine the within-firm temporal stability of the cultural heterogeneity measures across the full distributions of the measures. Figures A6 and A7 plot kernel density estimates of the distribution of each culture measure moving within-firm from time $t - 1$ to $t$. For both measures, Kolmogorov-Smirnov tests fail to reject the null hypothesis that the two distributions are different, providing statistical evidence that the culture measures exhibit relative stability over time.

Figure A6: Time Variation in Compositional Heterogeneity
Construct Validity

Beyond the face validity of the cultural topics that we demonstrated in Table 1, our heterogeneity measures themselves have construct validity as capturing variation along these culture dimensions. Table A2 shows the firms in the most represented industry in the data that score highest and lowest on both cultural heterogeneity measures. Firms are split into large and small firms because the culture measures vary to some degree with firm size. Xerox has high compositional heterogeneity, or high disagreement among employees about how to characterize the culture. This accords with lay accounts of Xerox’s culture in the study period, during which a newly appointed CEO vowed to redefine the culture. Conversely, Facebook has low compositional heterogeneity, or high agreement about the culture. This is consistent with the company’s well-known emphasis on maintaining a startup culture focused on innovation, autonomy, and open collaboration. The firms high and low on content heterogeneity similarly conform to intuition. For example, MicroStrategy has high content heterogeneity, meaning its culture is organized about a broad, diverse set of cultural topics. Instead of keeping its engineers behind desks, the company is known to encourage them to work in the field in collaboration with clients so as to expose them to more challenges and potential solutions.
Table A2: Business Service Firms with Highest/Lowest Cultural Heterogeneity Scores, 2008-2015

<table>
<thead>
<tr>
<th>Highest Compositional Heterogeneity</th>
<th>Large Firms</th>
<th>Small Firms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xerox</td>
<td>Kelly Services</td>
<td></td>
</tr>
<tr>
<td>SAP</td>
<td>Convergys</td>
<td></td>
</tr>
<tr>
<td>Paypal</td>
<td>TeleTech</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lowest Compositional Heterogeneity</th>
<th>Large Firms</th>
<th>Small Firms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amdocs</td>
<td>National Instruments</td>
<td></td>
</tr>
<tr>
<td>Facebook</td>
<td>Sapient</td>
<td></td>
</tr>
<tr>
<td>Wipro</td>
<td>Cornerstone OnDemand</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Highest Content Heterogeneity</th>
<th>Large Firms</th>
<th>Small Firms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microsoft</td>
<td>MicroStrategy</td>
<td></td>
</tr>
<tr>
<td>Harris Corp</td>
<td>National Instruments</td>
<td></td>
</tr>
<tr>
<td>Facebook</td>
<td>Intuit</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lowest Content Heterogeneity</th>
<th>Large Firms</th>
<th>Small Firms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wipro</td>
<td>Virtusa</td>
<td></td>
</tr>
<tr>
<td>Infosys</td>
<td>Syntel</td>
<td></td>
</tr>
<tr>
<td>CGI Group</td>
<td>IGATE</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Restricted to firms with at least 3 quarterly observations. Large and small firms delimited by industry median size.

Additional face validity is demonstrated in the association between the cultural heterogeneity measures and Glassdoor respondents’ subjective assessments of the quality of firm culture and values. Generally speaking, we expect more compositional heterogeneous cultures to engender lower approval from members, and more content heterogeneous cultures to elicit higher approval. In supplementary Coarsened Exact Matching models, compositional heterogeneity is associated with a lower average culture and values rating, and content heterogeneity a higher rating.
Appendix B : Glassdoor Data Details and Reviewer Robustness Checks

Employees reviewing their company are required to enter both positive (“pro”) and negative (“con”) comments. Since our objective was to identify the general cultural dimensions mentioned by employees without regard to valence, we combined the pro and con text when analyzing the reviews. Examining the most highly-weighted words for each LDA culture topics reveals that the model identifies cultural topics that are largely agnostic with respect to valence. In other words, both positive and negative review text contribute to most of the culture topics. A close reading of the review text for several firms revealed that pro and con text often characterize the culture the same way, even if an individual reviewer is mentioning a given topic because she believes it is either a positive or negative aspect of the culture.

Most visitors come to Glassdoor first and foremost to search for jobs rather than to post an employer review. Glassdoor employs a “give to get” model to solicit employer reviews from users. In order to receive unlimited access to the site’s content, users have to submit an anonymous employer review. Internal Glassdoor research has found that this method mitigates ratings bias by reducing the prevalence of extremely positive and negative reviews.

Since the employees who write Glassdoor reviews were not selected through random sampling from the population of firm employees, a concern is that systematic variation in the number or composition of reviewers is driving the observed associations between the cultural heterogeneity measures and firm performance. We conducted two checks to examine the robustness of our results to potentially non-random selection of employees into writing Glassdoor reviews: 1) modeling within-firm variation in the number and composition of reviews as a function of firm size and performance, and 2) modeling within-firm variation in the cultural heterogeneity measures as a function of number and composition of reviews, and firm performance and size. The sample includes firms with at least six quarterly observations so as to have enough within-firm observations to
include firm fixed effects.

Table B1 shows within-firm models of the number and composition of reviews used when calculating the cultural heterogeneity measures. These models test whether the number or composition of reviewers systematically changes during periods of high or low firm performance, which could bias our calculations of cultural heterogeneity. We examined reviewer composition by measuring the percentage of reviews in a given firm/quarter written by employees in managerial positions as opposed to lower-level employees, as indicated by non-missing job title information. Models 1 and 2 show that net of firm size, the number of Glassdoor reviews does not vary as a function of either lagged Return on Assets (ROA) or Tobin’s Q (TQ). Specifications 3 and 4 model the percentage of managers writing reviews as function of firm performance while controlling for number of reviews and firm size. Reviewer composition is insensitive to lagged ROA, but the percentage of managers decreases with increasing Tobin’s Q. This result prompted us to include the percentage of managers as a control in our multivariate Tobin’s Q models – inclusion of the control had virtually no impact on the size or significance of the content heterogeneity coefficient, so our findings are robust to changes in reviewer composition along this dimension.

Table B2 shows within-firm models of the cultural heterogeneity measures as a function of number and composition of reviews, and firm performance and size. These models directly test whether, net of firm size and lagged performance, the cultural heterogeneity measures vary with the number or composition of reviews. Models 1 and 2 show that both the compositional and content heterogeneity measures have strong positive associations with the number of reviews, which reflects the sensitivity of the calculation of the measures to the number of reviews as inputs. These results prompted us to explicitly match on number of reviews in our coarsened exact matching models so as to ensure that variation in the number of reviews is not driving the results. In contrast, the
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ln(# reviews)</td>
<td>ln(# reviews)</td>
<td>% Managers</td>
<td>% Managers</td>
</tr>
<tr>
<td>Lag ROA</td>
<td>-0.0075</td>
<td>0.0000018</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.54)</td>
<td>(0.00)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lag TQ</td>
<td>-0.091</td>
<td>-0.017*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.12)</td>
<td>(2.38)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log of Number of Reviews</td>
<td>-0.0029</td>
<td>-0.0034</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.61)</td>
<td>(0.69)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lag Log of Assets</td>
<td>0.19*</td>
<td>0.15*</td>
<td>-0.0031</td>
<td>-0.011</td>
</tr>
<tr>
<td></td>
<td>(2.53)</td>
<td>(1.83)</td>
<td>(0.28)</td>
<td>(0.97)</td>
</tr>
<tr>
<td>Constant</td>
<td>1.98***</td>
<td>2.52**</td>
<td>0.40***</td>
<td>0.51***</td>
</tr>
<tr>
<td></td>
<td>(2.73)</td>
<td>(2.86)</td>
<td>(3.59)</td>
<td>(4.37)</td>
</tr>
<tr>
<td>Year FEs</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Quarter FEs</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Firm FEs</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Firm/Quarters</td>
<td>2776</td>
<td>2733</td>
<td>2776</td>
<td>2733</td>
</tr>
</tbody>
</table>

Absolute $t$ statistics in parentheses
Standard errors clustered by firm

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

The percentage of managers is not significantly associated with the culture measures.
Table B2: Cultural Heterogeneity on Reviewer Characteristics and Performance

<table>
<thead>
<tr>
<th></th>
<th>(1) Compositional Heterogeneity</th>
<th>(2) Content Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log of Number of Reviews</td>
<td>0.19***</td>
<td>-0.23***</td>
</tr>
<tr>
<td></td>
<td>(4.55)</td>
<td>(5.20)</td>
</tr>
<tr>
<td>% Managers</td>
<td>0.30</td>
<td>-0.12</td>
</tr>
<tr>
<td></td>
<td>(1.13)</td>
<td>(0.33)</td>
</tr>
<tr>
<td>Lag ROA</td>
<td>-0.0023</td>
<td>0.0029</td>
</tr>
<tr>
<td></td>
<td>(0.48)</td>
<td>(0.49)</td>
</tr>
<tr>
<td>Lag TQ</td>
<td>-0.00016</td>
<td>0.055</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.65)</td>
</tr>
<tr>
<td>Lag Log of Assets</td>
<td>-0.050</td>
<td>0.087</td>
</tr>
<tr>
<td></td>
<td>(0.58)</td>
<td>(0.91)</td>
</tr>
<tr>
<td>Constant</td>
<td>-2.01*</td>
<td>1.42</td>
</tr>
<tr>
<td></td>
<td>(2.21)</td>
<td>(1.38)</td>
</tr>
<tr>
<td>Year FEs</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Quarter FEs</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Firm FEs</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Firm/Quarters</td>
<td>2730</td>
<td>2730</td>
</tr>
</tbody>
</table>

Absolute t statistics in parentheses
Standard errors clustered by firm
+ p < 0.10, * p < 0.05, ** p < 0.01, *** p < 0.001