## ONLINE APPENDIX

### Appendix A: Glassdoor Data Details and Additional Robustness Checks

Employees reviewing their company on the Glassdoor website are required to enter both positive ("pro") and negative ("con") comments. Because 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 latent Dirichlet allocation (LDA) culture topic reveals many topics that are largely neutral in valence.

Most visitors come to Glassdoor to search for jobs rather than to post an employer review. Glassdoor has a "give to get" model to solicit employer reviews from users: to receive unlimited access to the site's content, users have to submit an anonymous employer review. Research using the Glassdoor data has found that this method mitigates ratings bias by reducing the prevalence of extremely positive and negative reviews (Marinescu et al., 2018).

Because 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 drives 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 the 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 A1 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 a function of firm performance while controlling for the 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, and including this control had virtually no impact on the size or significance of the intrapersonal heterogeneity coefficient. Thus our findings appear to be robust to (observable) changes in reviewer composition.

**Table A1. Reviewer Characteristics on Performance***

| | **(1)**<br>**Log # reviews** | **(2)**<br>**Log # reviews** | **(3)**<br>**% Managers** | **(4)**<br>**% Managers** |
|---|---|---|---|---|
| Lag ROA | –.0075 | | .0000018 | |
| | (1.54) | | (.00) | |
| Lag TQ | | –.091 | | –.017$^{\bullet}$ |
| | | (1.12) | | (2.38) |
| Log of number of reviews | | | –.0029 | –.0034 |
| | | | (.61) | (.69) |
| Lag log of assets | .19$^{\bullet}$ | .15 | –.0031 | –.011 |
| | (2.53) | (1.83) | (.28) | (.97) |
| Constant | 1.98$^{\bullet\bullet}$ | 2.52$^{\bullet\bullet}$ | .40$^{\bullet\bullet\bullet}$ | .51$^{\bullet\bullet\bullet}$ |
| | (2.73) | (2.86) | (3.59) | (4.37) |
| Year FEs | Yes | Yes | Yes | Yes |
| Quarter FEs | Yes | Yes | Yes | Yes |
| Firm FEs | Yes | Yes | Yes | Yes |
| Firm/quarters | 2776 | 2733 | 2776 | 2733 |

$^{\bullet}$ $p < .05$; $^{\bullet\bullet}$ $p < .01$; $^{\bullet\bullet\bullet}$ $p < .001$.
* Absolute t-statistics are in parentheses; standard errors are clustered by firm.

Table A2 shows within-firm models of the cultural heterogeneity measures as a function of the number and composition of reviews, as well as 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 interpersonal and intrapersonal heterogeneity measures have strong positive associations with the number of reviews, which reflects the measures' sensitivity 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 percentage of managers is not significantly associated with the culture measures.

**Table A2. Cultural Heterogeneity on Reviewer Characteristics and Performance***

| | (1)<br>Interpersonal hetero. | (2)<br>Intrapersonal hetero. |
|---|---|---|
| Log # reviews | .19••• | –.23••• |
| | (4.55) | (5.20) |
| % managers | .30 | –.12 |
| | (1.13) | (.33) |
| Lag ROA | –.0023 | .0029 |
| | (.48) | (.49) |
| Lag TQ | –.00016 | .055 |
| | (.00) | (.65) |
| Lag log assets | –.050 | .087 |
| | (.58) | (.91) |
| Constant | –2.01• | 1.42 |
| | (2.21) | (1.38) |
| Quarter FEs | Yes | Yes |
| Firm FEs | Yes | Yes |
| Firm/quarters | 2730 | 2730 |

• $p < .05$; •• $p < .01$; ••• $p < .001$.
* Absolute t-statistics are in parentheses; standard errors are clustered by firm.

Next, we conducted additional robustness checks to examine whether potentially omitted variables are biasing our estimates of the effects of cultural heterogeneity on firm performance. First, we investigated whether our interpersonal heterogeneity measure is simply picking up on the existence of differentiated subcultures in organizations. It is important to note that interpersonal heterogeneity need not indicate a division into subcultures. In fact, our measure of interpersonal heterogeneity reaches extreme values when people are idiosyncratically different from one another. When they are divided into subcultures, the measure is significantly lower, given that groups of individuals are culturally similar to one another.

Nevertheless, we implemented a supplemental analysis in which we used a clustering procedure to assess the extent to which each organization-quarter observation can be divided into cultural subgroups. The clustering procedure uses the K-means algorithm, based on our measure of cultural distance, to divide the population into clusters and then uses the gap statistic to assess the optimal division into subgroups (Tibshirani, Walther, and Hastie, 2001). In only about 5 percent of the cases did we find evidence of multiple subcultures—at least based on Glassdoor reviews. Model 1 in table A3 reports the results of a robustness check in which we included a control variable for the number of clusters, as measured by this gap statistic. Interpersonal heterogeneity is negatively related to return on assets even when we account for the existence of subcultures (i.e., the number of clusters).

We also tested whether the association between interpersonal heterogeneity and ROA is actually being driven by organizations with subcultures produced by greater structural segmentation. Models 2 and 3 show that our results are robust to controlling for structural segmentation, whether operationalized as the number of business segments in the firm or the number of 4-digit SIC codes spanned by the firm, respectively. Overall, we believe this provides strong evidence that the link between interpersonal heterogeneity and profitability is not driven by the existence of subcultures.

**Table A3. ROA on Interpersonal Heterogeneity, Controlling for Subcultures (Number of Distinct Clusters), Number of Business Segments, and Number of SIC Codes***

| | **(1) ROA** | **(2) ROA** | **(3) ROA** |
|---|---|---|---|
| Lag interpersonal hetero. | –.37$^{\bullet\bullet\bullet}$ | –.39$^{\bullet\bullet\bullet}$ | –.38$^{\bullet\bullet\bullet}$ |
| | (3.67) | (3.62) | (3.57) |
| Lag/# subcultures | .19 | | |
| | (1.14) | | |
| Lag/# business segments | | .066 | |
| | | (1.40) | |
| Lag/# SIC codes | | | –.029 |
| | | | (.36) |
| Lag log assets | .22$^{\bullet}$ | .22$^{\bullet}$ | .24$^{\bullet}$ |
| | (2.13) | (2.01) | (2.23) |
| Lag log/# reviews | .24 | .26 | .30 |
| | (1.78) | (1.75) | (1.83) |
| Constant | –.69 | –.59 | –.85 |
| | (.67) | (.55) | (.80) |
| Quarter FEs | Yes | Yes | Yes |
| Industry FEs | Yes | Yes | Yes |
| Firm/quarters | 3283 | 2997 | 3001 |

$^{\bullet}$ $p < .05$; $^{\bullet\bullet}$ $p < .01$; $^{\bullet\bullet\bullet}$ $p < .001$.
* Absolute t-statistics are in parentheses; standard errors are clustered by firm.

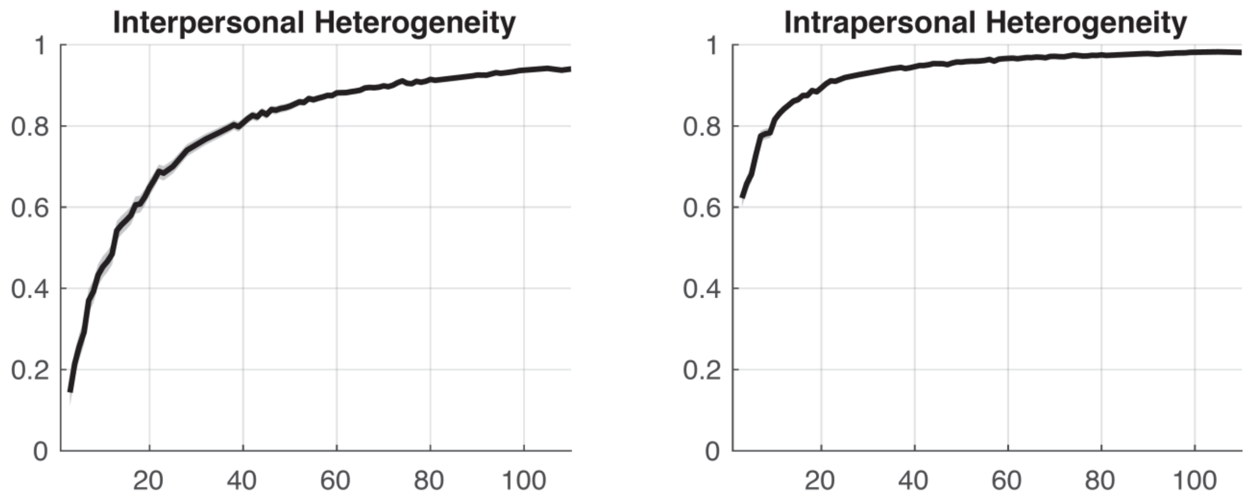
Further analyses we conducted sought to examine whether our results are driven by particular topics or are otherwise sensitive to culturally meaningless topics recovered by LDA. First, we examined whether our intrapersonal heterogeneity measure, which captures the extent to which reviews on average discuss a broad set of cultural topics, is sensitive to the possibility that some topics may be closer in conceptual meaning than others. To do so, we assumed that two topics that are frequently mentioned in the same review are likely to have lower conceptual distance than two topics that rarely co-occur, and therefore we applied the cosine distance metric to the review-by-topic matrix to produce distance weights for each topic pair. We then produced an alternative measure of intrapersonal heterogeneity that, for each topic pair, multiplies the estimated prevalence of topic 1 and topic 2, weighted by the pair's conceptual distance. We then summed across all topic pairs to measure the dispersion for an individual review and, as with the original measure, took the mean dispersion across all reviews for a firm/period. This alternative distance-weighted measure is highly correlated (.90 correlation coefficient) with the original measure, suggesting that there is little incremental value to accounting for the distance between topics and that our results are not skewed by the inclusion of similar cultural topics.

Second, we conducted a robustness check aimed at determining how sensitive our measures are to the inclusion of specific topics. To do so, we reproduced our measures by randomly selecting only a subset of k topics out of the 500 topics we recovered and used only these topics to calculate interpersonal and intrapersonal heterogeneity. We conducted analyses for values of k ranging from 5 to 100 topics. For each value of k, we repeated the process of randomly selecting topics 100 times, examining the correlation between the k-reduced measures and the original measures (based on the full set of 500 topics). The results reported in figure A1 plot the average correlation between the reproduced and original measures over these 100 random draws, as a function of the number of randomly selected topics. When k = 100, i.e., when we base our reproduced measures on only 20 percent of the original topics, the correlation with the original measures nears 1.

This analysis provides clear evidence that our measures are not determined by a handful of important topics and that culturally meaningless topics are not biasing our results. Even if we were to interpretatively determine that up to 80 percent of topics were culturally meaningless, we would still get effectively identical measures (our measures would retain .90 correlation even if we were to randomly select only 10 percent of topics).

These analyses provide strong evidence that even if some topics are culturally irrelevant or simply reflect linguistically aberrant patterns, these are not biasing our measures. They also help validate that our choice to rely on a 500-topic model is robust and preferable to models with a lower number of topics, which would likely render our results more sensitive to individual topics.

**Figure A1. Correlation between k-reduced heterogeneity measures and original (based on 500 topics) heterogeneity measures.**




## REFERENCES

Marinescu, I., N. Klein, A. Chamberlain, and M. Stuart
2018 "Incentives can reduce bias in online reviews." NBER Working Paper No. 24372.

Tibshirani, R., G. Walther, and T. Hastie
2001 "Estimating the number of clusters in a data set via the gap statistic." Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63: 411–423.

**Appendix B: Exploring the Interrelationships between Interpersonal and Intrapersonal Cultural Heterogeneity**

In the main body of the paper, we examine the independent relationship between the two cultural heterogeneity types—interpersonal and intrapersonal—and organizational performance. We show that the former predicts profitability and that the latter is associated with growth potential and innovation. A natural extension would be to explore how the two types of heterogeneity relate to one another in producing these outcomes.

Two considerations give us pause in performing such an analysis, however. First, we expect this relationship to be complex and nonlinear. Our theoretical framework leads to hypotheses about the independent main effects of interpersonal and intrapersonal heterogeneity on performance, but it does not straightforwardly imply what their joint effects would be. For example, greater intrapersonal heterogeneity implies a greater potential for interpersonal confusion: individuals who espouse multiple and potentially conflicting cultural frames might appear to their peers as behaviorally inconsistent across situations. At the same time, greater within-person cultural diversity mechanically increases, on average, the likelihood that two individuals would have cultural overlap. Thus intrapersonal heterogeneity might lead to both greater and diminished capacity for interpersonal coordination.

Second, although the two concepts are analytically distinct, their operationalization leads to a nonlinear, mechanical relationship between them. For example, at the extreme positive end of intrapersonal heterogeneity—i.e., when all organizational members equally espouse all possible cultural elements—interpersonal heterogeneity will, by design, be 0. But when intrapersonal heterogeneity is low, for example, when all members adopt one cultural element each, interpersonal heterogeneity can range from 0, when all members espouse the same element, to 1, when each espouses a different element. This relationship becomes even more complex when variability exists between members' levels of within-person heterogeneity.

To address the second, mechanical problem, we implemented a procedure to adjust interpersonal heterogeneity relative to what would be expected at random, holding intrapersonal heterogeneity constant. Let $N_{ot}$ be the number of individuals reviewing organization $o$ at time $t$. We can represent the set of reviews for organization $o$ at time $t$ as an $N_{ot} \times K$ matrix $X_{ot}$ (where $K$ is the number of LDA topics). Each row represents the probability distribution of each individual review over the set of $K$ topics. To induce the expected interpersonal heterogeneity, we randomly permuted each row in $X_{ot}$ to produce a permuted dataset labeled $X_{ot}^{i}$ and calculated interpersonal heterogeneity $B(X_{ot}^{i})$, as defined in equation 3 in the paper. This procedure preserves intrapersonal heterogeneity, as it does not change the shape of the distribution of probabilities over topics in each review. Nevertheless, it affects interpersonal heterogeneity by randomly shifting the topics over which these probability distributions are distributed. We repeated the process 1,000 times to produce a reference distribution of interpersonal heterogeneity. The mean of this reference distribution is the expected interpersonal heterogeneity, $E(B_{ot}) = \frac{1}{n}\sum_{i=1}^{n} B(X_{ot}^{i})$. We define the adjusted interpersonal heterogeneity as the distance between the observed and expected interpersonal heterogeneity, $\tilde{B}_{ot=} B_{ot} - E(B_{ot})$.

$\tilde{B}_{ot}$ is the interpersonal heterogeneity that is not mechanically explained by intrapersonal heterogeneity. We find that for all our firm-quarter observations this variable is negative, suggesting, as one would assume, that all firms exhibit a level of interpersonal heterogeneity that is lower than what would be expected at random. In other words, all firms have greater cultural consensus than would be the case if an equally sized set of individuals with random cultural preferences were to be assembled, keeping this set of individuals' level of intrapersonal heterogeneity constant. The analyses presented in this online supplement use the adjusted measure of interpersonal heterogeneity whenever the two heterogeneity variables are included in the same model.

We revisited the coarsened exact matching (CEM) models that test our two hypotheses while including both heterogeneity variables in the models. Model 1 in table B1 tests hypothesis 1 and reproduces model 2 in table 4. As expected, and consistent with hypothesis 1, (adjusted) interpersonal heterogeneity remains negatively predictive of ROA, but intrapersonal heterogeneity does not. Models 2 to 4 in table B1 revisit hypothesis 2, corresponding to models 2, 5, and 7 in table 5. Consistent with hypothesis 2, intrapersonal heterogeneity is predictive of greater market valuation and innovation output, but interpersonal heterogeneity is not. The only exception is model 3, which is marginally significant (at $p = .06$). In other model specifications—e.g., in an ordinary least squares (OLS) model with quarter and industry fixed effects, corresponding to model 3 in table 5—we find that intrapersonal heterogeneity is significantly predictive of the number of patents (at $p < .01$) but that interpersonal heterogeneity is not.

**Table B1. CEM Models with Adjusted Interpersonal Heterogeneity Measure***

| | (1) ROA matched | (2) TQ matched | (3) Log patent count matched | (4) # Back. cites matched |
|---|---|---|---|---|
| Lag intrapersonal hetero. | .026 | .098• | .085 | .25• |
| | (.21) | (2.10) | (1.88) | (2.46) |
| Lag interpersonal hetero. (adj.) | –41.8• | 12.5 | 8.86 | 8.82 |
| | (2.12) | (1.33) | (1.51) | (.67) |
| Constant | –.73 | 1.31••• | .21 | 1.21 |
| | (1.52) | (5.30) | (1.38) | (1.67) |
| Matching weights | Yes | Yes | Yes | Yes |
| Stratum FEs | Yes | Yes | Yes | Yes |
| Firm/quarters | 794 | 699 | 714 | 355 |

• $p < .05$; •• $p < .01$; ••• $p < .001$.
* Absolute t-statistics are in parentheses; standard errors are clustered by firm.

Overall, the results in table B1 indicate that our findings are robust to the inclusion of both heterogeneity variables in our models and that the two types of heterogeneity are differentially predictive of performance, as we hypothesize. Non-hypothesized relationships between cultural heterogeneity and performance are not driving our results. In additional analyses not reported here, we find that, when modeled separately in a CEM specification, intrapersonal heterogeneity does not predict profitability, and interpersonal heterogeneity does not predict market valuation or innovation.

Although theoretical predictions about an interaction effect between interpersonal and intrapersonal heterogeneity are not straightforward, we cautiously explore them in table B2. Our main hypotheses focus on the coordination disadvantages of interpersonal heterogeneity and the creative advantages of intrapersonal heterogeneity. Yet our outcomes are not perfect measures of coordination or creativity. For example, we use ROA as an indication of coordination efficiencies, but we expect that profitability is also affected, to some extent, by creativity. Similarly, patenting output depends on creativity, but the ability to translate creative ideas into useful innovation also depends on team coordination capabilities. We therefore expect that organizations that simultaneously exhibit low levels of interpersonal heterogeneity and high levels of intrapersonal heterogeneity—that is, firms with consensual and broad cultures—will also exhibit high profitability and innovation relative to other firms. Conversely, we expect that organizations with high levels of interpersonal heterogeneity and low levels of intrapersonal heterogeneity—that is, firms in which individuals espouse different and narrow cultural beliefs—will exhibit lower performance on both dimensions.

To evaluate this proposition, we explore interactions between the two measures in table B2. Due to expected nonlinearities in effects, we dichotomized the interpersonal and intrapersonal heterogeneity variables into high and low binary variables at their respective medians and examined their interaction, which produced four quadrants. This exercise led to a significant reduction in statistical power. We therefore explored only OLS models without coarsened exact matching of observations and included only one quadrant dummy in each model. The results reported in table B2 are broadly consistent with our assumption about the interaction between the two types of cultural heterogeneity. Firms with low interpersonal and high intrapersonal heterogeneity exhibit greater profitability, higher Tobin's Q, and higher innovation output than other firms. Culturally fragmented organizations, in contrast, whose employees adopt different and narrow cultural beliefs, exhibit lower profitability and market valuation. These firms also exhibit lower levels of innovation but at marginally statistically significant levels. Additional analyses wherein we included multiple quadrant dummies in each model did not yield consistent and statistically significant results. We therefore interpret the results in table B2 as suggestive but inconclusive evidence for the nonlinear, joint effects of interpersonal and intrapersonal heterogeneity on performance. We leave to future work further theoretical development and empirical exploration of the interrelationships between the two cultural heterogeneity measures.

**Table B2. Performance and Patent Outcomes on Heterogeneity Quadrants***

| | (1) ROA | (2) ROA | (3) TQ | (4) TQ | (5) Log # patents | (6) Log # patents | (7) # Back. cites | (8) # Back. cites |
|---|---|---|---|---|---|---|---|---|
| High inter., low intra. | –.50$^{\bullet\bullet\bullet}$ | | –.21$^{\bullet\bullet\bullet}$ | | –.14 | | –.16 | |
| | (3.69) | | (5.18) | | (1.87) | | (1.92) | |
| Low inter., high intra. | | .26$^{\bullet}$ | | .15$^{\bullet\bullet\bullet}$ | | .17$^{\bullet}$ | | .28$^{\bullet\bullet\bullet}$ |
| | | (2.00) | | (3.68) | | (2.30) | | (3.54) |
| Lag log assets | .18+ | .15 | –.094$^{\bullet\bullet}$ | –.10$^{\bullet\bullet}$ | .21$^{\bullet\bullet}$ | .21$^{\bullet\bullet}$ | .15$^{\bullet}$ | .14$^{\bullet}$ |
| | (1.73) | (1.46) | (2.79) | (3.04) | (2.72) | (2.67) | (2.29) | (2.37) |
| Lag log # reviews | .30$^{\bullet}$ | .31$^{\bullet}$ | .15$^{\bullet\bullet}$ | .15$^{\bullet\bullet}$ | .036 | .04 | –.19$^{\bullet}$ | –.18$^{\bullet}$ |
| | (2.08) | (2.03) | (3.12) | (3.04) | (.37) | (.40) | (2.30) | (2.27) |
| Constant | .60 | .51 | 3.01$^{\bullet\bullet\bullet}$ | 2.90$^{\bullet\bullet\bullet}$ | –.56 | –.77 | 2.40$^{\bullet\bullet\bullet}$ | 2.13$^{\bullet\bullet\bullet}$ |
| | (.65) | (.53) | (9.99) | (9.46) | (.76) | (.99) | (4.65) | (4.32) |
| Quarter FEs | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry FEs | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Firm/quarters | 3283 | 3283 | 3251 | 3251 | 3296 | 3296 | 949 | 949 |

+ $p < .10$; $^{\bullet}$ $p < .05$; $^{\bullet\bullet}$ $p < .01$; $^{\bullet\bullet\bullet}$ $p < .001$.

* Absolute t-statistics are in parentheses; standard errors are clustered by firm.

## Appendix C: 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 resulted in 904,613 sentences. We identified 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 we manually removed proper nouns, the document-term matrix tracked the frequency of 3,870 words.

This set of training sentences was analyzed using LDA—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, Ng, and Jordan, 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.

### Measure Variation

Organizational culture is stable but not invariant over time (Kotter and Heskett, 1992). As such, we examined the sources of variation in our measures of cultural heterogeneity. Figures C1 and C2 plot the within-firm variation in interpersonal and intrapersonal heterogeneity, respectively, moving from time $t-1$ to $t$. This visual evidence shows that cultural heterogeneity is relatively stable but not invariant over time.

**Figure C1. Within-firm variation in interpersonal heterogeneity.**

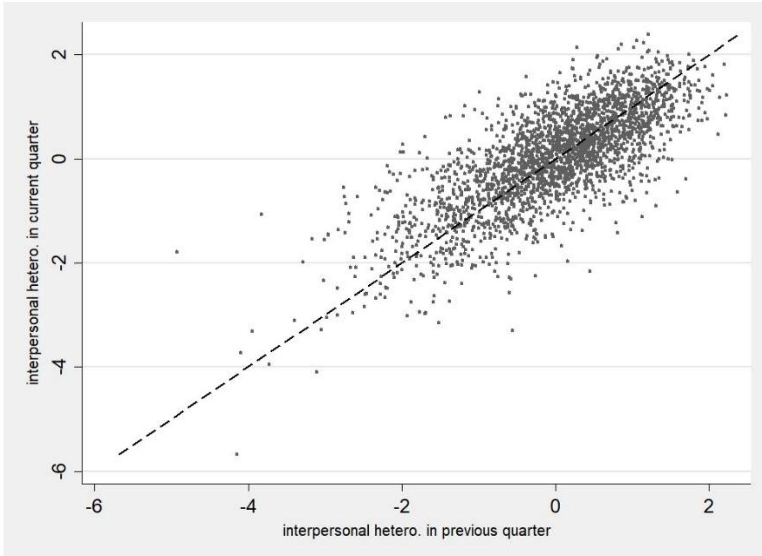
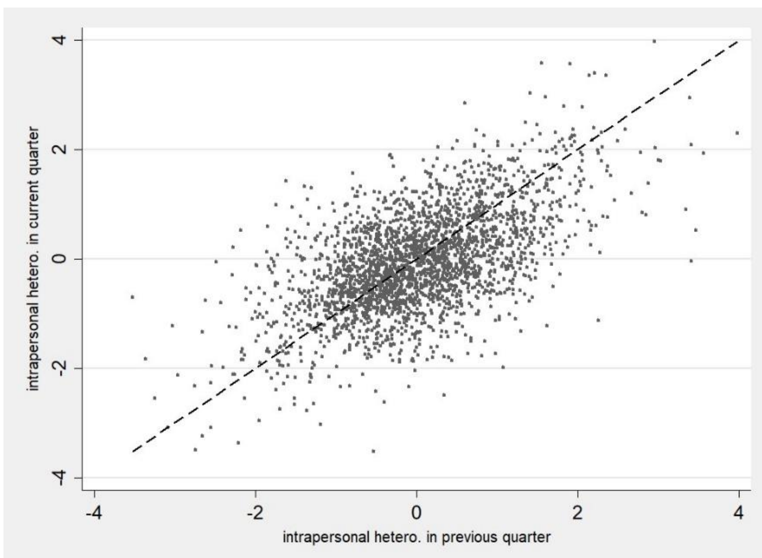



**Figure C2. Within-firm variation in intrapersonal heterogeneity.**




Additionally, we examined the within-firm temporal stability of the cultural heterogeneity measures across the full distributions of the measures. Figures C3 and C4 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 failed to reject the null hypothesis that the two distributions are different, providing statistical evidence that the culture measures exhibit relative stability over time.

**Figure C3. Time variation in interpersonal heterogeneity.**

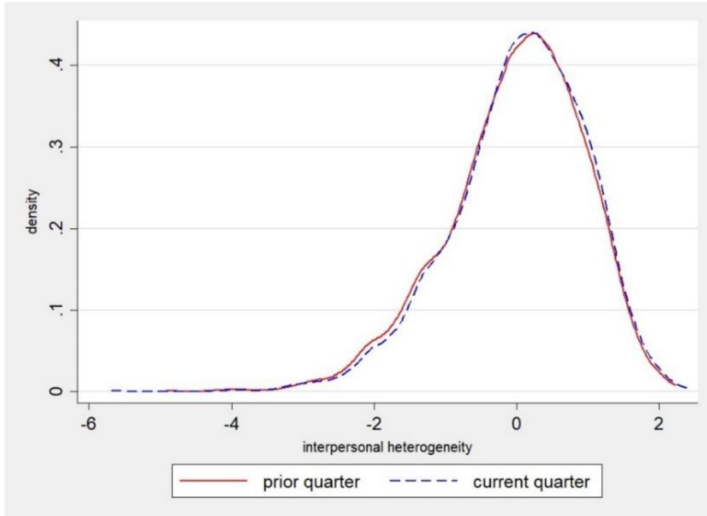
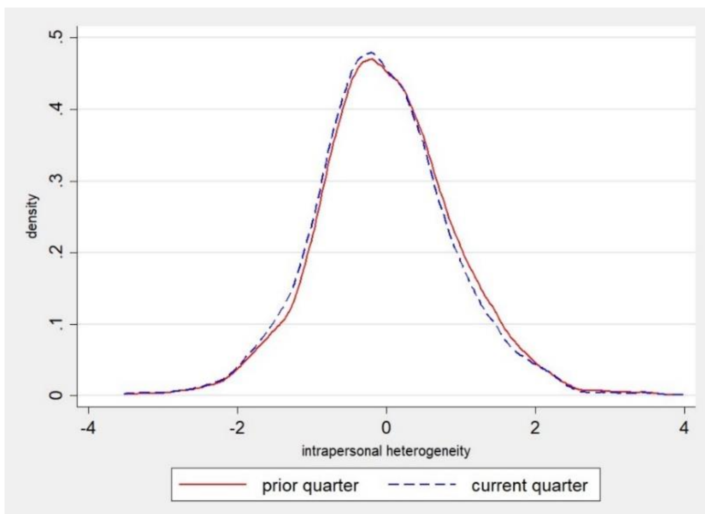



**Figure C4. Time variation in intrapersonal heterogeneity.**




## 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 C1 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 interpersonal 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 interpersonal 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 intrapersonal heterogeneity similarly conform to intuition. For example, MicroStrategy has high intrapersonal heterogeneity, meaning its culture is organized around 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 C1. Business Service Firms with Highest/Lowest Cultural Heterogeneity Scores, 2008–2015***

| | **Large firms** | **Small firms** |
|---|---|---|
| Highest interpersonal heterogeneity | Xerox | Kelly Services |
| | SAP | Convergys |
| | PayPal | TeleTech |
| Lowest interpersonal heterogeneity | Amdocs | National Instruments |
| | Facebook | Sapient |
| | Wipro | Cornerstone OnDemand |
| Highest intrapersonal heterogeneity | Microsoft | MicroStrategy |
| | Harris Corp | National Instruments |
| | Facebook | Intuit |
| Lowest intrapersonal heterogeneity | Wipro | Virtusa |
| | Infosys | Syntel |
| | CGI Group | IGATE |

* Restricted to firms with at least three 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 firms' culture and values. Generally speaking, we expect more interpersonally heterogeneous cultures to engender lower approval from members and more intrapersonally heterogeneous cultures to elicit higher approval. In supplementary CEM models, interpersonal heterogeneity is associated with a lower average culture and values rating and intrapersonal heterogeneity with a higher rating.

## REFERENCES

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

Kotter, J. P., and J. L. Heskett
1992 Corporate Culture and Performance. New York: Free Press.

## Appendix D: Comparing Selected Empirical Approach to One Based on Manual Coding of Cultural Categories

Our empirical approach relies on unsupervised learning of cultural content in employees' reviews based on phrases that contain culture or one of its synonyms. We also chose an arbitrarily large number of topics (500) and considered the distribution of these topics within and between individuals in a firm. A virtue of this approach is that it does not rely on subjective human judgment about which topics are or are not culturally meaningful. Instead, it implicitly assumes that topics derived from phrases that contain the word culture or one of its synonyms are culturally meaningful even if a given researcher cannot see this meaning. In other words, it allows for a cultural signal to be extracted from text even if researchers lack access to the cultural toolkit needed to detect that signal. As a consequence, our approach is more scalable and easier to port across empirical settings than typical interpretative approaches, and it is conducive for studying the distribution of a broad range of cultural beliefs when researchers are not focused on measuring the predefined linguistic signatures of specific cultural content. Moreover, our simulation analyses (see Online Appendix A) demonstrate that the resulting measures of interpersonal and intrapersonal heterogeneity are not sensitive to the specific topics used to derive these distributional measures and that measures based on considerably fewer topics are highly correlated with ones based on 500 topics.

Nevertheless, it remains unclear how our measures relate to ones derived from a more traditional approach that is based on human interpretation and coding of cultural categories. To evaluate the degree of correspondence, we undertook the following procedure. First, we began with a 100-topic model to make the task of manual coding more tractable. Second, each of the three coauthors independently reviewed each topic and the top 25 word stems associated with it to determine (1) which topics seemed to lack cultural coherence and should therefore be dropped; (2) which ones seemed to overlap and could therefore be combined; and (3) what label to assign to the remaining topics. Coders' independent judgments revealed substantial disagreement over both which topics should be discarded and which topics should be combined. For illustration, out of the 38 topics that at least one author deemed to be lacking coherence, only four topics were deemed so unanimously by all three coders. Similarly, after independently dropping and combining topics from the initial set of 100 topics, one coder proposed 33 topics, another coder 40 topics, and the third coder 57. As such, we sought to reconcile these independent attempts to manually code and interpret the topics using a systematic approach. In doing so, we began by applying the following decision rules: (1) discard topics that at least two coders identified as lacking coherence (20 topics were so identified); and (2) of the remaining 80 topics, combine topics when at least two coders had independently combined them. Applying these decision rules whittled the 80-topic set down to 59.

Next, we reviewed the resulting categorization. Of the 20 topics designated to be discarded, five were flagged by one coder as essential and were reinstated. In three additional cases we decided to move three retained topics to the discard list. We then revisited the combined topic list, erring on the side of separating combined topics when, upon further reflection, we could identify subtle differences between them. In the end, we ended up with 65 topics. Table D1 lists the 18 topics that were ultimately discarded, and table D2 lists the 65 topics (including combined topics) that were retained and the category labels we assigned to each.

**Table D1. 18 Discarded Topics**

| **Topic #** | **Top 25 words** |
|---|---|
| 5 | long term time employe compani short work stay sink swim lot period posit employ year train busi longer plan stand tenur promot staf contract |
| 7 | good work healthi ethic friendli stabl condit load peopl peer pressur enviro infrastructur benefit enviorn freindli env challang benifit pleasent ordin paymast remuner amic |
| 9 | good work learn salari facil fresher project technolog opportun onsit infrastructur exposur polici hike train hr technic apprais excel chanc transport join brand pressur |
| 18 | student research school colleg academ univers educ staff campu teach class graduat teacher faculti administr institut excel facil support world lab kid studi state |
| 23 | time work part full good employe flexibl relax friendli lot job enjoy spend stress spent worker hectic wast amount timer flexi pretti money alot |
| 27 | good work nice peopl friendli ppl ambienc disengag mark adapt christian crap compon sm employess dislik remind section won select smooth brand writer carpet |
| 31 | store custom fun discount retail manag employe product easi food cloth worker great sale servic sell restaur cowork deal job merchandis clean depend gener |

| | |
|---|---|
| 34 | compani make money daili corpor basi thing promis employe live run practic word save put heart preach complet show realiti manag absolut dollar true |
| 44 | depart depend team work group manag vari offic function divis locat hr cross project commun experi branch area lot collabor gener silo greatli interact |
| 47 | call center time day job phone question make custom expect talk meet manag supervisor person email answer cold desk constantli sit break stress sale |
| 48 | great work peopl team fun experi fab envior sweatshop stagnant hardcor stellar enviorn directli superior hill stodgi clean improp internet squeez american plain king |
| 63 | work good bad compani worst peopl profession ethic experienc thing experi absolut appl kind encount pathet recommend load join horribl imagin insid spoil posit |
| 69 | compani leav year employe peopl left month start hire ceo fire join stay replac employ due made end complet longer turn realiz desir feel |
| 73 | good work benefit salari pay peopl pretti compens infrastructur facil satisfi benifit reput ambienc cv freindli train prestig multipl mixtur bureacraci comapni sum ib |
| 89 | great peopl benefit work amaz compani awesom perk fantast product pay cowork locat ton fun terrif fabul phenomen flight stellar beat brand innov adob |
| 90 | peopl job talk thing bad make manag lot walk hand end kind put person head feel face fire smile stay pull dead leav boss |
| 96 | work compani make employe posit effort put feel creat reward contribut hard part impact step success extra made real individu huge recogn import show |
| 98 | compani con pro posit part corpor real start find startup found person type feel downsid kind tech biggest mention big import major fit love |

**Table D2. 65 Selected Topics (Including Combined Topics)**

| Topic # | Label | Top 25 words |
|---|---|---|
| 38 | aggressive | top compani line bank aggress corpor manag bottom start firm invest notch employe privat passiv financi industri front busi heavi public account gener client |
| 11 | poor leadership | lack manag leadership commun poor direct clear senior vision plan strategi account execut lead due depart process busi weak structur organ top chaotic reactiv |
| 71 | hostile management | manag employe hostil unprofession abus favorit hr behavior bulli bad horribl rude disrespect neg treat staff lie harass supervisor practic uneth toxic yell creat |
| 84 | blame culture | problem issu manag point blame review solv employe thing neg person fix address respons finger wrong mistak pass real account hr find understand deal |
| 24, 76 | work life balance | work life balanc good healthi excel flexibl person compens maintain brown perfect nose home worklif emphasi sheet compens superb strike conduc workahol memori emphas basic |
| 51 | fair compensation | good benefit pay decent great compens work fair packag salari competit excel locat perk train stabl offer cowork gener adequ fairli pto master descent |
| 58 | competitive compensation | salari benefit competit pay averag compens good industri compar low standard market lower packag offer work fair similar competitor par higher slightli area scale |
| 78 | benefits | benefit health great insur plan employe medic pay vacat paid offer match excel packag gener includ care time compens program stock bonu reimburs retir |
| 29 | bureaucracy | make decis slow process thing risk difficult lot made move busi red polit corpor tape bureaucrat improv conserv bureaucraci progress impact frustrat lead avers |
| 3 | caring environment | employe care treat manag famili staff compani patient genuin hospit nurs health owner equal number resid gener employ clinic children show import understand treatment |

| | | |
|---|---|---|
| 20 | challenge | difficult work time make challeng move thing hard find tough deal person peopl adjust face type bit quickli understand job adapt thrive statu busi |
| 22 | dynamism | work challeng project interest client dynam excit lot stimul divers intellectu technic technolog engag reward assign task team varieti involv intern great collabor offer |
| 91 | fast pace | fast pace challeng grow dynam move excit slow face learn quickli enjoy fun bore quick thrive demand extrem day constantli adapt handl toe intens |
| 60 | community | strong compani sens commun ethic employe teamwork integr corpor leadership focu excel safeti posit divers commit collabor promot pride profession core emphasi organ famili |
| 28 | performance | perform base polit promot highli competit reward review system driven recognit rank manag intern merit compens top individu compet peer talent result evalu process |
| 79 | cut-throat | cut technolog edg throat due cost constant layoff busi lead econom compani budget year unstabl competit creat recent frequent result continu reduc futur increas |
| 15 | safety | work control condit shop clean equip manag safeti air plant hot physic dirti system build area poor cold run qualiti floor manufactur mental offic |
| 33 | procedural | corpor polici govern rule union contract employe agenc due state standard oper employ time issu strict contractor procedur polit regul public organ militari ad |
| 88 | creative | creativ innov collabor encourag entrepreneuri support foster freedom challeng idea highli dynam motiv product independ talent individu inspir initi creat spirit teamwork design reward |
| 75 | consultative | busi firm client consult model market oper real practic partner unit run success understand side strong strategi focu develop area focus invest industri aspect |
| 4 | customer focus | custom servic product qualiti employe client focus focu compani base care deliv satisfact excel relationship experi improv result serv centric happi sale interact creat |

| | | |
|---|---|---|
| 95 | brand focus | compani industri product market brand strong leader mobil innov great reput global world big upward excit media lead competit interest digit field financi stabl |
| 35, 19 | learning and training | experi train learn program skill support knowledg resourc gain manag technic profession opportun develop lot excel improv intern set human abil mentor share job industri |
| 54, 56, 21 | development and growth | career opportun profession advanc growth limit room person develop lot learn progress great grow path compani support skill potenti plenti excel start move challeng promot |
| 77 | growth oriented | compani grow continu improv growth busi start challeng maintain quickli pain rapidli constantli startup evolv success move thing excit expand organ small rapid process |
| 72 | gender diversity | divers boy women club group promot corpor school cliqu domin male bit femal network type mental age social men part conserv inclus polit workplac |
| 66 | engineering | technolog develop product engin softwar system design process tool project latest tech technic agil date test innov comput market outdat practic learn data exposur |
| 57 | team excellence | great work excel team benefit support fantast colleagu worker train outstand leadership peer superb solid compens terrif enviro brilliant teamwork postiv ambienc rough etho |
| 37 | family oriented | famili orient friendli close team feel small type compani knit busi tight group part warm worker care friend commun cowork run felt owner home |
| 14, 61 | flexible | work flexibl home schedul hour friendli time great abil relax option fun remot pay casual easi independ benefit good cowork comfort decent freedom telecommut worker |
| 50 | demanding schedule | hour work long time flexibl stress expect pay week weekend day retail overtim requir shift season lot job demand holiday busi night hard extra |

| | | |
|---|---|---|
| 55 | serenity | work nice good friendli clean comfort pleasant worker safe facil profession peopl physic healthi campu modern quiet workplac peac calm neat cowork beauti condit |
| 16, 74 | friendly | work good friendli nice posit colleagu gener peopl profession boss enjoy cooper pleasant collegu relax salari worker helpful support payment upbeat workmat workplac relationship comfort |
| 52, 30 | collegial | staff nice friendli good work peopl support offic profession gener colleagu relax member locat warm worker excel pleasant collegi cowork cooper pretti experienc ambianc knowledg |
| 2, 17, 25, 42, 45 | fun | great make peopl work love benefit feel lot fun amaz team enjoy smart worker awesom fantast day meet friendli absolut thou friend nice cowork product |
| 10 | pet friendly | great work friendli worker fun peopl dog amaz eat love offic team famili cowork posit workplac super bring pet kind beat camaraderi anim bee |
| 43 | party | fun activ event compani lot social team employe parti offic sport commun happi build involv meet celebr perk outing regular includ engag holiday monthli |
| 53 | youth/energy | fun young work peopl energet lot profession excit dynam offic vibrant creativ upbeat energi age youth cool enthusiast colleg start motiv workforc make love |
| 8, 39 | global | opportun compani intern corpor divers local experi offic world american travel busi exposur countri global japanes wide understand multi india client bank project speak learn |
| 68, 70, 97 | laid back | back casual work laid relax friendli pretti dress easi offic code fun stab wear nice fairli busi cool gener comfort cowork peopl watch jean worker |
| 92 | exploration | learn lot great opportun thing experi work intern respons curv challeng freedom hand stuff exposur network alot resum chanc quickli scope teach steep travel |
| 0 | location | offic locat beauti area citi campu live great build park conveni downtown commut facil close site central easi san view access work small town |

| | | |
|---|---|---|
| 40 | low pay | pay low salari rais wage benefit job increas decent promot stress minimum year bonus bonu rate start poor hourli paid averag base worker expect |
| 41, 80 | hierarchy | manag level upper senior micro employe middl staff support posit style entri understand execut poor higher approach lower direct director touch top care mid listen |
| 83 | merger | compani year past recent corpor improv start acquisit made lost ago shift employe merger wors coupl acquir continu move major due left declin complet |
| 32 | sharing | idea open share employe encourag manag improv collabor opinion inform feedback listen innov heard suggest knowledg voic freedom express creativ team implement discuss posit |
| 65 | open door | open manag polici door friendli employe commun mind transpar hr approach compani honest support collabor close access revolv senior feedback easili easi leadership inform |
| 81 | open space | offic work open space quiet cubicl bore dull bit build moment distract wall floor desk comfort cool small modern cube light depress loud area |
| 64 | work hard/ play hard | work hard play reward fun game enjoy mental harder find recogn dedic type favorit worker beat recognit role prepar push notic music acknowledg parti |
| 86 | perks | free food lunch fun perk coffe snack gym drink game offic friday beer room break lot park tabl kitchen site cafeteria event parti cater |
| 87 | politics | corpor structur organ polit process flat larg hierarchi highli organiz compani intern bit rigid extrem organis hierarch lack bureaucrat system conserv defin big tradit |
| 36 | profit | sale commiss sell competit product make custom base goal train pressur market potenti earn money driven hit manag rep number target push retail incent |
| 46 | stable | job work secur stabl easi posit stabil stress safe employ satisfact bore enjoy steadi duti task respons find descript repetit part titl fairli comfort |
| 13 | size | compani small big larg corpor feel size famili firm smaller startup larger town part start organ busi group bigger number agenc compar typic resourc |

| | | |
|---|---|---|
| 94, 82 | stress | expect high work stress time pressur demand turnov meet low task school deadlin rate workload energi project level manag perform complet |
| 85 | supportive | employe posit compani creat support manag encourag care healthi promot relationship build engag happi maintain trust foster motiv workplac strive teamwork success genuin fellow |
| 6 | selection, turnover, and promotion | hire peopl compani fit posit job process promot recruit interview person talent find intern experi role candid type qualifi bring skill fire fill hr |
| 67 | individual excellence | peopl smart talent great work intellig motiv collabor passion incred colleagu group surround amaz highli interest individu dedic driven bright creativ hardwork young fantast |
| 49 | teamwork | team member support orient collabor strong manag work build spirit posit leadership player leader encourag excel foster true promot execut cohes mate cooper focus |
| 62 | goal oriented | team goal success result achiev driven individu orient motiv focus posit set common organ drive succeed perform compani win person reward competit collabor creat |
| 1, 59 | hostility | poor work manag creat employe stress moral hostil low neg bad extrem lack toxic terribl make pay tens commun made staff uncomfort horribl unhappi leadership |
| 93 | fear | fear creat manag employe toxic ceo neg constant fire lead intimid blame hostil leadership micromanag distrust base staff bulli trust run senior presid top |

| | | |
|---|---|---|
| 99 | gossip | polit offic manag corpor extrem toxic bad horribl terribl neg unprofession gossip drama workplac boss backstab micromanag complet ego immatur charg petti worst run |
| 26 | vacation | day time hour week work paid vacat shift leav holiday year month sick night earli break end start weekend home overtim schedul pm expect |
| 12 | vision | compani leadership team strong mission ceo vision leader execut organ product creat posit amaz core great solid senior inspir driven passion commit focus live |

Undertaking this exercise reinforced for us the advantages of our chosen empirical approach over manual coding for studying the distribution of cultural beliefs between and within individuals. First, it became clear that coders brought different interpretive lenses to the exercise as a function of their idiosyncratic past experiences in organizations and cultural backgrounds. For example, two coders initially excluded topic 0 on the basis that it appears to be about a firm's physical location rather than about the firm's culture. Yet one coder saw in it a culturally meaningful distinction between urban and suburban work cultures. Although there were 18 topics that we chose to drop, we suspect that a different research team might have seen some of these as culturally meaningful and instead chosen to include them. Second, some of the choices we made about which categories to combine or separate seemed arbitrary. For example, we ultimately chose to separate topic 51 (fair compensation) from topic 58 (competitive compensation), but we also think it would be completely reasonable to combine these topics. Similarly, although we combined topics 2, 17, 25, 42, and 45 into a category labeled "fun," we wondered whether a younger and hipper research team might have seen subtle distinctions among these topics that somehow eluded us.

Remaining mindful of these concerns about our manual coding procedure, we proceeded to create measures of interpersonal and intrapersonal heterogeneity based on the 65 topics identified in table D2. These measures were highly correlated with their corresponding measures from the 500-topic model (correlation coefficient of .76 for interpersonal heterogeneity and .83 for intrapersonal heterogeneity). Tables D3 and D4 reproduce our main empirical models in tables 4 and 5 using the 65-topic measures. The results are substantively identical to our main results, demonstrating that our empirical strategy recovers cultural signals even in the absence of human interpretation and manual coding.

This supplemental analysis supports our use of measures based on unsupervised learning of cultural topics for studying the distribution of cultural beliefs between and within individuals. Yet we believe an approach involving human interpretation and coding of cultural categories likely has some advantages for studying particular cultural *content*, such as when researchers have identified specific cultural themes and their linguistic signatures a priori. We encourage future work measuring culture using text to carefully consider the relative merits of each approach for different applications.

**Table D3. ROA on Interpersonal Heterogeneity***

| | (1) OLS | (2) Matched |
|---|---|---|
| Lag interpersonal hetero. (65 topics) | –.67••• | –.53• |
| | (4.34) | (2.05) |
| Lag log assets | .29•• | |
| | (2.78) | |
| Lag log # reviews | .18 | |
| | (1.33) | |
| Constant | –.48 | .50 |
| | (.48) | (1.14) |
| Quarter FEs | Yes | No |
| Industry FEs | Yes | No |
| Stratum FEs | n/a | Yes |
| Firm/quarters | 3283 | 735 |

• $p < .05$; •• $p < .01$; ••• $p < .001$.
* Absolute t-statistics are in parentheses; standard errors are clustered by firm.

**Table D4. Tobin's Q/Patent Outcomes on Intrapersonal Heterogeneity***

| | (1)<br>TQ<br>OLS | (2)<br>TQ<br>Matched | (3)<br>Log #<br>patents<br>OLS | (4)<br># Patents<br>Neg. Bin. | (5)<br>Log #<br>patents<br>Matched | (6)<br># Back. cites<br>OLS | (7)<br># Back. cites<br>Matched |
|---|---|---|---|---|---|---|---|
| Lag intrapersonal hetero. (65) | .098$^{\bullet\bullet}$ | .15$^{\bullet}$ | .22$^{\bullet\bullet}$ | .43$^{\bullet\bullet\bullet}$ | .13$^{\bullet}$ | .20$^{\bullet\bullet\bullet}$ | .25$^{\bullet}$ |
| | (2.77) | (2.44) | (3.07) | (3.62) | (2.12) | (3.43) | (2.00) |
| Lag log assets | –.11$^{\bullet\bullet}$ | | .22$^{\bullet\bullet}$ | .85$^{\bullet\bullet\bullet}$ | | .14$^{\bullet}$ | |
| | (3.03) | | (2.78) | (5.97) | | (2.38) | |
| Lag log # reviews | .16$^{\bullet\bullet}$ | | .067 | .084 | | –.14 | |
| | (3.13) | | (.67) | (.39) | | (1.79) | |
| Constant | 2.83$^{\bullet\bullet\bullet}$ | 1.05$^{\bullet\bullet\bullet}$ | –1.24 | 1.49$^{\bullet\bullet\bullet}$ | .023 | 1.92$^{\bullet\bullet\bullet}$ | 1.57 |
| | (8.60) | (6.06) | (1.48) | (8.57) | (.20) | (4.04) | (1.14) |
| Quarter FEs | Yes | No | Yes | Yes | No | Yes | No |
| Industry FEs | Yes | No | Yes | Yes | No | Yes | No |
| Strata FEs | n/a | Yes | n/a | n/a | Yes | n/a | Yes |
| Firm/quarters | 3253 | 679 | 3296 | 3296 | 703 | 949 | 349 |

$^{\bullet}$ $p < .05$; $^{\bullet\bullet}$ $p < .01$; $^{\bullet\bullet\bullet}$ $p < .001$.

* Absolute t-statistics are in parentheses; standard errors are clustered by firm.